Enhancing Bodily Expression and Communication Capacity of Telexistence Robot with Augmented Reality

Y Inoue[1], M Y Saraiji[2], F Kato[1] and S Tachi[1]
[1] Institute of Gerontology, The University of Tokyo, Tokyo 113-8656, Japan
[2] Graduate School of Media Design, Keio University, Kanagawa 223-8526, Japan

E-mail: y-inoue@tachilab.org

Abstract. This paper focuses on realization of embodied remote communication via telexistence robot with augmented reality (AR). Though a robot equipped with communication functions can realize natural conversation remotely, the robot has to reproduce bodily expressions of the robot user for achieving embodied communication. A humanoid is optimal shape for the purpose, but this kind of robot is currently expensive and difficult to popularize. By contrast, a 3DOF head-moving robot is easier to develop, but the bodily expression capacity is limited. To solve the trade-off problem, we propose an AR-based presentation system visualizing additional body-parts of head-moving robot. The developed system consists of a head-mounted display (HMD) worn by an operator, a 3DOF robot controlled by the operator's head movement, and see-through AR glasses worn by an interlocutor who faces the robot. For visualizing bodily expression, the system generates 3D-CG image of virtual avatar which copies operator's body movements, and projects the image onto both operator's HMD and interlocutor's glasses simultaneously. Consequently, proposed system provides body gesture functions to 3DOF robot, and achieves embodied remote communication.

1. Introduction

The conceptual core of telexistence [1] explains that a robot with multiple devices to substitute human communication function (i.e. camera as eye, microphone as ear, speaker as mouth, and so on) can be used for the surrogate body of an operating user. By using the robot, the user can act in a remote place spontaneously as if he/she were actually there. Also, the user can meet and communicate with people in a remote place directly via the surrogate body. This means that telexistence robot realizes natural face-to-face communication with multiple people in distant places, as if they were all in the same room. Unlike videophone, the robot is an existing object acting like human, therefore the presence of the communication partner is clearly for other meeting members and their interaction may be facilitated. That is, one major advantage of using a robot in telecommunication is increasing the sense of presence of a person inside the robot.

Recently, a number of robot-based telecommunication systems commonly referred to as “telepresence robot" such as Double2 [2] and BeamPro [3] have been produced and popularized. These kinds of robots typically provide both locomotion and communication functions, that is, they have rolling wheels to move across a room and are equipped with a computer screen to be used as the videophone. Furthermore, some other robots possess stereo camera in order to provide immersive 3D experience for the robot user (e.g. TELUBEE [4]). The two images taken from the pair of cameras are presented on the eyes through stereoscopic display device such as head-mounted display (HMD). In this case, eye contact between the robot user and the communication partner can be achieved spontaneously because the camera's direction corresponds to user's gaze, particularly when the robot also has head rotation function. Such a robot-head moves the neck joint to change the camera direction so that the user can look around the place. That is, the robot-head copies user's head movements. This means that the robot transmits the bodily expressions to the remote place. This capability is crucial for the achievement of embodied communication, because human's spontaneous actions during conversation such as nod and glance play an important role in human communication. Through the duplicated movements of the robot, the interlocutor facing the robot can understand the implicit intentions of the user who are not actually present.

It is considered that human's body action can be effectively replicated by human-like shape, therefore a humanoid robot is the optimal to transmit embodied information. However, many of popularized robots for telepresence/telexistence purpose are not perfect humanoids. This is probably because that a whole-body humanoid is currently expensive and therefore difficult to produce and popularize. Also, bodily expression is regarded as less important information than verbal and facial ones. Nevertheless, these communication robots desirably have additional ways to show bodily expressions of the user in order to realize embodied communication with the remote interlocutor. How does the robot acquire sufficient capacity to express embodied information with the minimum effort of the implementation? To equip mechanical add-ons are not reasonable. The most practical solution to achieve the enhancement is to apply virtual-reality (VR) and augmented-reality (AR) technology. Yamen et al. [5] demonstrate that visual projection of user's hand in the remote place can be used to compensate the embodied information in the robot-based telecommunication. To show user's hand image on the surface via a projector equipped in the robot, the robot user and remote interlocutor can see same hand image to joint their visual attention. This means that the robot user can point any object in the remote environment directly by using the virtual hand as long as the object is on the projectable surface (e.g. whiteboard and table). In this case, the physical arms and hands of the robot are not necessary for social activities between them. That is, VR/AR visualization of body can be helpful to transmit embodied information. Here, our proposal system visualizing additional body-parts of robot will be described below.

