Published verion at https://doi.org/10.1145/3562939.3565687

Investigation of User Performance in Virtual Reality-based Annotation-assisted Remote Robot Control

Thanh Long Vu
Dac Dang Khoa Nguyen
Sheila Sutjipto
Thanh.Vu@uts.edu.au
Khoa.Nguyen@uts.edu.au
Sheila.Sutjipto@uts.edu.au
University of Technology Sydney: Robotics Institute (UTS:RI)
Sydney, New South Wales, Australia

Dinh Tung Le
Gavin Paul
DinhTung.Le@uts.edu.au
Gavin.Paul@uts.edu.au
University of Technology Sydney: Robotics Institute (UTS:RI)
Sydney, New South Wales, Australia

ABSTRACT

This poster investigates the use of point cloud processing algorithms to provide annotations for robotic manipulation tasks completed remotely via Virtual Reality (VR). A VR-based system has been developed that receives and visualizes the processed data from real-time RGB-D camera feeds. A real-world robot model has also been developed to provide realistic reactions and control feedback. The targets and the robot model are reconstructed in a VR environment and presented to users in different modalities. The modalities and available information are varied between experimental settings, and the associated task performance is recorded and analyzed. The results accumulated from 192 experiments completed by 8 participants showed that point cloud data is sufficient for completing the task. Additional information, either image stream or preliminary processes presented as annotations, was found to not have a significant impact on the completion time. However, the combination of image stream and colored point cloud data visualization modalities was found to greatly enhance a user's performance accuracy, with the number of target centers missed being reduced by 40%.

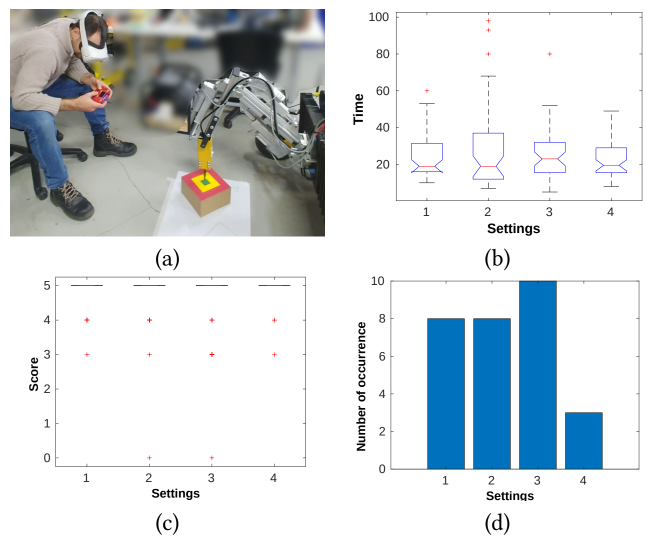


Figure 1: (a) A participant manipulating the manipulator while in VR. (b) - (d) Experimental results for each setting: (b) Completion time; (c) Precision quantified score; (d) Histogram of repetitions where participants missed the center

1 INTRODUCTION

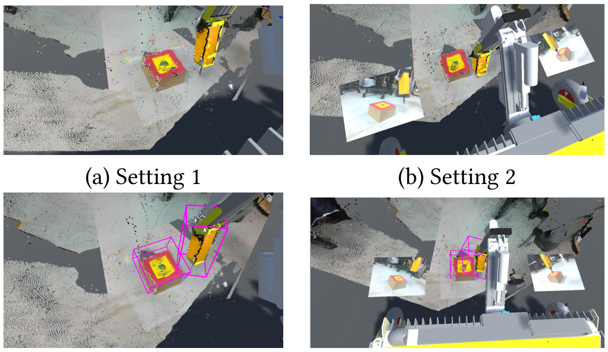
This poster presents the findings ascertained from the observation of users performing real-time remote control manipulation tasks in a VR environment. The VR-based system consists of a custom-built robotic manipulator with remote joystick control and two calibrated static Intel Realsense D435 RGB-D cameras with different viewpoints and partially overlapping fields of view. Both color and depth images are utilized to generate point clouds in real-time within the VR environment. A point cloud processing algorithm is then implemented to generates three-dimensional bounding boxes as annotations.

Previous works focus on quantifying human performance during manipulation tasks to obtain insights into the effect of different visualization settings and control modalities [7][3]. Additionally, the user's presence in a VR environment can be leveraged to provides users with graphical overlays on existing objects in sight. The overlaid information provided can constitute documentation or guidance for the ongoing task [1] [8], or processed sensor data [2]. More sophisticated methods of interaction, such as using a gesture-based interface, enable operator movements to be converted into control commands in real-time [4]. However, such systems lack the haptic feedback provided by external controllers, which has been shown to aid the user in following a planned path [5][6].

2 VR-BASED USER STUDY

A human participation study was designed to investigate the practicality of including a point cloud overlay to provide additional information in VR. The user study conducted involved a remote manipulation task that required participants to move the custom-built robotic arm's end-effector to specified targets using a joystick controller whilst wearing a VR headset. Within the VR environment, camera data ascertained from the 2 RGB-D cameras is rendered as a point cloud. Depending on the task configuration, the bounding boxes and color image streams may be presented. The participants were asked to perform the experiments pertaining to 4 sets of six repetitions. The scene layout was varied after each repetition of a set. The four different settings, shown in Figure 2, each visualizing the environment with a combination of data modalities, were used.

