FRAMEWORK FOR VIRTUAL COLLABORATION EMPHASIZED BY AWARENESS INFORMATION AND ASYNCHRONOUS INTERACTION

Tomo Matsuda, Naoki Shibata, Keiichi Yasumoto, and Minoru Ito*

Graduate School of Information Science, Nara Institute of Science and Technology
Nara 630-0192, Japan
*Department of Information Processing and Management, Shiga University
Shiga 522-8522, Japan

ABSTRACT

In this paper, we propose a framework which allows remote users to form conversation groups based on spatial relationship in a shared virtual space. Our proposed framework can transport awareness information of real world by capturing and transferring user's audio visual information. Our framework also provides functions useful to CSCW, which allow each user to simultaneously join different conversation groups, and communicate with others asynchronously exchanging awareness information. We show a reference implementation architecture to realize the framework in an ordinary computing and networking environment.

Index Terms— DVE, CSCW, collaboration, awareness

1. INTRODUCTION

Widespread broadband network has changed the type of information exchanged among people from text to rich audio visual data. Nowadays, video conferences are held among ordinary people in their homes. The large advantage of video conferencing is that *awareness information* can be exchanged by audio visual conversation. Awareness information is knowledge gained from surrounding environment. In the case of video conferencing, the information includes gestures, expressions and tone of the voice, which helps distant users to smoothly communicate with each other [1]. However, video conferencing is not superior to conferences held in real space in terms of transferring awareness information, and communication efficiency [2]. One of disadvantages of video conferencing against real space conferences is that in real space, attendees can freely change their positions to form sub-groups and/or whisper to right next person [3].

On the other hand, Distributed Virtual Environment (hereafter, *DVE*) such as network games and Second Life [4] are systems where distant users can freely move in the virtual space, find communication partners, and communicate with the partners in appropriate distance. However, capability of exchanging awareness information is limited.

There are other weak points of these communication methods. These methods assume that users join the system at the same time and they communicate like phone. However, asynchronous communication through DVEs which are capable of transferring awareness information would also be useful. There is no existing communication method which supports all of these communication functions.

In this paper, we propose a new DVE framework which realizes the following functions: (1) users can intuitively and naturally form conversation groups by changing his/her position in a space and communicate with others with awareness information; (2) each user can participate in multiple conversation groups and have conversations at the same time; and (3) each user can asynchronously communicate with others by showing his/her presence and awareness. For (1), we prepare a virtual space shared among users like network games where a user changes his/her positions in the virtual space by manipulating his/her *avatar*, and allow users to exchange audio visual information depending on their virtual positions. For (2), we propose a function to allow each user to replicate his/her avatar, place the replicas at different positions in the space, and switch focus among the replicas. For (3), we propose a function that each avatar replica records audio visual messages from other users, and the corresponding user can play the messages back at arbitrary time.

In the following Sect. 2, we address the related works. In Sect. 3, we present the details of the proposed framework. Sect. 4 describes implementation issue. Sect. 5 provides experimental validation. Finally, we conclude the paper in Sect. 6.

2. RELATED WORK

When we have many users in a shared virtual space and each user receives all information transmitted by all the users, it will be hard to have a conversation due to too much information. So, we need a mechanism to select only necessary information for each user. There are some studies on efficient communication in a virtual space in terms of this problem. In DIVE (Distributed Interactive Virtual Environment) [5], three concepts called *aura*, *focus*, and *nimbus* were introduced. Aura for object o is defined as a spherical region whose center is o. Only when auras for two objects have intersection, interaction between two objects is allowed. Focus defines a region in which objects can be seen by the corresponding user. Nimbus is a region in which an object can be seen. Only when nimbus and focus has intersection, the user can see the object. The concepts of aura, focus, and nimbus are widely used in various DVEs. RAVITAS [6] is a system in which users communicate with voice chat in a virtual space using 3D sound effect, where cocktail party effect is emphasized based on the concepts of focus and nimbus.

There are some studies on tele-immersion where real spaces at multiple different locations are composed into one unified virtual space and the users can view the virtual space from their virtual positions/directions. TEEVE [7] realizes such a tele-immersion environment by using 3D cameras and high-speed network. However, this approach is costly when we want to realize virtual space communication with awareness in an ordinary computing and networking environment.

This work was partly supported by HAYAO NAKAYAMA Foundation for Science and Technology and Culture.

3. FRAMEWORK

Our framework targets applications in CSCW area such as knowledge creation and e-learning, where rich communication including facial expressions and gestures plays an important role.

