RAIN: A Vision Calibration Tool using Augmented Reality

Andre Cleaver
Tufts University
andre.cleaver@tufts.edu

Elaine Short
Tufts University
elaine.g.short@usc.edu

Jivko Sinapov
Tufts University
jivko.sinapov@tufts.edu

ABSTRACT

Robots rely on noisy sensors to accomplish their tasks. For example, depth camera sensors output 3D data which are then interpreted by visual algorithms to segment objects of interest. Robots often need their sensor algorithms re-calibrated due to environmental changes, malfunctions, or periodic maintenance to continue operation which is only accomplished with expertise. In this work, we developed an Augmented Reality based calibration tool that enables users of all experiences to calibrate computer vision algorithms. We used common filtering and segmentation techniques for 3D data sets to adjust the rendered visual output. Finally, we describe the design of a study to evaluate four different visualization designs and their effect on user satisfaction and effectiveness.

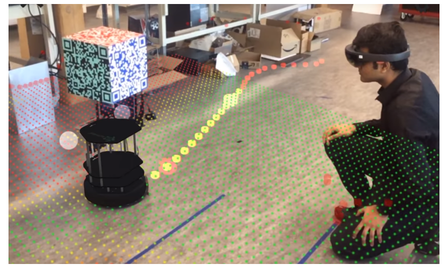

Figure 1: Interactive virtual space shared by a human and a mobile robot. The Hololens renders the robot's Laser scan data (red markers on human), Occupancy-Grid data (red/green dots in grid formation), and planned path trajectory (red/yellow markers).

1 INTRODUCTION

Robots rely on noisy sensors to accomplish their tasks. For example, depth camera sensors output 3D data which are then interpreted by visual algorithms to segment objects of interest. These algorithms often have many parameters, each of which needs to be set appropriately for a given task and environment. The success of a robot's task depends on this calibration which is often performed with human expertise. Robots often need their sensor algorithms re-calibrated due to environmental changes, malfunctions, or periodic maintenance to continue operation. However, current approaches to parameter-setting and calibration require significant expertise, leaving a robot out of commission for periods of time unless a new, user-friendly approach to vision calibration is introduced.

In this paper, we propose and describe the development of an Augmented Reality (AR) system designed to allow non-expert users to calibrate a robot's visual perception software pipeline. AR has helped robots convey their motion intent as well as display their cognitive and sensory data onto the real-world giving humans an intuitive visual aid from the perspective of the robot [3, 17]. AR technology can create an interactive virtual space containing the robot's sensory data and cognitive output (see Figure 1) that a human can access to detect and address malfunctions.

We introduce a "shared reality" framework for understanding the role of AR in human-robot interaction (HRI) for vision calibration, describe the design of a system for using AR to calibrate a multi-parameter RGBD object segmentation pipeline, and describe a planned study to validate the described system. Our proposed framework, RAIN (Robot-parameters Adjusted In No-time), allows the robot to share the state of its sensor data through an AR interface available to the human user. Our proposed system includes a novel approach to parameter-setting which allows naive users to appropriately calibrate complex visual perception pipelines. Using our approach, users can adjust the filtering/segmentation parameters through the AR device that directly modifies the computer vision pipeline running on the robot. The users can see the results of their modification over-layed onto the real world through the AR device, allowing them to choose appropriate parameters without necessarily understanding how the underlying pipeline works.

2 RELATED WORK

Some recent work on automatic sensor calibration has attempted to eliminate the need for a human in the loop entirely. For example, algorithms have been developed that do depth segmentation for object grasping using supervised and reinforcement learning techniques [15, 20]. Other segmentation techniques have leveraged prior image processing techniques and integrated depth information [4]. However, in real-world environments that change over time, it will still be helpful (or even necessary) to have an end user validate that the final parameterization is correct. Our work provides a method for doing this, and relates to work in HRI on establishing common ground between humans and robots and to the literature on augmented reality for robotics. The contribution of this work is (1) identifying an intuitive visual representation of depth information and (2) the development of AR-mediated algorithmic calibration tool for all user experience levels.