2. Proposed system

2.1. Overview

Figure 1 shows the overview of proposed system. The system configuration is similar to previous study [5] besides the display of virtual body. The user can see and hear what happens at the remote place through the robot, and the interlocutor can meet the user who controls the robot. Furthermore, this system allows them to be aware of each other's bodily expressions. This system is divided into two subsystems of local side (left) and remote side (right). Both sides are connected to each other via the network.

The subsystem for local side consists of three components below: (L-1) a head-mounted display (Oculus CV2, Oculus VR) worn by the robot user to measure head movements and to present audio/visual information received from the robot; (L-2) a pair of position tracking sensor (Oculus touch, Oculus VR) held by the user's hands to obtain the hand position/attitude; (L-3) a laptop computer placed near the user to control user's equipment, record user's voice and communicate with the robot. The subsystem for the remote side consists of two components below: (R-1) a 3DOF head-moving robot with communication functions (TX-toolkit, JST-ACCEL Embodied Media Project, Figure 2) controlled by user's head movements, speaking the user's voice and sending audio/visual data recorded in the remote environment; (R-2) an optical see-through AR glasses (HoloLens, Microsoft) worn by the interlocutor to show an AR-image superimposed on the robot.

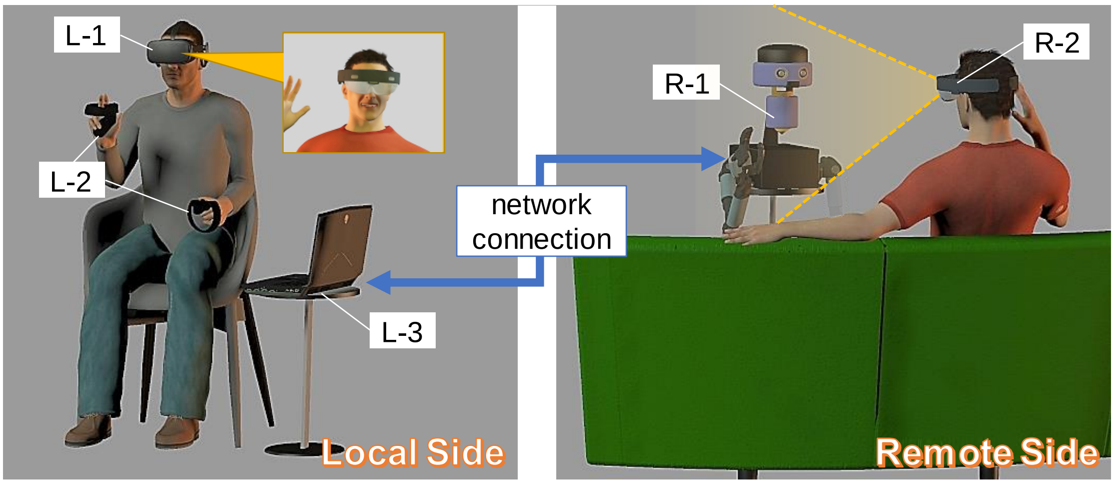

Figure 1. Overview of the proposed system.

Figure 2. "TX-toolkit", developed by JST-ACCEL Embodied Media Project, is a 3DOF robot-head equipped with communication functions for the experimental platform of telexistence study.

Specification:

Size	362mm x 167mm x 232mm
Weight	2.5kg
Devices	Camera x 2, Microphone x 2, Speaker x 1, USB3.0 x 2, Ethernet x 1, WiFi x 1
Motion	Yaw ±90° Pitch ±20° Roll ±20°
ViewAngle	H 100° V 98°
Resolution	960 x 950 pixel/eye @ 60fps

2.2. Virtual Arm: Control

The system generates 3D-CG image of robot-arms which are controlled by user's arm movement, and presents the image on the stereoscopic display devices of the user and the interlocutor simultaneously in different manner (explained later). The virtual arm has 7DOF, the same as human's arm, and moves according to the tracking sensor in order to match the hand position/attitude between the user and the robot. Note that though the sensor input is insufficient to determine the posture of robot-arm uniquely, the unconstraint 1DOF of arm movement can be adjusted by using extra input (e.g. thumbstick).