Our basic ideas for controlling awareness transmission are as follows: (1) information required to recognize presence and location of participants (users) is shown to each user based on the spatial relationship among the users in a virtual space; and (2) a communication group is automatically formed by users in the neighborhood of the virtual space, and presentation of the information in the communication group is emphasized. In Fig. 1, interactions between A and B are emphasized because their auras are overlapped and they form a communication group.

Our framework have the following functions: (1) virtual space sharing; (2) voice communication; (3) video communication; and (4) avatar replication.

Virtual space sharing function This function provides a 3D visible space as a place for communication. Our framework displays an image called *avatar* which plays a role of user's substitute in the virtual space. A user moves his/her avatar freely in the space with arrow keys or a mouse device. Avatar tells his/her presence and location in the virtual space to other users. In real world, a person can see others when they exist in his/her view. Therefore, we set a user's focus as the corresponding avatar's view in the virtual space so that the user can see avatars which are in the user's focus.

Voice communication function This function allows users to talk with each other in the virtual space via voice. In real world, an area in which a speaker's voice is heard depends on the volume of the voice. When a user exists near the speaker, the user hears the voice. Therefore, our framework dynamically sets an appropriate area for a speaker's nimbus based on the volume of his/her voice. When an avatar exists in a speaker's nimbus, the corresponding user can hear the speaker's voice.

Our framework controls the volume of voice so that a user can hear the voice uttered by a speaker who belongs to the user's communication group more clearly. In Fig. 1, user A can hear user C's voice because A exists in nimbus of C. Moreover, auras of A and B intersect and they form a communication group, so A and B can hear each other's voice more clearly.

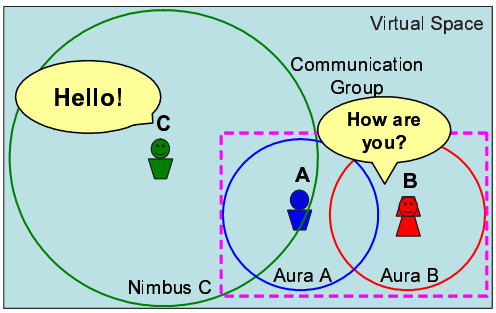

Fig. 1. Model of Voice Communication Control

Video communication function This function displays users' videos captured by their cameras and allows each user to acknowledge facial expressions and gestures of other users in a virtual space. In real world, we generally focus on facial expressions and gestures of only persons we are currently communicating with. Therefore, our framework displays, on a user's display, only videos of users in the user's communication group.

Avatar replication function In CSCW, it is convenient if a person such as a leader or a teacher can participate in multiple groups and interact with members of all the groups at the same time. Also, it is convenient if asynchronous communication is possible if real-time interaction is not required. For example, if someone has to manage multiple projects which progress in different places at the same time, he/she would want to participate in the corresponding communication groups.

Avatar replication function allows a user to make replicas of his/her avatar at multiple places in a virtual space. The user can put the replicas at the places where the avatar has visited as shown in Fig. 2. One of the replicas is specified as the *main body*, and the user can freely change the main body among the replicas. The user can see the multiple views from all the replicas, but the voice and video communication functions are available only through the main body. Fig. 3 shows a screen image, where a window on left top side shows the view from the main body, and two windows on left bottom side show the views from the other replicas. Two windows on right side of Fig. 3 show the virtual space map and the list of replicas, respectively, through which the user can select the main body.

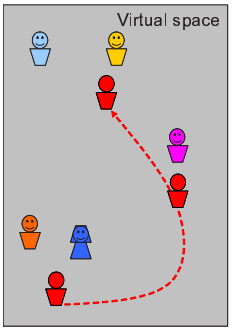

Fig. 2. Avatar Replica

Fig. 3. Multiple Views from Replicas

Our framework allows only the main body to communicate with other users via voice and video, because it is difficult for a user to follow multiple conversations at the same time. However, the user can grasp situations at multiple places from replica windows and move to a place with necessity for communication by dynamically changing the main body.

Moreover, our framework provides an asynchronous communication function through avatar replicas. Other users can send a voice or video message to a replica. When a message is put on a replica, the corresponding user receives a notification and is able to read the message later.

