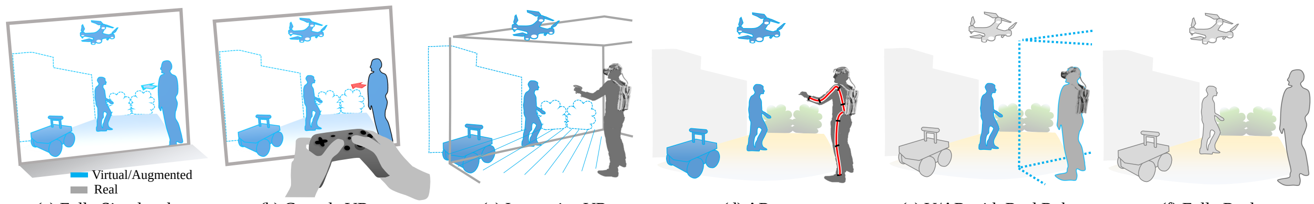
Mixed-Reality Testbed for Robotic Systems with Human Interaction

Bhoram Lee, Jonathan Brookshire, and Supun Samarasekera



(a) Fully Simulated (b) Console VR (c) Immersive VR (d) AR (e) V/AR with Real Robots (f) Fully Real

Fig. 1: Mixed-Reality system with varied realism of robotic systems and human behavior. Virtual components are illustrated in blue (best viewed in color): (a) Fully simulated, (b-e) human characters are controlled by real humans (via (b) console-based Virtual Reality (VR), (c) immersive VR, (d) Augmented Reality (AR) interfaces, and (e) VR/AR with real robots), and (f) fully real.

***Abstract*— Advances in reinforcement learning are being brought to physical robotic systems. Enabling them to deal with humans in the real world is critical, yet how to do so safely is an open question. This paper presents a Mixed-Reality (MR) system toward human-centered development of robotic systems emphasizing benefits as a Reinforcement Learning (RL) data collection and testbed tool. It allows real humans to interact with various levels of simulation to maintain both realism and safety. Collected data can be used for improvement/evaluation of algorithms and simulation models.**

I. Introduction

Interests are increasing in bringing advances in machine learning to physical robotic systems such as autonomous driving cars [1] and manipulators [2]. Although we have witnessed successful algorithms to control and operate robots, an issue still remains when it comes to their interaction with humans [3]. Since the ultimate use of physical robots would happen in the real world, enabling them to learn to deal with humans is essential. However, developing and testing robotic systems operating alongside humans in a safe, yet realistic, environment is challenging in two aspects. First, human models in simulation are very limited. Second, testing physical systems with real humans can be unsafe. Efforts are being made to learn safe algorithms [4], but we still lack a good way to verify safety before deployment.

One way to increase behavioral realism while safely verifying performance is to directly use human input via Mixed Reality (MR), referring the virtuality continuum ranging between the completely virtual and the completely real as defined in [5]. Researchers have suggested ideas for integration of MR with complex robotics systems for development and test [6–9]. We can further exploit MR technology as a tool to train and test robotic systems with humans. A recent study [10] presented a MR-based framework for validating an autonomous vehicle's performance in the presence of pedestrians. By creating a virtual environment shared between the vehicle and pedestrian, the testbed provided a way to test algorithms under model-free human behavior. While appreciating the previous studies revealing the value of MR testbeds in robotics applications, we further contribute more systematic views of MR testbeds as means of training and testing algorithms involving human-robot interactions and also share our implementation design.

Authors are with the Center for Vision Technologies, SRI International, Princeton, NJ, USA (e-mail: `<first>.<last>@sri.com`)

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. 140D6319C00007. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

II. MR for Robotic System with Realism

In a MR environment, there are virtual and real entities that interact with each other and are spatially mapped from real to virtual, and vice-a-versa. Varied ways to implement MR interfaces allow different levels of virtuality or reality as shown in Fig. 1. Based on virtualization levels of robots, humans, and the environment, we may consider (a) fully simulated, (b-d) MR with virtual robots, (e) MR with real robots, and (f) fully real. Note that humans can interface with the system in several ways. Virtual humans mean programmed human characters. Real humans can be immersed in the virtual scene using Console, VR, and AR interfaces, or physically present in the real world.

The foremost concern when deploying or testing robots near humans is safety. Only in (f), robots can physically harm people when they share the same environment. We consider this as the final deployment stage rather than a developing approach.

The completely virtual approach, Fig. 1 (a), affords significant experimental repeatability. This is very common and useful way to obtain a massive dataset for training as well as to test low-level algorithms especially in the initial stages of development. Data collection can be conducted faster than real time, possibly in parallel, without any spatial restriction. In configurations (b)-(e) where the virtual characters are controlled/mapped by real human players, there are three ways to use the data. First, the data can be added to the replay buffer during RL training. The data size would be small because human inputs will not be faster than real time nor parallelizable, yet this may be useful for refinement or

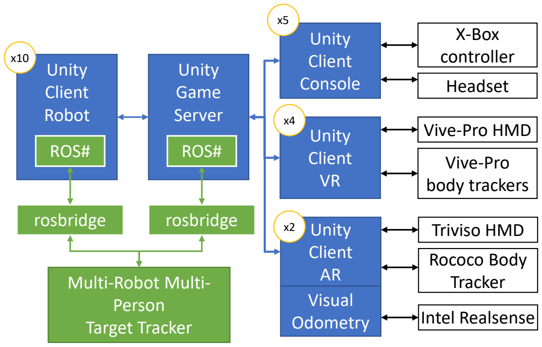


Fig. 2: The architecture includes ROS (green) and the Unity game engine [13] (blue). It consists of a Game Server and clients connected to the server. The clients include robot clients, and Console/AR/VR player clients.

meta-testing [11]. Second, the recorded human inputs may be used to improve the programmed human character for configuration (a). Third, the data can be used for evaluation of the algorithm. This is what simulation alone cannot achieve and what a fully real, physical system cannot achieve safely. As we envision continued learning problems of a longer term, all the usages of human data are meaningful.

