第 22 回日本バーチャルリアリティ学会大会論文集（2017 年 9 月）

対話のための視線と手の動きを自然に伝える
非没入型遠隔ロボット操作環境

A Non-immersive Robot Teleoperation Environment for Natural Communications of Gaze and Hand Motions for Dialogs

Haoyan LI[1), Samratul FUADY[1), Hironori MITAKE[1) and Shoichi HASEGAWA[1)

1) 東京工業大学 未来産業技術研究所 （〒 226-8503 神奈川県横浜市緑区長津田町 4259 R2-624 長谷川晶一研究室,
lihaoyan@haselab.net）

Abstract: Recently, robots have been widely used in several situations for entertainment purpose. Robots used for dialogs offer an effective way to connect remote and local side. To operate a robot during dialogs, an appropriate teleoperation environment is needed. Compared with an immersive environment, the non-immersive environment shows its possibility of allowing operators to handle some multiple tasks simultaneously. However, those non-immersive systems always use an unintuitive method to control robots which could increase operators' loads and distract their focus from the main task. In this paper, we proposed a non-immersive robot teleoperation environment for natural communications of gazes and hand motions for dialogs. In the remote side, we implemented a natural behavior control interface to reduce operators' loads. In addition, the semantic transformation based on gaze and hand motions realized natural movements of a robot on the local side.

Keywords： teleoperation, communication robot, entertainment robot

1. Introduction

Nowadays, more and more researchers and experts have committed to work in the research of interactive robots. The interactive robot is also starting to be widely used in various fields and playing an important role in people's daily life. In particular, interactive robots for entertainment, family communication, education, and healthcare, are rapidly spreading. For example, Sony's AIBO[1], which is dog-like, could autonomously recognize human's voice, sense user's touch, and react with users by some pre-programmed motions. Because of preprogramming, the interaction that can be performed is very limited and monotonous. At the same time, lacking real-time nonverbal communication in this type of robot is also believed to be critical in communication[2]. Another approach to realize an interactive robot, such as Quasi[3], is to perform a movement based on a remote operator's commands. However, the main problem of this type of robot is the user needs to input the response or emotion manually, which results in an unnatural interaction.

Generally, to realize a natural and interesting interaction by an interactive teleoperation robot, especially in a dialogue communication, nonverbal information, such as turn-taking, eye-gaze direction, and gesture, is also essential to express operator's intention and increase the robot's perceived lifelikeness[4]. Therefore, we intend to develop a friendly user environment for the interactive robot which can give non-experts the ability to operate or create robot's motions easily.

In this paper, we propose a non-immersive robot teleoperation environment for dialogs which is using gaze and hand motions to realize a natural and interactive communication. The system is aimed to reduce an inexperienced operator's loads of learning how to operate the robot by his/her eye-gaze or hand motions. In addition, the operator also should be able to handle multiple tasks simultaneously, such as taking notes or reading materials. For the local side, the remote user's existence should be perceived by participants. Also, the avatar robot could perform a natural and interactive eye-gaze and hands movement.

The remainder of this paper is organized as follows. In section 2, we will briefly introduce our previous work about an asymmetric telepresence system using a stuffed-toy robot. Section 3 is the system overview of non-immersive robot teleoperation environment herein. Then, a system experiment will be introduced and discussed in section 4. Finally, we conclude this paper with a summary of our research and discuss some future work in section 6.

2. Previous Work

In our previous work [5], we proposed an asymmetric telepresence system using a stuffed-toy robot as the representative of the remote user, as shown in Figure 1. The telepresence system showed that people on the both sides are enabled to have a natural interaction, even without immersing the user into a virtual environment.

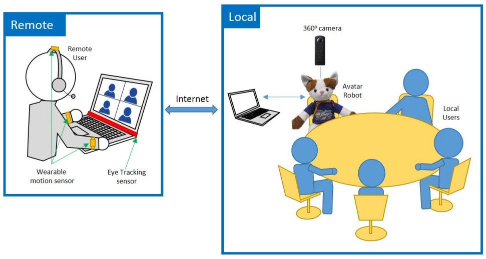

Figure 1: System Overview

We used a stuffed-toy robot[6] as the avatar robot which is able to perform some envelop feedbacks with fast response. The stuffed-toy robot is built by cotton-based soft materials for the moving part, as shown in Figure 2. Each moving part is controlled by three strings. Since the end point position of robot's moving part is mapped and interpolated to the length of control strings, the robot can perform a smooth motion.

