A Robotic Arm-based Telepresence for Mixed-Reality Telecollaboration System

Le Luo* [1] Dongdong Weng ✉ † [1] Jie Hao [1] Ziqi Tu [1] Bin Liang [2] Haiyan Jiang [1]

[1] MRAD of Beijing Institute of Technology; [2] China Software Testing Center

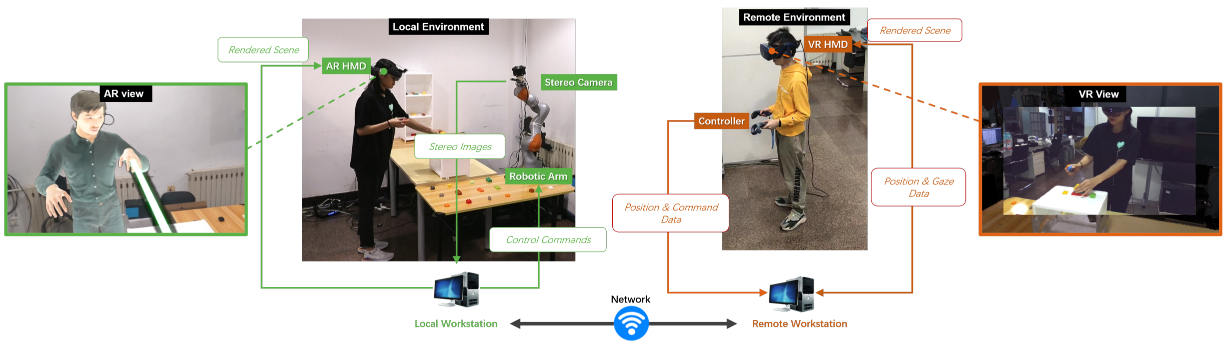

Figure 1: Remote collaboration between a local user wearing an AR HMD and a remote user wearing a VR HMD.

Abstract

In mixed-reality telecollaboration, it is often difficult for remote users to actively and naturally control the viewpoint. We propose a viewpoint controllable telepresence with a robotic arm that carries a stereo camera in the local environment so that the remote user can control the robotic arm by moving his or her head to actively and flexibly observe the local environment. Additionally, we built a mixed-reality telecollaboration prototype that adds avatar and nonverbal cues for the remote user to the local user side, enabling the remote user to better guide the local user. In this paper, we discuss our prototype, make design recommendations, and discuss configuration and implementation.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality;

1 Introduction

Mixed-reality telecollaboration has a high sense of immersion and coexistence, which can provide a nearly face-to-face interaction experience in remote collaboration. In mixed-reality telecollaboration, a remote user wearing a virtual reality (VR) head-mounted display (HMD) observes a remote representation of the local space and collaborates with a local user wearing an AR HMD [5]. Video capture devices are set up in the local environment to transmit local scene information to remote users in real time. In traditional mixed-reality telecollaboration, video capture devices are usually placed in a fixed location in the local environment to provide a standalone view [4] or placed on the local user's head-mounted display to provide a first-person view of the local user [2], but the remote user may not be able to actively and freely switch the view. In studies of telepresence, researchers have equipped video capture devices on wheeled robots [1] or drones [3], which can be operated by remote users to move the viewpoint within the local space. However, these systems either lack vertical control or operate with relatively low accuracy and high latency, preventing remote users from controlling the viewpoint precisely and naturally.

In this paper, we propose a viewpoint controllable telepresence that allows remote users to obtain a precisely controllable view in the local space for more active and natural observation of the local environment (Figure 1). At the same time, we demonstrate a prototype system that adds avatar and nonverbal cues for the remote user to the local user side, enabling the remote user to better guide the local user.

2 Design

We propose a viewpoint controllable mixed-reality remote interaction system that provides an interactive environment that allows remote users to freely control the viewpoint, observe the local physical space, and guide local users to collaborate on tasks. At the same time, the local user is able to sense the presence of the remote user and is able to receive voice and nonverbal cues from the remote user. The basic structure is shown in Figure 1. Remote users wear VR HMDs to view local scenes, which show a combination of stereoscopic video and 3D models of the local scene. When the arm is within the working range, the image transmitted by the stereo camera in real time is fixed in the user's central visual field, and the peripheral vision is complemented by a 3D scene reconstructed from the local space. When the arm is outside the working range, the local space environment perception is fully provided by the reconstructed 3D scene. A tracker set on the HMDs captures the remote user's head pose data, which are used to drive a robotic arm in the local space, enabling the remote user to actively change the viewing perspective by moving their head to control the arm. Remote users can also use a pair of controllers to communicate nonverbal cues.