2.1 Reaching Common Ground with a Robot

"Theory of Mind" is a set of social skills humans use to attribute beliefs, goals, and desires to other individuals. Within HRI, humans form mental models of robots that express social cues and gestures with the same attributions. Furthermore, the *Common-ground Theory* suggests that mutual understanding between humans during communication is due to coordination of shared information [9]. The common ground theory is then essential towards effective human-robot interactions as it can elicit finer mental models of a robot from a person [14]. Common ground in HRI is reached once the beliefs and perceptions of the physical world as observed by the robot match those observed by the human. Therefore, we focused on an interface design that allows users to visualize the robot's sensor data in the real world to establish a form of common ground and allow the user to set parameters for the vision pipeline.

Related works include common-ground frameworks for human-computer and human-robot interactions using verbal [1, 6, 18, 23] and non-verbal communications [7, 10]. However, this work has not been extended to human-robot interactions beyond natural language. A pilot study by Cheli *et al.* [8] explored AR as an education tool for K-12 students. Middle school students were observed debugging their assigned robots (EV3 Kit) through tablets and initiated group discussions around sensor readings. In this work, we similarly utilize AR technology to establish common ground with a robot. We designed a system that users of all experiences can use to quickly calibrate complex sensor pipelines.

2.2 Augmented Reality for Robotics

AR enables the rendering of computer graphics on to the real-world in real-time [5, 12]. This differs from Virtual Reality (VR) that renders computer graphics over the entire physical environment which places a user into a fully immersive virtual world [16]. Thus, robots can leverage AR as a medium for communication and interaction in HRI. AR interfaces such as the Microsoft HoloLens have shown promise in further enhancing natural HRI by aligning the perspectives of humans and robots [2, 21]. For this study, relevant works include approaches that have robotic systems that communicate through an AR interface.

Walker *et al.* [3] explored an AR design space that conveyed a UAV's motion intent through various explicit and implicit designs. Their study showed that a design that renders more visual explicit information on the environment resulted in better task efficiency and was rated as more clear, usable, and the robot seen as a better member of the team. Muhammad *et al.* [17] had developed an AR design communication tool that displays a robot's cognitive and sensory data as well as prompting for human intervention in order to create a shared reality between humans and robots for communicating and problem-solving. Gadre *et al.* [11] enabled users to create waypoints through an AR-interface to control the motion trajectory of a robotic arm.

Although these prior works have developed tools to visualize a robot's perception of the environment, no prior work has studied how AR can be used to improve perception-pipeline calibration by non-expert users. In this work, we describe a set of design possibilities for such AR-enabled calibration, and show that using our system can allow a non-expert user to correctly and quickly calibrate a perception pipeline.

3 AR FOR SENSOR PIPELINE CALIBRATION

Formally, the problem we are trying to address is the alignment of observations between humans and robots in a shared environment. In particular, all robots with visual pipelines require a user to modify parameters and observe those changes to ensure that the robot performs its function, such as object detection, adequately. Typically, this modification is done on a computer screen and will require some knowledge of coding to know which changes to make. A "good" calibrated system is determined by the user or when the output matches the design specifications.

Figure 2 shows the core challenge of this type of human-robot interaction for calibration. Humans cannot directly access or interpret the complex data within the virtual space of the robot, while the

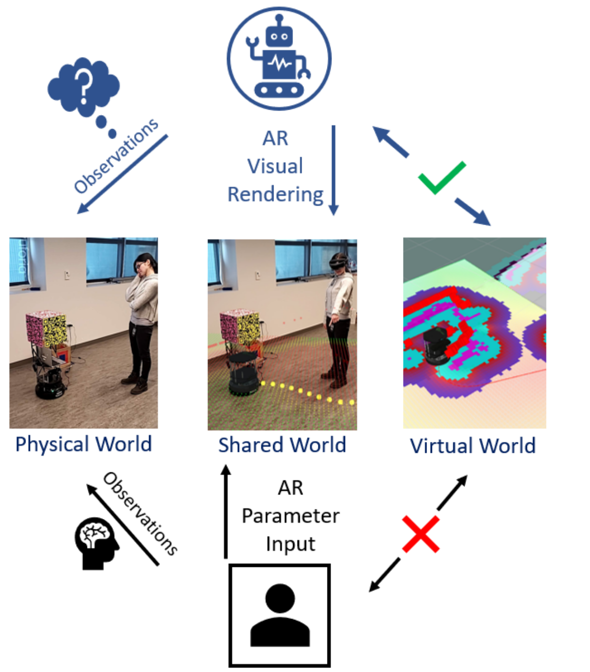

Figure 2: (Right) Physical world dominated by a human, robots are unaware of most objects within environment, (Left) Virtual World dominated by the Robot, humans may not have direct access to the robot's algorithms, (Center) Shared reality world accessed by both parties. Robot renders the virtual markers seen by human with an AR device; parameters can be set by human through a User-interface.