4. IMPLEMENTATION

We implemented a prototype of our proposed framework based on the architecture shown in Fig. 4. The architecture consists of (1) position information manager, (2) GUI, (3) sound controller, (4) video controller, and (5) message controller. Position information manager manages each user's position in a virtual space. GUI manages display of virtual space information and avatar movements. Sound controller controls voice chat among users. Video controller controls video communication among users in a communication group. Message controller controls asynchronous communication of

avatar replication function. Each module communicates with other modules through position information manager as shown in Fig. 4. We implemented the above modules in Java language. Position

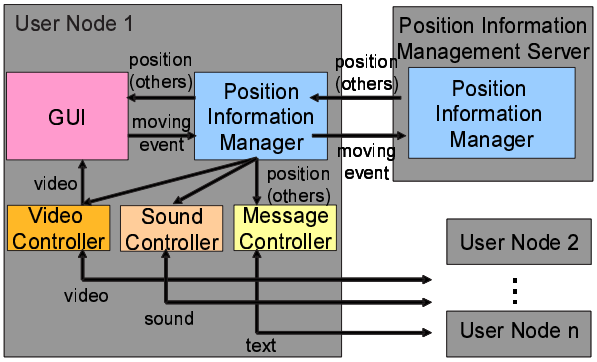

Fig. 4. Implementation Architecture

information manager allows users to share positions of their avatars and replicas. Management of position information is realized in a client/server fashion by using a dedicated server assigned for all/part of the virtual space. Position information manager on each user node notifies the server when GUI detects an event of the avatar movement. The server manages all avatar positions notified by user nodes. The server broadcasts all avatars' positions to the user nodes at regular intervals. Position information manager also retains the IP address of the user node for each avatar, which is used in unicast communication between sound and video controllers of user nodes. Position information manager on a user node also manages communication groups. When it finds avatars in its aura based on position information, it automatically adds them to a communication group.

The GUI module displays the virtual space and avatars in 3D graphics. We suppose that images and 3D geometry data of objects are distributed to all user nodes beforehand. 3D graphics rendering was implemented using JOGL (Java bindings for OpenGL).

Sound controller controls capturing/transfer of voice data from/among user nodes, 3D sound effects based on spatial relationship among avatars, and voice output at user nodes. This module was implemented based on the mechanism proposed in [6].

Video controller controls capturing a video from a web camera at a user node and video data transfer among user nodes in a communication group. When position information manager notifies the video controller that a new member joins a communication group, it establishes a unicast connection to the corresponding user node (new member) and starts transmitting a video to the user node. We used JMF (Java Media Framework) to capture video data from a web camera. Video controller buffers received video, and asks GUI to display the video in a balloon over a virtual space as shown in the main body window of Fig. 3.

Message controller controls multi-media message exchange between a replica and a user. It identifies the user node corresponding to an avatar replica to which a message was sent, and transmits the message to the node. When a message arrives at a user node, the user is notified through GUI.

5. EVALUATION

In this section, through some experiments, we show the usefulness of the two main features of our framework: (i) video communication function and (ii) avatar replication and asynchronous communication functions.

5.1. Evaluation of Video Communication Function

In general, whether collaborative activities will smoothly progress or not depends on how well participants can exchange necessary information. Therefore, we conducted an experiment according to the following procedures to evaluate whether the video communication function added to a 3D virtual space improves an efficiency of information exchanges.

First, we asked six testees to participate in the experiment and form three pairs arbitrarily in the virtual space using our prototype system. Then, we gave them a "Spot the difference" puzzle like Fig. 5 as their task. For each pair, we gave one picture of the puzzle (i.e., left side picture in Fig.5) to one testee of the pair and the other picture (i.e., right side picture) to the other testee. Two testees of each pair cooperated with each other to solve the problem using the system (case 1).

For comparison, we also asked the testees to conduct the above task using a video conference system (skype, case 3) and using our prototype system without video communication function (case 2). In case 3, all testees participated in a conference session and decide pairs on the session, then each testee established a new connection with his/her partner and started the above task.

We measured the total time to form pairs and solve the task for each pair for the above three cases, then compared the measured time. Also, we conducted a questionnaire for testees on their subjective impression on our system.

The experimental environment was composed of laptop PCs equipped with Intel Pentium M CPU (933MHz) and webcams with 300K pixel CMOS for clients and a desktop PC equipped with Intel Core2 Duo CPU (2.40Ghz) for a server on the 100BASE-TX LAN network.

Fig. 5. Spot the Difference[8]

We show experimental results in Fig. 6. The measured time of case 1 (virtual space with video) and case 3 (video conference) was shorter than the time of case 2 (virtual space without video). This suggests that a function for exchanging information visually by video improves efficiency of information exchange. Actually, we saw that testees often turned the picture toward the camera to show it to the partner.

The time of case 1 (virtual space with video) was slightly shorter than that of case 3 (video conference). Here, note that the time of case 3 includes the time for connection with the partner, and it took about 15 second. On the other hand, in our system, connection to the partner was done in a moment. Difference of total time between case

1 and case 3 will be greater when more than 3 people will conduct collaborative work since reconnection time will be larger as the number of participants increases. When comparing the time taken for "Spot the difference" task itself, our system (case 1) took slightly longer time than the video conference system (case 3). We guess that the reason is that the image captured by webcam and provided to each testee on our system is smaller than the image of the video conference system. Adding a function to adjust size and quality of the video image and evaluating its impact are our future work.