We set up a robotic arm in a local physical space with a stereo camera to obtain stereo video of local scene information in real time. The robotic arm is controlled by the head pose data of the remote user. The data acquired by the stereo camera are transmitted to the remote user in real time over the network. The local user wears an augmented reality (AR) HMD, which will allow them to view the real local environment and be able to see the overlaid augmented information, including the avatar of the remote user

*E-mail: l_luo@bit.edu.cn
†Corresponding author,E-mail: crgj@bit.edu.cn

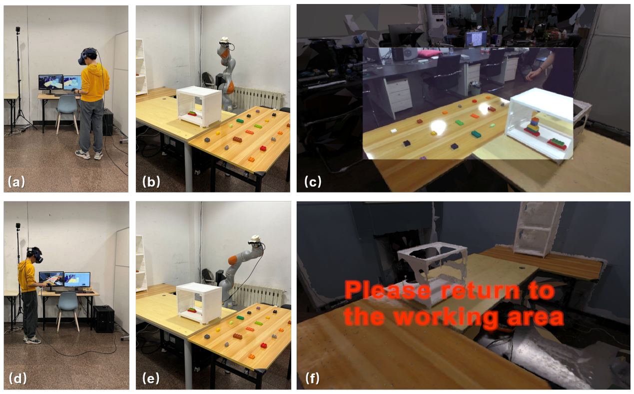

Figure 2: The remote user controls the robot arm to view the local space. From left to right: user position, robotic arm position and user view. From top to bottom: user is inside the working area, user is outside the working area.

and the nonverbal cues conveyed by the remote user. We align the remote space with the local space, display the remote user's avatar superimposed on the location of the robotic arm, and animate the avatar using the remote user's head, hand, and gaze data.

3 System Overview

3.1 Implementation

In the local space, a Kuka LBR iiwa 14 R820 robotic arm is placed in the scene. The robotic arm is equipped with a ZED mini stereo camera. The local user wears a HoloLens2, and the remote user's avatar is rendered using Unreal Engine 4.26. All devices in the local space are connected to one device (Intel Core i9-7900X, NVIDIA GeForce RTX3080, 32 GB RAM and 1 TB SSD).

In the remote space, the user wears an HTC Vive Pro Eye to observe images of the local environment and holds a pair of matching controllers, with the user's head and controller positions tracked in a pair of universal world coordinate systems established by a Lighthouse transmitter. We used the camera to take pictures of various angles of the local space in advance and used Reality Capture to generate 3D models that were rendered on the remote user side. We used Unity3D 2020f3 to develop our system, with the entire remote side running on a PC (Intel Core i7-6850K, NVIDIA GeForce GTX1080Ti, 32 GB RAM and 1 TB SSD).

3.2 Interaction interface for remote users

The remote user wears a VR HMD and is placed in a virtual space. The stereoscopic video is displayed in the user's central field of view through a plane that is fixed in front of the user, while the edge field of view is filled by a 3D reconstruction of the local space (Figure 2(c)). In addition, the remote user uses a pair of handheld controllers, which are displayed synchronously in the HMD, and the remote user can use the handheld controllers to create a virtual ray that conveys pointer information to the local user.

Remote users can control the viewpoint by moving their heads to observe the local space. When the remote user moves beyond the working range of the local robotic arm, the real-time updated stereo video is hidden; the user can only observe the local environment through the reconstructed 3D space but can continue to roam in it. At the same time, a prompt message will appear in the remote user's field of view, indicating that the remote user has moved beyond the working range of the robotic arm and guiding the user to return to the working area of the robotic arm (Figure 2(d)-(f)). When the remote user returns to the working range of the robotic arm, the display of the stereo video turns on and continues to provide the remote user with a real-time stereo view of the local space (Figure 2(a)-(c)).

3.3 Interaction interface for local users

We selected a high-realism full-body avatar as the remote user's avatar to be rendered in the local user's AR HMD. We calibrated the coordinate system of the remote user space with the local space to display the remote user's avatar superimposed on the robotic arm position and follow the remote user's motion. We captured the positional information of the remote user's head and hand nodes through the VR HMD and a pair of handheld controllers and drove the avatar's motion through inverse kinematics (IK).

In addition to the remote user's avatar, we provided visual communication cues for the local user so that the remote user could help the local user perform the task more efficiently and accurately. We created a vector-direction projected ray from the location of the controller held by the remote user in the local space, determined the direction of the ray projection based on the controller orientation, and then rendered it in the local AR HMD, where the virtual ray emanated from the hand of the remote user avatar in the local user's field of view. The hiding and display of the ray was controlled by the remote user.

4 Conclusion and Future Work

In this paper, we propose a robotic arm-based movable telepresence system that allows remote users to actively control the view of the local space by moving their heads to observe the local environment more naturally. In future work, we will try to combine multiple view-sharing schemes and conduct research for more application scenarios while taking full advantage of the robotic arm to allow remote users to physically influence the local space and refine the design of a highly realistic avatar to improve the interaction experience between remote and local users.

Acknowledgments

This work was supported by the Beijing Municipal Science & Technology Commission and Administrative Commission of Zhongguancun Science Park under Grant Z221100007722002 and the National Natural Science Foundation of China (No.62072036) and the 111 Project (B18005).

References

[1] S. Kratz, J. Vaughan, R. Mizutani, and D. Kimber. Evaluating stereoscopic video with head tracking for immersive teleoperation of mobile telepresence robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, HRI'15 Extended Abstracts, p. 43–44. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2701973.2701982

[2] G. Lee, H. Kang, J. Lee, and J. Han. A user study on view-sharing techniques for one-to-many mixed reality collaborations. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 343–352, 2020. doi: 10.1109/VR46266.2020.00054

[3] K. Matsumoto, E. Langbehn, T. Narumi, and F. Steinicke. Detection thresholds for vertical gains in vr and drone-based telepresence systems. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 101–107, 2020. doi: 10.1109/VR46266.2020.00028

[4] T. Rhee, S. Thompson, D. Medeiros, R. dos Anjos, and A. Chalmers. Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1923–1933, 2020. doi: 10.1109/TVCG.2020.2973065

[5] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock. Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, OzCHI '18, p. 406–410. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3292147.3292200