Augmented Reality with Multi-view Merging for Robot Teleoperation

Bidan Huang*
bidanhuang@tencent.com
Tencent Robotics X Lab
Shenzhen, China

Nicholas Gerard Timmons*
ngt26@cam.ac.uk
University of Cambridge
Cambridge, UK

Qiang Li
kaiserli@tencent.com
Tencent Robotics X Lab
Shenzhen, China

ABSTRACT

This paper proposes a user-friendly teleoperation interface for manipulation. We provide the user with a view of the scene augmented with depth information from local cameras to provide visibility in occluded areas during manipulation tasks. This gives an improved sense of the 3D environment which results in better task performance. Further, we monitor the pose of the robot's end effector in real-time so that we are able to superimpose a virtual representation into the scene when parts are occluded. The integration of these features enables the user to perform difficult manipulation tasks when the action environment is normally occluded in the main camera view.

We performed preliminary studies with this new setup and users provided positive feedback regarding the proposed augmented reality (AR) system.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Empirical studies in interaction design**; • **Hardware** → **Emerging technologies**.

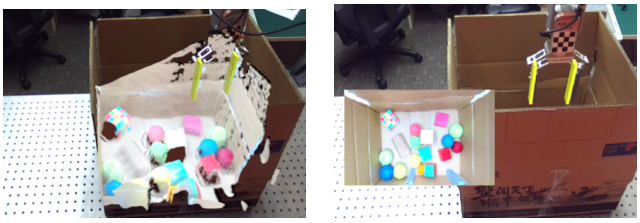
KEYWORDS

augmented reality, teleoperation, multi-view vision, object manipulation

ACM Reference Format:
Bidan Huang, Nicholas Gerard Timmons, and Qiang Li. 2020. Augmented Reality with Multi-view Merging for Robot Teleoperation. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20 Companion), March 23–26, 2020, Cambridge, United Kingdom.* ACM, Cambridge, UK, 3 pages. https://doi.org/10.1145/3371382.3378336

1 INTRODUCTION

Teleoperation enables humans to control robots remotely and access areas that would be difficult or dangerous for workers. This technology has wide applications in industrial and medical tasks such as repairs, inspection and surgeries[3, 8, 10]. Typical teleoperation systems includes two parts, a master and the slave. While the master provides an interface to be operated by the human, the slave is a robot that follows the human commands. Sensors such as cameras are deployed on the slave to monitor the surrounding environment. Information gathered from these sensors is transmitted to the master and allows the user to perceive the operational environment. The technology that makes the users feel as if they were present in the remote scene is referred to "telepresence"[9].



(a) PCDP (b) PiP

Figure 1: Comparison of the user view for PCDP and PiP interfaces of the same operation scene.

Good performance in a teleoperation task largely relies on having high-quality telepresence, especially for tasks in complex environments involving unknown obstacles. In these environments views of the operation scene can be easily blocked by the environment or the robot itself. To tackle this problem it is common for engineers to mount multiple cameras to the slave[3, 5]. Choosing the best viewpoint to display the scene remains an open question.

In this paper, we focus on the visual telepresence for manipulation tasks. For manipulation tasks, physical interaction between the robot and the environment is essential, therefore, to achieve a successful manipulation it is important to display the local view of the scenes to the user in a meaningful way. To this end, we employed the latest immersive Virtual Reality (VR) technology to display the operation scenes. We propose a novel Augmented Reality (AR) approach for merging camera information to facilitate manipulation. In the proposed system, a global and a local camera are mounted to the robot to provide views from different perspectives which then merged into a single view which is easier for the user to understand and work with.

2 AUGMENTED MULTI-VIEW TELEOPERATION INTERFACE

In this section, we detail the system design and configuration for teleoperation. For the master, we used the HTC Vive headset and controller to track head and hand motion. Head movement was mapped to the view in the headset and the hand movement was mapped to the robot end effector. The robot gripper operation was controlled by button press.

*Both authors contributed equally to this research.

HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3378336

For the slave, the hardware components include a UR5 6 d.o.f robot arm, a RG2 two finger robot gripper, a RealSense depth camera[4] and a ZED stereo system[6]. The ZED stereo camera was mounted to a fixed location related to the UR5 and provides a global view of the scenes. An RG2 gripper was mounted to the robot end effector, while a RealSense depth camera was attached to its side as a local view camera.

2.1 Global View Visualization

Our VR presentation was setup in the Unity game engine. With this we are able to configure our 2D, VR and augmented reality display modes and present them to the user through a HTC Vive display.

The visualisation of the scene contains five main components which all need to be integrated with each other: (1) VR rendering of the main view camera, (2) Augmented Reality View of the Robot with Occlusion, (3) Augmented Reality View of the Depth Information from the local view camera, (4) Presenting 2D camera views to a VR Display, (5) Allowing the user to look around in AR 3D.

2.2 Local View Visualization

The local view camera mounted at the end effector is provided by a Intel RealSense depth camera. The RealSense provides a single RGB image along with the depth from its infrared cameras.

The goal of a local camera is to increase the users physical understanding of the scene near to the operation being performed. Past research has often used PiP (Picture-In-Picture) to present separate views to the user[2, 7].

We believe that the information should be presented as a single output to the user to maximize the ability to understand the scene with the relative positions of all objects. To test this hypothesis we provide three ways to visualise the local camera within the global cameras viewpoint:

PiP: For comparison to other approaches we provide a PiP mode where the view from the local camera is presented in the corner of the view for the user. This picture is able to be toggled on and off.

