Real-time Human Proxy : An Avatar-based Interaction System

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. 3213, pp.419-425, 2004-09. Springer Berlin, Heidelberg
バージョン：
権利関係：(c)2004 Springer

Real-time Human Proxy: an Avatar-based Interaction System

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. This paper describes techniques to improve human representation on an avatar-based interaction system, using real-time motion sensing and human action symbolization. Avatar-based interaction systems with computer-generated virtual environments have difficulties in acquiring user's information, i.e., enough information to represent the user as if he/she were in the environment. This mainly comes of high degrees of freedom of human body and causes the lack of reality. Since it is almost impossible to acquire all the detailed information of human actions or activities, we, instead, recognize, or estimate, what kind of actions have occurred from sensed human motion information and other available information and re-generate detailed and natural actions from the estimated results. In this paper, we describe our approach, Real-time Human Proxy, on both side of acquiring and representing human actions. Also we present experimental results.

1 Introduction

Modern communication technologies have made it usual to have communication for people who are distant from each other. For example, telephone is one of the most useful tools to communicate using acoustic information, and e-mail is one of them using literal information. Furthermore, information-richer tools are also available. Video-chat or video-phone are becoming available by data compression techniques and broader network bandwidth. Many people may consider video-phone to be the richest communication tool, since video-phone seems to provide us the feeling of being connected to the other speaker's place, because appearances of speakers are directly transmitted to each other. It has problems, however, in case that many (10 or more) people are participating at once, as on a conference or a meeting. Every participant in a conference must have numbers of windows showing the other participants on his/her display. Each window collocated on the same plane has its own geometrical coordinate system, showing his/her companion who is always just facing the front. This situation causes the user to have spatioperceptual inconsistency and causes the lack of reality. For instance, this does not make participants feel as if they were present alongside. To solve this problem and to establish the spatioperceptual consistency, we have chosen the way to get all participants into a virtual space.

There are several researches on virtual environments for human interaction. In these researches, a 3-D virtual space is reconstructed, in which each participant is represented as an avatar by computer graphics techniques. Through the reconstructed virtual space, each participant sees and hears other participants' activities from the position where his/her avatar is represented. Their positional relations are consistent virtually. This means that each participant can understand where other participants stand, look at, and point to, or understand where a sound comes from, and also move around in the virtual space. In contrast to a video chat system, participants can easily feel coexistence. However, these avatar-based interaction systems have difficulty in controlling avatars, where the degrees of freedom of human body are so high that legacy input devices are not sufficient to acquire or input participants' activities.

In this paper, we describe Real-time Human Proxy; our concept to provide realistic virtual-space-based communication.

2 Real-time Human Proxy

2.1 Human Motion Sensing

In avatar-based interaction, an avatar is expected to reflect activities of a participant into a virtual space as if he/she were there. An avatar is expected to turn its head to a particular person when a participant does so, for example. Nevertheless, as already mentioned above, legacy input devices are not sufficient to acquire participants' activities in aspects of quality and quantity. Using such devices, participants have to keep feeding their own activities into a system by hand, and acquired information may not be precise. Special input devices are often used as solution to this problem[1]. We have developed a vision-based motion capture system (MCS)[2] as an input device, for acquiring more information of participants without compelling them annoying operations.

2.2 What is Real-time Human Proxy?

Although using an MCS as an input device, it is not possible to reflect all of participant's motions into an avatar. Although more information could be acquired in comparison with legacy devices, an MCS cannot acquire all information such as shape of hands, or movement of eyebrows and a lip. Just feeding captured data into an avatar, the lack of information may cause unnatural avatar motions. On the other hand, it is not necessary for an avatar to act exactly the same as participant's motions, since participants usually want to know only what others are doing, but how others moving. Real-time Human Proxy (RHP) is a new concept for avatar-based interaction, which makes better use of an MCS, and makes avatar act more meaningfully.