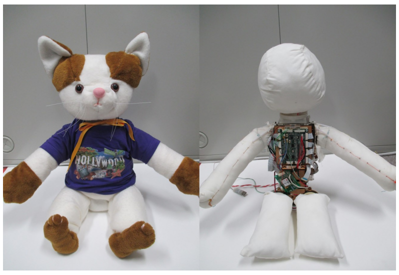

Figure 2: Stuffed toy robot

On the remote side, a user needs to take three wearable devices. Two watch-like sensors are used for capturing both hands' movement, and one headphone-embedded sensor is for the head movement. The headphone is also used for voice communication during the teleconference. The sensor can capture data from an accelerometer, gyroscope, magnetometer, and pressure sensor. Here, we only used gyroscope data to get the angular velocity of the remote user's movement and transformed it to the end point position. Moreover, we placed an eye-gaze tracker in the remote user's PC to capture his/her eye-gaze movement. Participants' face was shown on the remote user's display monitor. The looking point captured by the eye-gaze tracker would be transformed to robot's head rotation, so the robot could realize the eye-contact with local participants.

On the local side, we prepared an avatar robot, omni-directional camera, microphone, and speaker. Similar to a video chat, camera, microphone, and speaker are used for video and audio communication. The avatar robot would be operated by the remote user.

According to the questionnaire results of the experiment, the proposed system showed that it could enhance teleconference experience in some aspects. The turn-taking and beat gesture perceived better than a general video conference, whereas other aspects didn't show significant difference with a video conference.

In our previous work, we did not focus on the remote user's experience. The remote user should be well-trained in operating robot. Considering the practical use, the implementation of the remote side must be optimized. Moreover, we found that there were several practical problems negatively influenced our system's performance, such as network latency and video quality. Therefore, how to solve those problems and implement a better user interface has been the next steps for us.

3. Non-immersive Robot Teleoperation Environment

Since our previous system is designed for a teleconference, the proposed system needs to be improved and adapted to the new purpose. To explore the possibility of a general communication configuration, this paper presents (1) improvement of software architecture, (2) modifications of user interface, and (3) motion transformation method, based on previous work. On the local side, the avatar robot is placed on a stage rather than the seat in the conference room. The participants will be front of the robot and interact with it.

3.1 Improvement of Software Architecture

We improve the software architecture of system from previous design to adapt to a dialog based application, such as an entertainment game. The software architecture of this system is shown in Figure 3.1.

On the remote side, the reading of motion sensors and eye-tracker are collected and reformatted by the application running on the remote PC. The angular velocity of remote user's movement got from motion sensors will be sent to the local side via Internet. The eye-gaze tracker

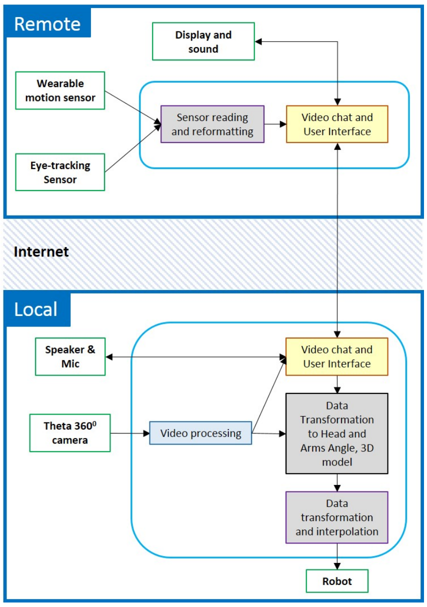
Remote
Display and sound
Wearable motion sensor
Sensor reading and reformatting
Video chat and User Interface
Eye-tracking Sensor
Internet
Local
Speaker & Mic
Video chat and User Interface
Theta 360⁰ camera
Video processing
Data Transformation to Head and Arms Angle, 3D model
Data transformation and interpolation
Robot

Figure 3: Software Architecture of System

gives the coordinate of remote user's looking point on the display. It is also sent to the local side via Internet.

On the local side, the transmitted data will be handled and processed by the local program. The sensor data, which consists of the angular velocity of the head and hand movement and the position of eye-gaze, will be transformed based on the robot structure and reproduced in a real-time 3D model.

