A Movie-watching System by Muliple Telepresence Robots

Ken Yonezawa and Hirotada Ueda

Division of Frontier Informatics, Kyoto Sangyo University

Abstract: Telepresence systems which paid their attention to the meeting have been developed. We pay attention to entertainment application and implement a system using a dialog robots. The system which users can feel as if enjoying the movie together in the same room even if they are in different a remote places is proposed. By distinguishing and transmitting the person's verbal behavior and nonverbal behavior, the telepresence robot can express the person's presence strongly as the person's avatar. By the technique of measuring the position and posture of users and robots by sensors, and establishing virtual spatial relationship, the user can place robots freely and gaze coincidence of a look is maintained correctly. We designed the three robot arrangement / control methods that flexibility differed. We also proposed the technique of exaggerating a motion of a robot by the feeling presumed from a user's gesture, the tone of voice, and volume of voice as a system which raises presence.

1 Introduction

Face-to-face conversation is one of the most basic forms of communication. To support communication among remote places, videoconferencing systems have been developed. Otsuka et al. [1] developed the MM-Space. In this system, multiple rooms are connected by a network, and the displays in each room are arranged as if users are sitting in a circle. This display screens synchronously move depending on user's head movements to convey speaker's intention. Guizzo et al. [2] proposed the interactive robot which is remotely controlled. They placed priority on what the robot expresses as non-verbal human behavior. Kashiwabara et al. developed the robot named TEROOS [3] which is placed on user's shoulder. This robot faithfully conveys the gesture of the user who is at remote place. Their demonstration showed that this robot can provoke a reaction in social manner like bowing from surrounding people. Hasegawa et al [4]. developed the robot which can move just like user's movements during conversation by recognizing user's unconscious behavior, using Kinect. It proved that their approach has an effect on avoiding the speech collision. The previous researches mainly regarded meetings. Unlike those researches, we focus on the field of entertainments. By using multiple robots, we design the system that can raise the presence so that the users can comfortably and freely enjoy movies with the remote people together as they are in the same room.

(a) movie-watching with real people

(b) movie-watching with telepresence robots

Fig.1. Proposal of the telepresence movie-watching

2 Concept

2.1 Analysis of Demands

An illustration of our concept is shown in Fig.1. In this system, a priority is placed on ralization of the same sense as watching a movie with multiple people in the same room (a), using multiple robots. The presence of human who are enjoying the same movie should be realized. In order to do that, making clear who the speaker is (b-1) and representing emotion of the users through the robots (b-2) are important. This system should also allow the users to sit in a comfortable position to watch movies (c-1) and to freely place the robots in their rooms (c-2).

The approach of Otsuka [1] can be applied to the demand (a) of conveying the presence as if a user can watch a movie with many people in the same room. However, this approach does not meet the demands (b) and (c). These demands should be tackled as a new research task. The approach of Kashiwabara [3] and Hasegawa [4] might be useful for the demand of detecting speakers (b-1) and conveying the emotion (b-2). However these methods work only dialogue on a one-to-one basis and are not enough for the system requirement. The demands of sitting at comfortable position for the user (c-1) and placing the robots wherever the user prefers even if the room layouts are different (c-2) are also the new tasks which have not been discussed in any previous research. In addition, there is the following task. In previous researches, the main focuses was to support meetings or dialogue on a one-to-one basis, therefore, the relationship between a camera and a user is implicitly kept while using the system. In the case of our application of enjoying a movie, such a relation may not be kept anymore. If the system requires that users should remain at a position within angle of view of camera while watching a movie, the users cannot freely eat snacks and drink alcohol, so the users cannot enjoy watching the movie at ease. The new demand is to allow the user to be free to move (d).

2.2 Proposal of Basic Solutions

We propose the following way to solve all of the problems: speaker detection (b-1), estimation of emotion (b-2), allowing users to sit everywhere they prefer (c-1), the freedom of robot deployment (c-2), and allowing users to freely move (d). Visual sensors are installed in each user's room in order to keep on recognizing the position and the posture of the users and robots. Thus even if a user moves around, the position of the user is continuously detected and traced. Audio sensors are also installed in the room. By comprehensively determining the visual and audio information from those sensors, the user's verbal behavior and non-verbal behavior are distinguished, and even emotion of the user is recognized. The robots in the other rooms receive that information and represent the user's presence. In such a way, the system creates the atmosphere as if the user watches a movie surrounding by good friends, sitting at the most comfortable position.

There is an important fact that DOF of robots is much smaller than that of human. All gestures of the user such as finger pointing, clapping hands, raising hands, twisting body, and leaning forward are interpreted into the appropriate robot behaviors. According to the present technology, feeling like joy, anger, humor and pathos can be read in the expression of speech information or a face, and a motion of the head, hand, and foot. Therefore, it seems that it is easy to make a robot imitate the detected motion. However, even if the system transposes a motion of man to a robot's behavior as it is, feeling information is not transmitted well. Here, in addition to suitable interpretation, a suitable exaggeration of motion is indispensable. Contrary to exaggeration, there is a case restraint of the motion is better. It will be only a mere noise to make a robot reproduce a meaningless user's motion, but it will break the atmosphere of comfortable movie appreciation. Therefore, it is necessary to improve the image recognition rate especially about judgment whether the user see the display screen or look at the robot which is an avatar of the friend in other rooms.