RHP is a concept which virtualizes a human in the real world in real-time. The aim is to make an avatar act as if he/she in a distant place is present in a virtual space. Therefore, we focus on acquisition and representation of human

action or nonverbal information. We use symbolization on acquisition, and use motion model and behavior model on representation, that they are illustrated in Fig.1 and described below.

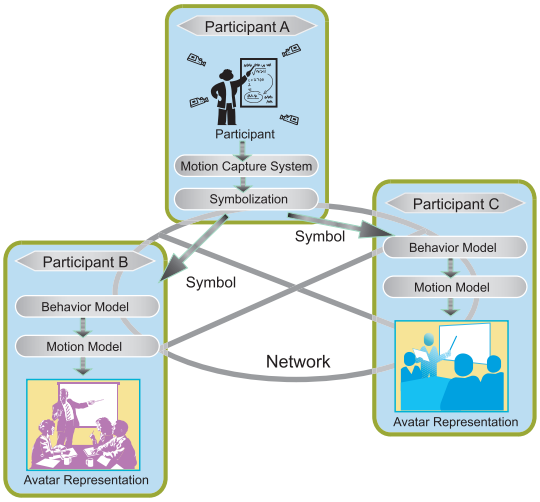


Fig. 1. The concept of RHP.

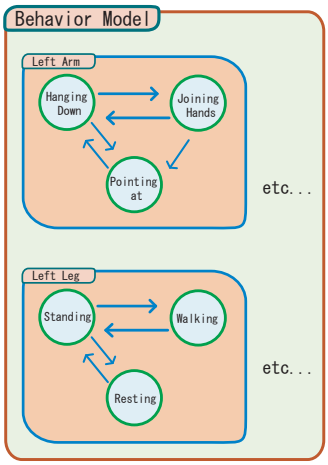


Fig. 2. Behavior Model (example).

Symbolization On RHP, we acquire human actions instead of human motions. We categorize motion sequences into pre-defined actions, expressing them as symbols. Each symbol is formed by a label of an action and its parameters, such as "walking (ν_x, ν_y)" where (ν_x, ν_y) is the velocity of participant. After recognizing human actions from captured motion data, the system transmits the symbols to the representation side of the virtual space.

Motion Model Since each symbol transmitted from a participant is based on a label of an action, to make an avatar act, therefore, pre-defined knowledge is needed, which is a set of trajectories or motion sequences of each body part, tied to each symbol. We call the knowledge *motion model*.

Behavior Model We have another model called *behavior model* for describing the human actions (see Fig.2). Behavior model is a set of state transition graphs that decides the next action an avatar is going to do. Each graph corresponds to a body part such as right and left arm, right and left leg, etc. and each state in the graphs corresponds to an action or a symbol such as "walking", "raising hand", and "pointing with finger." When a symbol is transmitted, the current state of the graph is forced to transit to the state corresponding to the symbol. In addition, an avatar often freezes if the avatar acts only when symbols are transmitted, since no symbols are transmitted when a participant

does not make any pre-defined actions. Needless to say, such avatar's behavior does not seem natural. To solve this problem, behavior model has some actions invoked spontaneously such as "folding arms" or "sticking hand into a pocket." These actions may work for time filling, and may make participants feel more natural[3]. And of course, these actions do not indicate participant's intentions in order not to influence interaction between participants.

Representation of Virtual Space Using the above models, an avatar behaves in a virtual space according to transmitted symbols. At first, a behavior model changes transition probabilities in its transition graphs so that the avatar's state transits to the state corresponding to the symbol just received immediately. Secondly, the avatar performs an action referring to the motion models tied to the state just have been activated in the behavior model. A participant is able to see a virtual space in which any participants, including him/herself, are represented as avatars. Therefore, the viewpoint is anchored where the avatar was represented.