3.2 Modification of User Interface

On both sides, we develop a server-based video chat web application for displaying and transmitting audio, video, and sensor data, depends on the WebRTC protocols. WebRTC is designed to enable rich, high-quality RTC(Realt-Time Communications) applications to be developed for the browser. Considering that video, audio, and data communications are essential to our system, we believe that WebRTC should be an appropriate choice for the purpose. The data flow of video processing is shown in Figure4.

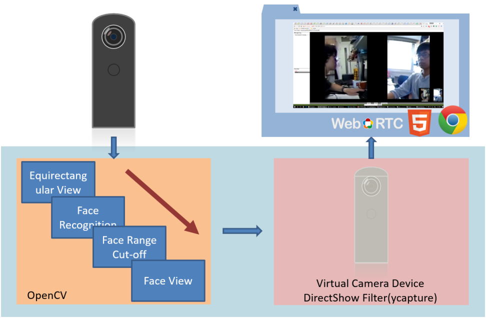

Figure 4: Data Flow of Video Processing

On the local side, the video captured by the omnidirectional camera will be processed by an OpenCV program. The OpenCV program can automatically detect or manually select the faces showing in the local side view. Then, the program cuts these faces out from the view and generates a face view for the DirectShow filter. The DirectShow source filter, a modified version based on ycapture[7], gets the output from the OpenCV program and send it by web application as a virtual camera device.

3.3 Motion Transformation Method

For each moving part of the robot, we transform end point velocity v_{ep} into string length velocity v_s which is needed to control the robot. Jacobian Matrix J, which is the first derivative of string length with respect to the end point, is used to calculate v_s.

$$v_s = kJv_{ep} \tag{1}$$

where $v_s = \begin{bmatrix} \dot{L_1} \\ \dot{L_2} \\ \dot{L_3} \end{bmatrix}$, $v_{ep} = \begin{bmatrix} \dot{\theta} \\ \dot{\phi} \end{bmatrix}$ and k is constant. L_1, L_2, and L_3 describe the string length while θ and ϕ are the endpoint angle of each moving part.

The jacobian matrix is calculated numerically as,

$$J \approx \begin{bmatrix} \frac{L_1(\theta+\Delta\theta,\phi)-L_1(\theta-\Delta\theta,\phi)}{2\Delta\theta} & \frac{L_1(\theta,\phi+\Delta\phi)-L_1(\theta,\phi-\Delta\phi)}{2\Delta\phi} \\ \frac{L_2(\theta+\Delta\theta,\phi)-L_1(\theta-\Delta\theta,\phi)}{2\Delta\theta} & \frac{L_2(\theta,\phi+\Delta\phi)-L_1(\theta,\phi-\Delta\phi)}{2\Delta\phi} \\ \frac{L_3(\theta+\Delta\theta,\phi)-L_1(\theta-\Delta\theta,\phi)}{2\Delta\theta} & \frac{L_3(\theta,\phi+\Delta\phi)-L_1(\theta,\phi-\Delta\phi)}{2\Delta\phi} \end{bmatrix} \tag{2}$$

where $L_i(\theta, \phi)$ is the length of string i required to produce end point position at angle θ and ϕ. The calculated string length velocity is then sent to the robot.

Because we use the velocity data, the position error will be accumulated. To avoid this kind of problem, the robot always returns to the neutral position when no angular velocity is detected from the remote user. This approach is suitable to mimic beat gesture of human when talking.

The eye gaze position data will be transformed into the turning angle of the robot's head because the current robot does not have moving eyes. The transformation will depend on the type of video that is shown to the remote user. When using the equirectangular video, the transformation is calculated by proportionally matching the gaze position with robot's turning angle. But when using the faces image, the transformation is calculated based on face position data which is obtained from the video processing part.

4. Experiment and Discussion

We setup a word quiz game which is designed to test the performance of proposed system. Figure 5 is shown the actual configuration of the experiment.