robot has limited human-interpretable knowledge of the physical world. Although anything in AR can be shown on a 2D screen, AR shows data in context of the real-world allowing users to see their immediate changes and avoid the disconnect caused by switching back and forth from a screen to the real-world. Thus, users can directly compare the robots data structures by visualizing them onto the real-world as opposed to on a screen. Our framework enables users and robots to access this "Shared Reality" world where translatable information exchange is most effective.

4 METHODOLOGY

We designed an AR mobile system to render processed robotic sensory data as visual particles with an interactive UI allowing users to modify the distribution of the particles, and we plan to deploy the system in a user study as described below:

4.1 Calibration Task

We consider the task of segmenting objects from a 3D point cloud. Common issues include mistaking two objects as one, losing objects within the background, or misidentifying an object due to a noisy measurement. We use the Point Cloud Library (PCL) [1] as the segmentation pipeline, which contains a variety of filtering and

[1] http://pointclouds.org/

segmentation algorithms for detecting objects in 3D data. The specific steps in the pipeline include a voxel grid filter, a pass-through filter, outlier removal and cylinder segmentation. Each of these components has a set of parameters (e.g., the size of each voxel, radius limits for cylinder detection, etc.).

4.2 System Design

In this section, we describe the AR system which allows users to visualize 3D point clouds while calibrating the robot's visual processing module. The system is composed of three main components:
Depth Camera Sensor: We used the Astra Camera sensor for gathering point-cloud data sets which are then processed by visual algorithms. This camera is commonly used with the Turtlebot2 robotic platform. The depth camera sensor was plugged into an ASUS laptop installed with Ubuntu 16.04 LTS and launched by Robot Operating System (ROS) [19].

AR Device: We used a Samsung Galaxy S9 Android smartphone device with a 12 MP Camera to run the application. It should be noted that the application that runs the proposed design space can be installed on any smartphone or tablet device. The application has the ability to render visual markers similar to those displayed in RViz, the default visualization software that is part of ROS.

Communication Framework: A customized framework ensured that the visualizations and interactive input fields were appropriately displayed. First, the visual prototypes described in 5.2 were developed in *Unity* [2]. Displaying the visual information accounting for the pose of the robot is handled through *Vuforia* [3], an augmented reality software development kit, which enables the rendering of computer graphics onto real-world environments. A websocket connection links the robot platform to the AR device. This link enables ROSBridge to send topic messages as JSON types within ROS to Unity. ROS-Sharp is responsible for the conversion of ROS topics to JSON messages. C# scripts that run within Unity handle the feedback responses.

5 PROPOSED STUDY

We plan to study the effect of several different visualization types on users' ease and effectiveness at calibrating the vision pipeline.

Figure 3: Uniform distribution of rendered particles that correspond to the filtered output of the raw data. (Left) Low density and (Right) High density.

[2]https://unity.com/
[3]https://developer.vuforia.com/

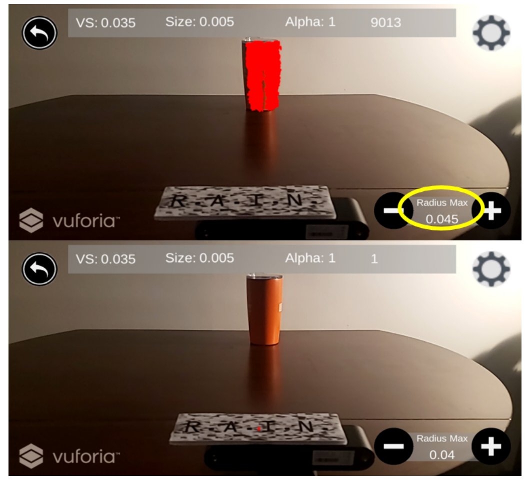

Figure 4: Red particles rendered over a detected real-world object. The particles correspond to points in the raw data that fit a cylinder model of a given radius. A user adjusts the radius parameter to detect the object.