2.3 Benefits

The use of an MCS liberates participants from annoying operations of controlling an avatar. On the symbolization process, the system recognizes the action that a participant makes. This operation is intuitive and natural since participants make the same action as they make in the real world. The behavior model can maintain consistency in transition of action. Unnatural transition of action (or a behavior) will not be provided, where any transition from an action to another is a production of transition probabilities in the behavior model. That means impossible transition of action can be prevented. The concept, i.e., symbolization of actions, and representation of an avatar using the motion model and the behavior model, makes representation process simple. As described above, acquired information about participant's motion does not have detailed information of several body parts such as angles of fingers, which must be compensated proving conformity with other body parts to provide natural representation. Using RHP, actions are already recognized, and it is possible to generate full body motion sequence in an arbitrary manner referring to the motion model.

Moreover, this allows avatars to be designed beyond constraints of physical structure. Ordinarily, each avatar is designed to fit an actor whose motions are captured, so that 3-dimensional positions of body parts acquired from an MCS is correctly represented in computer graphics. Some errors in interpretation of posing can occur from mismatches in the sizes of body parts, such as the height of body or the lengths of arms. RHP transmits the symbolized action, not the 3-dimensional positions directly, and consequently RHP largely relaxes the constraints. We can use any kind of avatar including higher body, shorter arms, bigger head, etc. This improves not only usability of the system, but also variety of avatars in interaction.

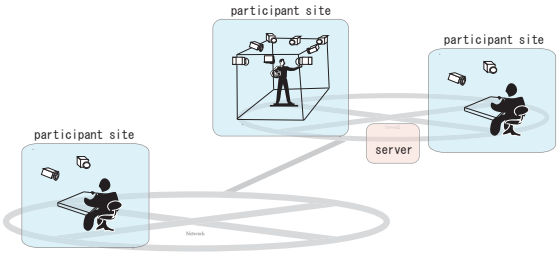


Fig. 3. System configuration of the game

3 Prototype System of RHP

3.1 Interactive Game

First, we have developed a simple interactive game system to verify whether participants can naturally interact each other with RHP in case that all actions necessary for interaction can be listed to recognize and display.

In this experiment, there are three participants as shown in Fig.3. Motion information about each participant is captured by a full-body motion capture system[2] or a upper-body motion capture system[4] we have developed. Each participant wears a HMD to see what he/she want to see in a virtual environment.

The rules of the game, which is a simplified version of a famous game in Japan, are as follows:

1. One of participants becomes a leader.
2. The leader says “A” and points to one of participants.
3. The participant who is pointed to at step 2 says “B” and points to one of participants.
4. The participant who is pointed to at step 3 says “C” and puts his/her hands up.
5. The participant who puts his/her hands up and becomes a leader.
6. Return to step 2 until someone fails.

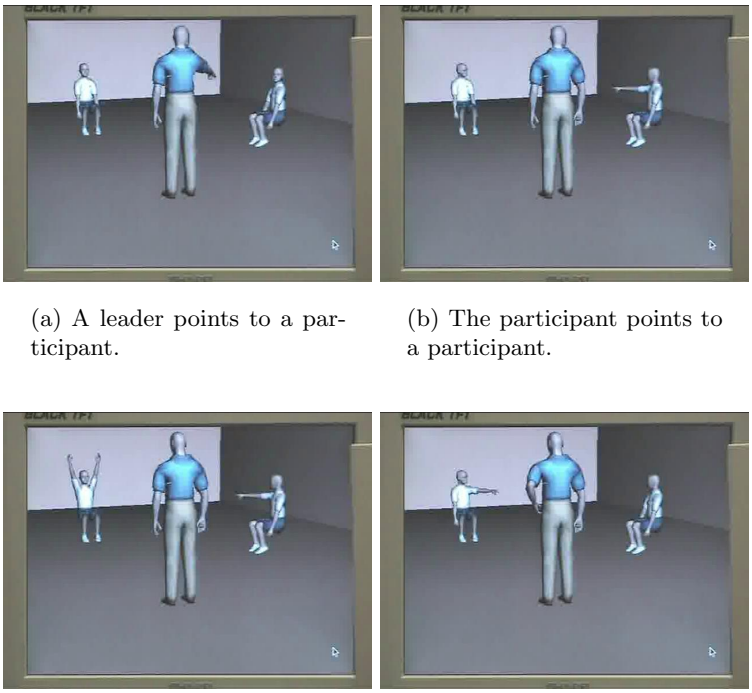
Symbols transmitted among participants are as follows:

- finger pointing
- hands up
- head turn

And, voice information is transmitted via microphones and headphones.