The participant's aim is to move the robot's end-effector to the center of the targets as quickly as possible (Figure 1a). Experimental results collected from participants are shown in Figure 1b to 1d.



(a) Setting 1 (b) Setting 2

(c) Setting 3 (d) Setting 4

Figure 2: Participant's view in VR with differing settings: (a) Point cloud only; (b) Point cloud and two image streams; (c) Point cloud and annotations; (d) Point cloud, two image streams, and annotations

Objective Task Completion Time. The task completion time is determined as the time from when the participant starts manipulating the robotic arm to when they verbally indicate that the task is completed. A one-way ANOVA test was conducted with the null hypothesis being that there is no significant difference between the mean completion time of the various settings implemented during the study. An F-statistic value of 1.26 and a p-value of 0.2887 were obtained from the analysis, showing no statistical significance.

Objective Task Precision. To quantify the accuracy of the participants during the experiment, they received a point score based on if the end-effector touches the target. A one-way ANOVA test was again conducted with the null hypothesis being that there is no significant difference between the mean accuracy of the above-mentioned settings. An F-statistic value of 1.79 and a p-value of 0.15 were reported, supporting the previous statement.

3 DISCUSSIONS

The experimental results indicate that the availability of the information presented through various mediums, namely 2D images, and bounding box annotations, did not significantly impact the objective task completion time, as seen in Figure 1b. The median completion time of the participants collectively for all settings is similar, despite the varied fluctuation range. Figure 1c shows that among the four settings investigated, participants are able to manipulate the end-effector to the target's center in most trials. However, Figure 1d demonstrates a difference in the resulting score when images are available. In settings where images are available (Setting 2 and 4), participants failed to hit the target's center 11 times in total. Within the 11 misses, there were eight occurrences (72%) where the end-effector touched the edge of the green target area, resulting in a score of 4. On the other hand, in other settings, participants missed the center 18 times and scored 4 a total of 11 times (61%). Amongst all the investigated settings, participant performance improved with Setting 4, as indicated by the low number of times the participants missed the green area.

The experimental results suggest that the non-annotated visualization, the presented colored point cloud alone, was sufficient for this remote-controlled manipulation task. However, this may be due to the controlled and ideal indoor setup where the depth sensors are not exposed to excessive infrared noise, such as from the sun. Furthermore, bounding box annotations did not impact the overall performance as the target could be easily identified in the scene. In other scenarios, where the target cannot be distinguished through color or if the color information is unavailable, thus making the target less recognizable, annotations may better assist the user.

4 FUTURE WORK

Future work requires undertaking further experiments regarding the implementation of annotations in VR, especially for scenarios where the targets are difficult to see without additional assistance. Furthermore, this research necessitates experiments to be conducted in more challenging environments. Outdoor environments, particularly those with bland colors, are expected to result in more RGB-D camera data error, which is predicted to significantly reduce a user's performance and increase the need for annotations.

ACKNOWLEDGMENTS

This research is supported by UTS:RI, The Commonwealth of Australia's Department of Industry, Science, Energy and Resources (Innovative Manufacturing CRC Ltd) and Perenti, via its subsidiary Ausdrill. T.L. Vu, D.D.K. Nguyen, D.T. Le and S. Sutjipto are supported by Australian Government Research Training Program Scholarships.

REFERENCES

[1] L. L. Gong, S. K. Ong, and A. Y. C. Nee. 2019. Projection-Based Augmented Reality Interface for Robot Grasping Tasks. In *Proceedings of the 2019 4th International Conference on Robotics, Control and Automation* (Guangzhou, China) *(ICRCA 2019)*. Association for Computing Machinery, New York, NY, USA, 100–104.

[2] Michael Gradmann, Eric M Orendt, Edgar Schmidt, Stephan Schweizer, and Dominik Henrich. 2018. Augmented reality robot operation interface with Google tango. In *ISR 2018; 50th International Symposium on Robotics*. VDE, 1–8.

[3] Dinh Tung Le, Sheila Sutjipto, Yujun Lai, and Gavin Paul. 2020. Intuitive Virtual Reality based Control of a Real-world Mobile Manipulator. In *16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. 767–772.

[4] Yuan Lin, Shuang Song, and Max Q-H Meng. 2016. The implementation of augmented reality in a robotic teleoperation system. In *2016 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. IEEE, 134–139.

[5] D Ni, AWW Yew, SK Ong, and AYC Nee. 2017. Haptic and visual augmented reality interface for programming welding robots. *Advances in Manufacturing* 5, 3 (2017), 191–198.

[6] Sheila Sutjipto, Yujun Lai, Marc G Carmichael, and Gavin Paul. 2020. Fitts' law in the presence of interface inertia. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 4749–4752.

[7] David Whitney, Eric Rosen, Elizabeth Phillips, George Konidaris, and Stefanie Tellex. 2020. Comparing robot grasping teleoperation across desktop and virtual reality with ROS reality. In *Robotics Research*. Springer, 335–350.

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