5.1 Interaction Design

Preliminary validation of the system was conducted in a tabletop workspace with a cylinder detection task. The workstation consisted of a workbench and a depth camera sensor attached to an adjustable arm-bar secured to the workbench resulting in a front view of the workbench as seen in Figure 3. Various cylindrical objects are placed on the workbench which can be detected by the sensor. A user can access the visualizations by placing the target image within the AR devices field of view which will render the graphics on the screen. Users then can adjust the parameters displayed on the AR device. Figure 4 shows an adjustable radius parameter at the bottom right of the interface.

5.2 Visualization Designs

We consider four different visualization modes for displaying the 3D point cloud data set to the human user:
Uniform density: Raw point-cloud data points are rendered uniformly after voxel-grid filtering, as shown in Figure 3.
Non-Uniform density: This visualization identifies the ground plane and renders the plane with a sparse distribution of markers, while identified objects will be rendered with a dense surface area of markers.
Heterogeneous density: Identified objects are rendered with visual markers that differ from the ground plane. (i.e. slightly larger, white in color, and spherical in shape, etc.)
Negative Space density: Within this visualization, the particles are rendered around the space of detected objects.

5.3 Outcome Measures

Objective and subjective measurements will characterize the effectiveness of the calibration tool. Task efficiency measures the total time participants spent completing the task (lower times indicated better performance/efficiency). A participant's recorded time will only be included in the analysis if no technical issues occur. Technical issues include any freezing of the AR-Device or on the robot end. Consistent start and end times for each participant will ensure fair results. A percentage of the objects detected will be recorded to measure accuracy. As for the subjective ratings, a 7-point scale Likert-style questionnaire common among HRI papers, will be utilized to measure user perceptions and preferences. Scales will rate *Design Clarity*, *Design Usability*, and *Recommendation for other users*. Open-ended responses will be recorded for any improvements or changes to the study and will be administered at the conclusion of the study. They will include statements that refer to "working with the AR interface", "difficulty to correct the robot perception", "AR-device provided enough information". We will then analyze the data using a one-way Analysis of Variance (ANOVA) with experimental condition as the fixed effect. Post-hoc tests will evaluate the design methods against a baseline condition. In addition, the NASA-TLX workload survey metric will measure cognitive workload when analyzing the visualization techniques [13, 22].

5.4 Procedure

(1) First, participants are read identical instructions for the study and the task they will perform, and a consent form is then obtained. In addition, the participant is randomly selected a visualization (i.e. AR Visualization Designs) they will utilize. (2) Next, participants familiarize themselves with the AR-device and the functionality of the application; this will take approximately 5 minutes. Participants who are assigned to the baseline condition are verbally instructed that the robot will make observations in the world and can be seen through a non AR-device. (3) Participants will then perform the instructed task within 10 minutes. Specifically, participants will try to identify all cylindrical objects placed on a tabletop by only using the interactive buttons on the AR interface. (4) After the allotted time has elapsed, participants will be notified that the task is over, and a post-questionnaire will be administered.

6 DISCUSSION AND CONCLUSION

In this work, we developed an AR-based calibration tool that enables users of all experiences to calibrate computer vision algorithms. We used common filtering and segmentation techniques for 3D data sets to adjust the rendered visual output. Finally, we describe the design of a study to evaluate four different visualization designs and their effect on user satisfaction and effectiveness.

There are some limitations to this work: The visualizations developed so far are only a small section of possible ways to visualize depth information. We plan to explore other visualization techniques such as lifetime of visual rendering and substitution of point-clusters to gauge user performance in a similar experiment. One other limitation is the narrow bandwidth that results in lag times when sending large messages through a web-socket connection. Changing the parameters in the device interface does not immediately update the PointCloud visuals which can be a source of confusion. To address this limitation, we plan to develop an adaptive visualization approach to render 3D data at various resolutions. Our research will demonstrate the effectiveness of AR technology by providing users of all levels of expertise an intuitive visual aid to calibrate sensor algorithms and as a result improve depth-sensor based tasks.

ACKNOWLEDGMENTS

This research was supported in part by the Tufts Center for Applied Brain and Cognitive Science, PTC inc., The National GEM Consortium (GEM), and the Verizon Foundation.