2.3 Approach for Installation

The function described above is realized by the rule database which stored a number of descriptions about the meanings of user's movements. It will be also appropriate to prepare a look-up table storing user's feelings and behavior versus robot movement sequences. Thus the robot in the different room can perform the user with the realistic user's behavior and emotion by transmitted information on the behavior and emotion acquired by sensing the user.

3 System Design

3.1 Position, Posture, and Voice Detection

The position (3 DOF) and posture (3 DOF) of robots and users are recognized by the system developed by Nakazato et al [5]. Kinect may be also combined to increase the accuracy. Sound is conveyed by VoIP technology, and mixing is conducted so that users can easily detect a speaker's intention. Headphones are used to avoid overlapping user's voice and TV sound so that user's voice can be recognized. As mentioned above, user's emotion like the level of excitement and disappointment are estimated by tone and volume of user's voice.

3.2 Gaze and Behavior Control

Gaze and Behavior control of robots is shown in Fig.2. Information of position and posture of users and robots are continuously detected and sent to servers. The server in each room determines which direction the user and the robots are looking at, and the information is sent to the other servers. The server also determines intention of the user from verbal and non-verbal information and sends this intention to the other servers. After receiving that information, the server calculates the 3D direction and amount of the head movement and controls the robots. In this calculation, the interpretation and exaggeration mentioned in the former section is also handled. The case that there are two or more users or robots on the line of the gaze may cause misunderstanding. In such cases, the system restricts the movement of the robot.

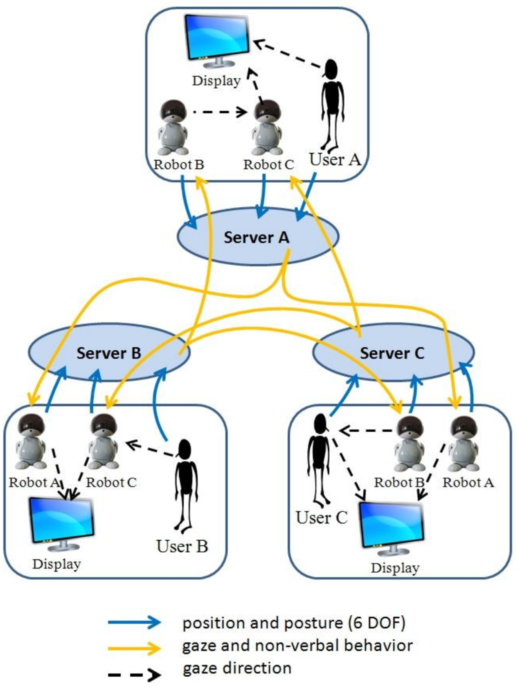

Fig.2. Robot's behavior Control

3.3 Robot Deployment

A user's gaze direction and behavior are detected by the sensors. Without by losing such information, virtual spatial relationships are established in each user's room, and it can allow for the freedom of deployment of the user and the robots. Moreover, we considered how much the flexibility of the deployment is necessary and the problem of robot control which arises as the flexibility increases. Robot deployment is divided into three ways: (a) completely-flexible deployment which users can have their own way to deploy the robots, (b) flexible deployment except for the order of neighbor relation, and (c) completely-fixed deployment, which is the same way of Otsuka [3].

4 Conclusion

We considered entertainments as an application of telepresence system and proposed the system which can allow for watching a movie with users who are at distant places. In this system design, the advantage that the user can view and listen at the favorite place just because he or she is in the separate room was also pursued besides the quality improvement of the presence. The method to maintain the coincidence of the gaze direction was developed in addition to guaranteeing user's freedom of the robot arrangement. The position and posture of the head of the user and robots are recognized by the sensor of the room, and the gesture and the gaze direction of the robots in the other rooms are calculated based on the result of the sensor. Interpretation of the motion which conveys atmosphere well was proposed taking into consideration the difference of the flexibility of the user's and the robot's motion. The method to calculate the robot control which is most suitable for the user's robot arrangement based on what the user sees is employed. The mechanism which judges whether the robot behaves exaggeratedly or not, based on user's gesture, the tone of voice, and volume of voice is also installed.

References

[1] K. Otsuka.: MM-Space: Recreating Multiparty Conversation Space by Using Dynamic Displays, NTT Technical Review, Vol. 10, No. 11, pp. 1-5 (2012)

[2] Guizzo. E.: When My Avatar Went to Work, IEEE Spectrum, IEEE Spectrum, Vol. 47, pp. 26-50 (2010)

[3] T. Kashiwabara, H. Osawa, K. Shinozawa, M. Imai.: TEROOS: A Wearable Avatar to Enhance Joint Activities, 30th International Conference on Human Factors in Computing Systems (2012)

[4] K. Hasegawa, Y. Nakauchi.: Effects of Telepresence Robot's Gestures on Turn Talking (in Japanese), Human-Agent-Interaction 2012 (2012)

[5] Y. Nakazato, M. Kanbara and N. Yokoya.: Localization of wearable users using invisible retro-reflective markers and an IR camera, SPIE Electronic Imaging, Vol. 5664, pp. 1234–1242 (2005)