For the remote side, the operator will be tested on whether he or she can handle multiple tasks simulta-

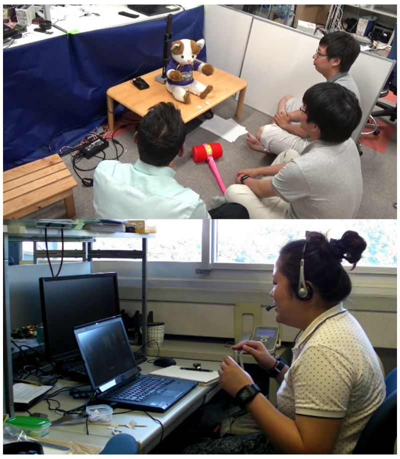

Figure 5: Participants interacting with avatar robot during game

neously and if the system is easy to use. He/She was asked to take all of the sensors and choose a word randomly from a word list. Then, he/she could start to describe the chosen word by any mean, except for mentioning some particular words. When local participants raise their hand, the remote operator needs to choose one person to answer or ask. Calling participant's name is not allowed, which means the remote user needs to look at the person whom he would like to select. As one responsibility of the remote operator, he/she is also asked to take a note of game record. A natural communication by verbal and non-verbal cues, such as dialogs, feedback gestures, and turn-taking are expected in this part.

For the local side, participants are asked to guess the word which the remote operator is describing. Before giving their answer or asking for details, they need to raise their hand first, in order to let the remote user choose the person who should answer or speak.

We observed that participants in the local side could understand well during the game and catch up with the turn-taking. The remote operator also handled multi-tasks simultaneously, like note writing or card taking. Because we only explained the rules to the remote user and did not mention how to control the robot before the game, it seems that the remote operator can easily understand how to use their natural behaviors to control the robot, even without a training. Additionally, Compared with the previous system, the robot's looking at behavior is more stable and accurate.

5. Conclusion and Future Work

We developed a non-immersive robot teleoperation environment for dialogs that enables a natural gaze and hand motions interaction. We use wearable motion sensors and eye-gaze tracker to obtain the essential feedback behaviors of the remote user: eye-gazing, turn-taking, and beat gesture. The stuffed-toy robot is used to be the representative of a remote user. According to the observation of experiment, the proposed system could reduce the operator's loads of learning how to operate the robot and handle multiple tasks simultaneously. In addition, the proposed system also could make users achieve a natural and enjoyable interaction.

In the future, we would like to improve the richness of the robot movement feature so it can show better life-like movement. Further investigation about the meaning of user movement also needs to be conducted so that the avatar robot can produce more accurate and meaningful gesture. Furthermore, the influences of applying this system in some specific situation, such as entertainment show, child education, and healthcare, could be investigated.

Acknowledgement This research is (partially) supported by the Center of Innovation Program from Japan Science and Technology Agency, JST.

References

[1] SONY Corporation: http://www.sony-aibo.com/, Sony Aibo Tribute Site，2017.

[2] Justine Cassell, Kristinn R. Thorisson: The Power of a Nod and a Glance: Envelope vs Emotional Feedback in Animated Conversational Agents, Applied Artificial Intelligence, Vol.13, No.4, pp.519-538, 1999.

[3] Haskell Sabrina, Hosmer Andrew, Leu Eugenia: An Extensible Platform for Interactive, Entertaining Social Experiences with an Animatronic Character, Proc. of the 2005 ACM SIGCHI Int. Conf. on Advances in computer entertainment technology, pp.141-148, 2005.

[4] Nehaniv Chrystopher, Dautenhahn Kerstin: Imitation and Social Learning in Robots, Humans, and Animals: Behavioural, Social and Communicative Dimensions, Cambridge University Press, 2007.

[5] Samratul Fuady, Masato Orishige, Haoyan Li, Horonori Mitake, Shoichi Hasegawa: Natural Interaction in Asymmetric Teleconference Using Stuffed-toy Avatar Robot, Proc. of the 26th Int. Conf. on Artificial Reality and Telexistence and the 21st Eurographics Symposium on Virtual Environments, pp.93-98, 2016.

[6] 高瀬 裕, 山下洋平, 石川 達也, 椎名 美奈, 三武 裕玄, 長谷川 晶一：多様な身体動作が可能な芯まで柔らかいぬいぐるみロボット, 日本バーチャルリアリティ学会論文誌, Vol.18, No.3, pp.327-336, 2013.

[7] 谷沢 智史：http://yzwlab.net/ycapture/，ycapture(わいきゃぷちゃ) version 0.1.1，2017.