The results show the participants can interact each other. Some snapshots of the game are shown in Fig.4. The impressions of participants are as follows:

- At the beginning of interaction, participants doubted that avatars were controlled by other participants since they was able to see only avatars represented in computer graphics. However, such feeling became fainter in a short time since avatars acted in time to the game flow.



(a) A leader points to a participant.

(b) The participant points to a participant.

(c) The participant puts his/her hands up.

(d) The participant becomes a leader and the next turn starts.

Fig. 4. Snapshots of the game

- Each participant was able to see what he/she wanted to see.
- Each participant felt that others looked at him/her since he/she saw that avatars turned its head.
- Participants were able to easily understand where avatars looked and pointed.
- Using an HMD made the interaction more natural than a desktop display since each participant was able to easily control his/her view.

3.2 VEIDL

We are also developing a system called VEIDL (Virtual Environment for Immersive Distributed Learning), which is a prototype system for estimating RHP. VEIDL is a virtual classroom environment where geographically dispersed people can attend through avatars of teachers or students represented, as shown in Fig.5.

On VEIDL, every participant has his/her own microphone and cameras as input devices, which acquire verbal and nonverbal information from him/her, and

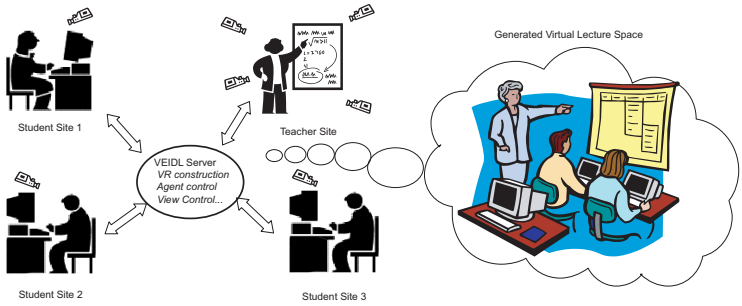


Fig. 5. The Concept of VEIDL.

a display unit (possibly HMD) and speakers as output devices, which present a scene of the classroom generated by a computer. Acquired information from each site is transmitted via network to one another. Each participant can see a computer-generated scene of the classroom from the viewpoint of his/her avatar, and, of course, can also see the other participants behaving according to the transmitted information. This makes participants able to interact each other through the virtual environment. The advantage of dealing with a virtual classroom, as a prototype of RHP, is that it is easy to decide which information (or symbol) should be transmitted, since the objective of interaction in a classroom is clear.

4 Conclusion

In this paper, we propose a concept of real-time human proxy for avatar-based interaction systems, which virtualizes a human in the real world in real-time, and which makes the virtualized human behave as if he/she were present. To verify the effectiveness of RHP, we have developed an interactive game system. And, we introduce VEIDL, a virtual environment for immersive distributed learning.

References

1. Roussos, M., Johnson, A.E., Leigh, J., Vasilakis, C.A., Barnes, C.R., Moher, T.G.: Nice: combining constructionism, narrative and collaboration in a virtual learning environment. ACM SIGGRAPH Computer Graphics **31** (1997) 62–63
2. Date, N., Yoshimoto, H., Arita, D., Yonemoto, S., Taniguchi, R.: Performance evaluation of vision-based real-time motion capture. In: Proc. of Workshop on Parallel and Distributed Computing in Image Processing, Video Processing, and Multimedia, in IPDPS CD-Rom Proceedings. (2003)
3. Arita, D., Taniguchi, R.: Non-verbal human communication using avatars in a virtual space. In: Proc. of International Conference on Knowledge-Based Intelligent Information and Engineering Systems. (2003) 1077–1084
4. Yonemoto, S., Taniguchi, R.: A direct manipulation interface with vision-basedhuman figure control. In: Proc. of HCI International. (2003) 811–815