REFERENCES

[1] [n. d.]. Collaborative Effort Towards Common Ground in Situated Human-robot Dialogue *(HRI '14)*. New York, NY, USA.

[2] [n. d.]. Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays. 38 ([n. d.]).

[3] [n. d.]. Communicating Robot Motion Intent with Augmented Reality.

[4] Alexey Abramov, Karl Pauwels, Jeremie Papon, Florentin Wörgötter, and Babette Dellen. 2012. Depth-supported real-time video segmentation with the Kinect. In *2012 IEEE Workshop on the Applications of Computer Vision (WACV)*. 457–464. https://doi.org/10.1109/WACV.2012.6163000 ISSN: 1550-5790.

[5] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications* (2001), 14.

[6] Susan E. Brennan and Eric A. Hulteen. 1995. Interaction and feedback in a spoken language system: a theoretical framework. *Knowledge-Based Systems* 8, 2 (April 1995), 143–151. https://doi.org/10.1016/0950-7051(95)98376-H

[7] Andrew G Brooks and Cynthia Breazeal. 2006. Working with robots and objects: Revisiting deictic reference for achieving spatial common ground. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 297–304.

[8] Mark Cheli, Jivko Sinapov, Ethan Danahy, and Chris Rogers. 2018. Towards an Augmented Reality Framework for K-12 Robotics Education. In *1st International Workshop on Virtual, Augmented and Mixed Reality for Human-Robot Interaction*.

[9] Herbert H. Clark. 1992. *Arenas of language use*. University of Chicago Press.

[10] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. 2013. Legibility and predictability of robot motion. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 301–308.

[11] Samir Yitzhak Gadre, Eric Rosen, Gary Chien, Elizabeth Phillips, Stefanie Tellex, and George Konidaris. 2019. End-User Robot Programming Using Mixed Reality. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, Montreal, QC, Canada, 2707–2713. https://doi.org/10.1109/ICRA.2019.8793988

[12] Scott A. Green, Mark Billinghurst, XiaoQi Chen, and J. Geoffrey Chase. 2008. Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design. *International Journal of Advanced Robotic Systems* 5, 1 (March 2008), 1. https://doi.org/10.5772/5664

[13] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[14] S. Kiesler. 2005. Fostering common ground in human-robot interaction. In *RO-MAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* https://doi.org/10.1109/ROMAN.2005.1513866

[15] Ian Lenz, Honglak Lee, and Ashutosh Saxena. 2015. Deep learning for detecting robotic grasps. *The International Journal of Robotics Research* 34, 4-5 (April 2015), 705–724. https://doi.org/10.1177/0278364914549607

[16] P. Milgram, S. Zhai, D. Drascic, and J. Grodski. 1993. Applications of augmented reality for human-robot communication. In *Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '93)*, Vol. 3. 1467–1472 vol.3. https://doi.org/10.1109/IROS.1993.583833

[17] F. Muhammad, A. Hassan, A. Cleaver, and J. Sinapov. 2019. Creating a Shared Reality with Robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 614–615. https://doi.org/10.1109/HRI.2019.8673191

[18] Tim Paek and Eric Horvitz. 1999. Uncertainty, Utility, and Misunderstanding: A Decision-Theoretic Perspective on Grounding in Conversational Systems.

[19] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. 2009. ROS: an open-source Robot Operating System. In *ICRA workshop on open source software*, Vol. 3. Kobe, Japan, 5.

[20] Deepak Rao, Quoc V. Le, Thanathorn Phoka, Morgan Quigley, Attawith Sudsang, and Andrew Y. Ng. 2010. Grasping novel objects with depth segmentation. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2578–2585. https://doi.org/10.1109/IROS.2010.5650493 ISSN: 2153-0858.

[21] Eric Rosen, Nishanth Kumar, Daniel Ullman, David Whitney, George Konidaris, and Stefanie Tellex. [n. d.]. Bridging the Semantic Gap for Robots: Action-Oriented Semantic Maps via Mixed Reality. ([n. d.]).

[22] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. 2006. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. 33–40.

[23] Tom Williams, Gordon Briggs, Bradley Oosterveld, and Matthias Scheutz. 2015. Going beyond literal command-based instructions: Extending robotic natural language interaction capabilities. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.