World Space PiP: In this mode the view from the local camera is presented on a 2D plane in front of the local cameras position. This gives the user information regarding the relative angle and pose of the camera so they can better reason about where the camera is looking and what can be seen. The rendering of the view from the camera is never occluded even when the end effector is occluded. A limitation of this approach is when the view from the local camera is perpendicular to the global camera then the plane displaying the PiP is also perpendicular, which means the information is not visible.

Point Cloud Depth Projection (PCDP): This mode is similar to 'World Space PiP' but also provides depth information. This gives a 3D object generated from camera depth values which can be rendered into the scene to show hidden geometry1. This object is only visible where the distance from the global camera is greater than that of the depth visible to the global camera. Effectively, this draws geometry that is being occluded when that geometry is visible to the local camera. This gives the user a sort of "x-ray vision" to perform tasks that would be more difficult with occluded views. It also suffers from the same problem as "World Space PiP" with perpendicular geometry, to resolve this we provide a mode which allows the user to disable the global camera view and use head movement to inspect the 3D geometry.

2.3 Telemanipulation

The teleoperation system is coordinated using the *Robot Operating System* (ROS)[1]. The network uses two ROS nodes. One node is located in Unity on a *Windows 10* machine to collect user input and another node is running ROS (Kinetic) on an *Ubuntu 16.04* machine to control the UR5 using URScript. The communication frequency is 125Hz and implemented by the ROS Unity bridge [1]. The UR5 is working in the joint servoing mode.

3 EXPERIMENTS AND RESULTS

To evaluate the performance of the proposed AR system we carried out a user study comparing it with the conventional PiP interface. In both interfaces the RGBD images from the global camera were displayed on the headset to provide the user an immersive view of the operation scenes. The view from the local camera were displayed in two forms:

Interface 1: A PiP style image in the corner of the view

Interface 2: A Point Cloud Depth Projection merged based on occlusion as described above.

Using both interfaces participants were asked to repeatedly perform a manipulation task with two stages:

(1) Stage 1: Pick up target objects on the table and place them into a large box which occludes the global camera.
(2) Stage 2: Pick up target objects from the large box which occludes the global camera and put them back on the table.

For the study, we recruited 12 participants from the local community (9 male, 3 female, ages 25-60). Before the experiment each participant was shown how to operate the robot and given 5 minutes to practise without the headset. After familiarisation with the controls the participants put on the headset and then were instructed to complete the task with the two different interfaces.

After the experiment, participants were asked to describe their experience with the system and point out their preferences between PiP and PCDP. 8 out of the 12 participants admitted that the PCDP was preferred as it provides a better sense of depth compared to the PiP interface. They also felt that PCM provided more intuitive manipulation and view of the scene, while PiP require users to jump between local and global views which created extra cognitive burden. The main reason given by those who preferred PiP was the higher resolution given by the 2D view than the RGBD view.

4 CONCLUSION AND DISCUSSION

In this report, we presented our latest AR development for teleoperation. In this system, we merged the local and global view of the scene to give users an augmented field of view in a VR environment. With the PCM interface the users can see through obstacles that would otherwise hinder tasks. This allows users to perform manipulation even when the main view of the scene is occluded. Compared to the conventional PiP interface the PCDP interface provides a more intuitive perspective in our users experience. In our future work, we will focus on improving the resolution of PCDP so that the video quality of the occluded view matches that of the standard PiP.

REFERENCES

[1] [n.d.]. Ros Unity Bridge. https://github.com/siemens/ros-sharp. Accessed: 2019-12-01.

[2] Gloria L Calhoun, Mark H Draper, Jeremy T Nelson, and Heath A Ruff. 2006. Advanced display concepts for UAV sensor operations: Landmark cues and picture-in-picture. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 50. SAGE Publications Sage CA: Los Angeles, CA, 121–125.

[3] Soichiro Iwataki, Hiromitsu Fujii, Alessandro Moro, Atsushi Yamashita, Hajime Asama, and Hiroshi Yoshinada. 2015. Visualization of the surrounding environment and operational part in a 3DCG model for the teleoperation of construction machines. In *2015 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 81–87.

[4] Leonid Keselman, John Iselin Woodfill, Anders Grunnet-Jepsen, and Achintya Bhowmik. 2017. Intel realsense stereoscopic depth cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1–10.

[5] Fumio Okura, Yuko Ueda, Tomokazu Sato, and Naokazu Yokoya. 2013. Teleoperation of mobile robots by generating augmented free-viewpoint images. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 665–671.

[6] Luis Enrique Ortiz, Elizabeth V Cabrera, and Luiz M Gonçalves. 2018. Depth data error modeling of the ZED 3D vision sensor from stereolabs. *ELCVIA: electronic letters on computer vision and image analysis* 17, 1 (2018), 0001–15.

[7] A Pitman, Curtis M Humphrey, and Julie A Adams. 2007. A picture-in-picture interface for a multiple robot system.

[8] Long Qian, Anton Deguet, Zerui Wang, Yun-Hui Liu, and Peter Kazanzides. 2019. Augmented Reality Assisted Instrument Insertion and Tool Manipulation for the First Assistant in Robotic Surgery. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 5173–5179.

[9] Thomas B Sheridan. 1992. Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments* 1, 1 (1992), 120–126.

[10] AWW Yew, SK Ong, and AYC Nee. 2017. Immersive augmented reality environment for the teleoperation of maintenance robots. *Procedia CIRP* 61 (2017), 305–310.