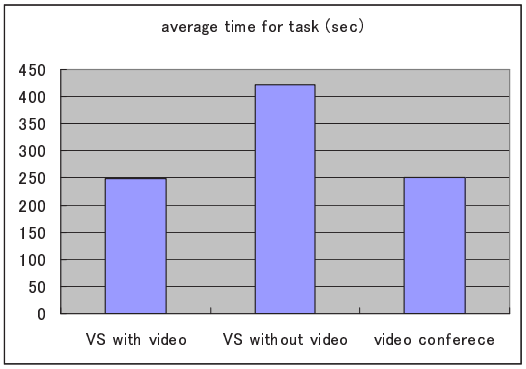

Fig. 6. Comparison among Virtual Space with Video, without Video, and Video Conference System

For the questionnaire, we asked the following ranking questions to testees (1: low rating, 5: high rating): Is the system easy to use? (2) Do you recognize partner's feeling and emotion? (3) Do you feel friendliness to the partner in the system? (4) Do you feel that you are sharing working environment with the partner? (5) Do you feel that you are working as if you were with the partner at the same place?

Table 1 indicates average ranking for our system with video, our system without video, and the video conference system (skype). As a result, our system got an good evaluation for improving senses of feeling friendliness and sharing the same place.

Table 1. Comparison of Subjective Impression

Question	with video	without video	Skype
(1)	3.6	3.6	3.8
(2)	4.0	3.0	4.0
(3)	4.2	2.8	4.0
(4)	4.5	4.2	3.2
(5)	3.3	3.0	2.5

5.2. Evaluation of Avatar Replication Function

To investigate whether the avatar replication function with asynchronous communication is useful or not for collaboration among multiple groups, we conducted the following experiment.

First, we divided seven testees to an instructor and three groups with two persons. Second, we gave to each group three quizzes with a high degree of difficulty. Then, we let the instructor to give some hints to each quiz when requested by a group member.

We asked the testees to conduct the above experiment for two cases: with the avatar replication function (including asynchronous communication function) and without this function. In case without this function, the instructor walked around the three groups. We measured the time until all the groups solved the given quizzes. The experimental configuration is the same as in Sect. 5.1. The experimental result is shown in Table 2.

Table 2. Comparison of Time to Solve Quizzes (with and without Avatar Replication Function)

	with replication	without replication
Time to solve	315 sec	432 sec

The table shows that the time to solve the quizzes was greatly reduced by using the avatar replication function. The reason is that the instructor put his replicas to all the groups and could give hints to an appropriate group by knowing the situation of discussion at each group. Also, with the replication function, each group member could readily send a message to the replica of the instructor even while the instructor was talking to a member of the other group. The instructor could easily respond to the message and gave hints to the message sender by switching his main body to the group. It greatly helped to reduce the total time to solve the quizzes.

6. CONCLUSIONS

In this paper, we proposed a new framework for efficient collaboration in a virtual space by adding a video communication function to an avatar-based distributed virtual environment. Through user study, we showed the usefulness of introducing video communication function and avatar replication function to a virtual space communication. More precise evaluation of usefulness of the proposed frame work is our future work.

7. REFERENCES

[1] Mehrabian, A.: "Silent messages, Implicit Communication of Emotions and Attitudes, 2nd Ed.,", Wadsworth Pub. Co. (1981).

[2] Gaver, W. W.: "The affordances of media spaces for collaboration," Proc. of ACM conference on Computer-supported cooperative work (CSCW92), pp.17-24 (1992).

[3] Hall, T. E.: "The Hidden Dimension Doubleday & Company," NY (1966).

[4] Second Life Home Page, http://secondlife.com/

[5] Benford, S. and Fahlén, L. E.: "A Spatial Model of Interaction in Large Virtual Environments," Proc. of Third European Conference on CSCW (ECSCW'93), pp.107-122 (1993).

[6] Yasumoto, K. and Nahrstedt, K.: "RAVITAS: Realistic Voice Chat Framework for Cooperative Virtual Spaces," Proc. of IEEE International Conference on Multimedia and Expo (ICME2005), (CD-ROM) (2005).

[7] Yang, Z., Nahrstedt, K., Cui, Y., Yu, B., Liang, J., Jung, S., and Bajscy, R.: "TEEVE: The Next Generation Architecture for Tele-Immersive Environment," Proc. of the 7th IEEE International Symposium on Multimedia (ISM'05), pp. 112–119 (2005).

[8] Spot the difference – Wikipedia, http://en.wikipedia.org /wiki/Spot_the_difference/