2.3. Virtual Arm: Display

Figure 3 shows examples of subjective appearance of the virtual arms presented on HMD for the user (left) and AR-glasses for the interlocutor (right). For the user, the virtual arms are superimposed on the video image streaming from the eye-camera as the first-person-perspective view. Therefore the user feels to be in the place as the robot because the virtual arms are seen as if they were user's own. For the interlocutor, the virtual arms are presented around the actual robot as if these arms were attached to the robot. Thus the interlocutor feels that he/she faces to an upper-body humanoid, despite facing to a robot-head. Consequently, even though the robot does not have actual arms or hands, the user can show his/her arm gestures to the interlocutor.

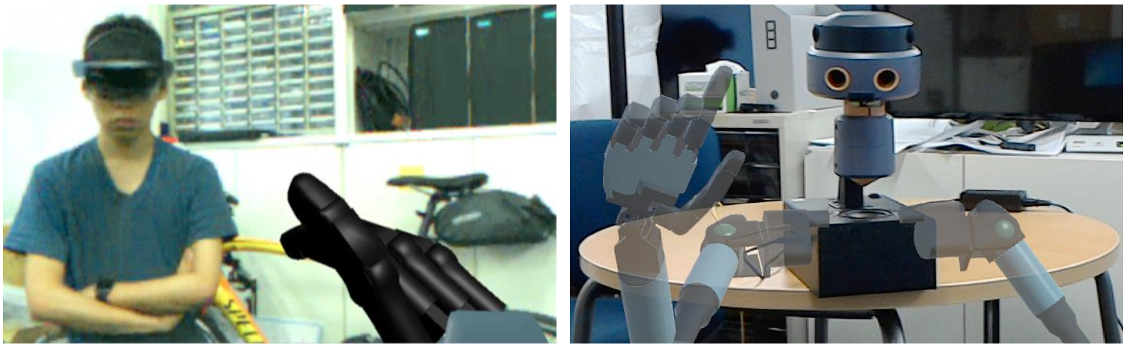

Figure 3. Examples of AR images of the user's view (left) and the interlocutor's view (right).

3. Conclusion

In this paper, we propose an AR-based enhancement system of the communication robot which has limited capacity of bodily expression. This system realizes embodied remote communication. Owing to the system's aids, persons who are actually in different places can share the same surroundings, and they have close communication through their bodily expressions. In contrast to projected hand image presented in previous study [5], the proposed system presents AR hand object on the 3D space, therefore, the robot user can point any object (not limited to be on the surface) by using the virtual hand. Because the proposed system works as an AR add-on application, this is available to any other communication robot if the robot has wireless network devices (e.g. WiFi and Bluetooth) to transmit user's body posture data to the AR glasses worn by the interlocutor. Also, by detecting spatial contacts between user's virtual hand and interlocutor's actual hand, it is possible to realize tactile communication between them such as high-five. To conclude, the combination of robotics and VR/AR has great potentials to realize new communication tools.

Acknowledgments

This research is supported by ACCEL Program from Japan Science and Technology Agency, JST.

Reference

[1] Tachi S 2015 *Telexistence 2nd Edition* (Singapore: World Scientific)

[2] Double Robotics, Inc. (n.d.) Double 2 – Features, Retrieved July 31, 2018 from https://www.doublerobotics.com/double2.html

[3] Suitable Technologies, Inc. (n.d.) Meet BeamPro®, Retrieved July 31, 2018 from https://www.suitabletech.com/products/beam-pro

[4] Laval Virtual. (February, 1, 2013) ReVolution 2013 – Ubiquitous Telexistence, Retrieved July 31, 2018 from https://www.youtube.com/watch?v=Q5zigZDKZo8

[5] Saraiji M Y, Fernando C L, Minamizawa K and Susumu Tachi 2015 Development of Mutual Telexistence System using Virtual Projection of Operator's Egocentric Body Images, *Proc. 25th Int. Conf. on Artificial Reality and Telexistence (Kyoto)* pp.125-132