The similarity between the human's behavior in the virtual and real worlds, or "transfer" [12], may depend on many factors. Among the MR interfaces, a joystick usually provides the lowest level of transfer, and VR and AR modalities are capable of a better transfer. Photo-realism, higher frame rates, or audio effects may affect transfer as well. Overall, as real components increase (going from (a) to (f)), realism improves in terms of transfer and fidelity. Improved realism, however, comes with added cost and expert knowledge. For example, console interfaces, consisting of just a laptop, joystick, and headset are relatively inexpensive and easily deployable, compared with an AR system. Thus, it will be possible to recruit more human players if with the same budget. Depending on the development phase or the purpose of study, we may benefit from a proper choice of mixed-reality setting.

Other technical considerations beside handling MR devices include that, when virtual and real entities are combined, all of them must spatially align. A low quality alignment may degrade the user experience of human players or alter robots' behavior in an unexpected way.

III. System Implementation

In this section, we share a short description of our implementation of a MR testbed presented in the previous section. The system architecture is shown in Fig. 2. The Game Server, responsible for maintaining the game state and scenario loading, publishes a ground truth message that includes states of all entities and sensor outputs. Autonomous Non-Player Characters (NPCs), controlled by behavior trees, can be added to scale up the number of characters in addition to real human players. Each robot client subscribes to a command ROS message allowing the publisher to move the robot and publishes another message that includes sensor data and its states. Other clients include Console, VR, and AR clients

Fig. 3: The test environment is derived from photogrammetric reconstructions [20, 21] of parts of SRI's Princeton campus (spanning over $180 \times 120m^2$). (Left) The actual site, (Right) its virtual replica as a mesh model.

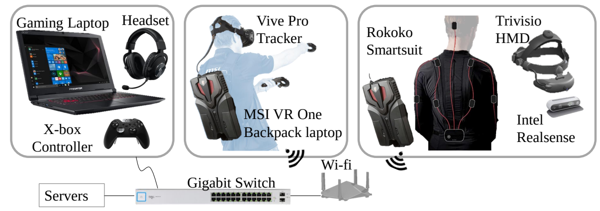


Fig. 4: Console (left), VR (middle), and AR (right) interfaces.

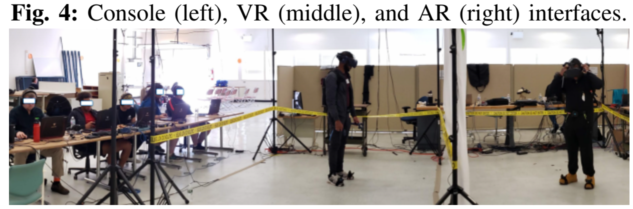


Fig. 5: Experiment with real human players (five console players and two VR players are seen in the figure.)

for human players. The console-based interfaces (Fig. 4, left) include a gaming laptop, Xbox controller [14], and a headset. Using the controller, the player can walk, run, pickup/deposit an object, and more. VR players (Fig. 4, middle) wear an MSI VR One [15] computational backpack and a VIVE Pro HMD [16]. They hold a VIVE controller in each hand, and trackers are mounted on the front waist and both feet. They can freely walk around a 3.5m x 3.5m play area. The AR interfaces (Fig. 4, right) are also wearable devices including a MSI VR One backpack and a Rokoko Smartsuit [17] for full-body motion capture. The AR headset consists of a Trivisio HMD [18] paired with a physically mounted Intel Realsense camera and provides video see-through capabilities. Video frames and IMU data from the camera are fused with GPS via visual odometry [19] to provide pose. In our other study (not published at the moment), we performed experimental demonstration of the system for a multi-robot multi-person tracking and monitoring application Fig. 5.

IV. Discussion

We have presented an approach toward human-centered development of robotic systems using a Mixed-Reality testbed as well as our implementation of the system. This provides us a safe way to iterate training/developing/testing algorithms involving human interactions. Our study was driven by a safety requirement and a need to have real humans interact with the system: MR offered the best of both worlds. We believe that MR testbeds will become increasingly common given a need for close interaction between robots and humans.

REFERENCES

[1] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani, and P. Pérez, "Deep reinforcement learning for autonomous driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[2] O. Kroemer, S. Niekum, and G. Konidaris, "A review of robot learning for manipulation: Challenges, representations, and algorithms.," *J. Mach. Learn. Res.*, vol. 22, pp. 30–1, 2021.

[3] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, 2019.

[4] M. El-Shamouty, X. Wu, S. Yang, M. Albus, and M. F. Huber, "Towards safe human-robot collaboration using deep reinforcement learning," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4899–4905, 2020.

[5] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Transanctions on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[6] A. H. Göktoğan and S. Sukkarieh, "An augmented reality system for multi-uav missions," in *the Simulation Conference and Exhibition, SimTect*, 2005.

[7] I. Y.-H. Chen, B. MacDonald, and B. Wunsche, "Mixed reality simulation for mobile robots," in *IEEE International Conference on Robotics and Automation*, pp. 232–237, 2009.

[8] I. Y. H. Chen, B. Macdonald, and B. Wuensche, "A flexible mixed reality simulation framework for software development in robotics," 2011.

[9] W. Hoenig, C. Milanes, L. Scaria, T. Phan, M. Bolas, and N. Ayanian, "Mixed reality for robotics," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5382–5387.

[10] M. R. Zofka, M. Essinger, T. Fleck, R. Kohlhaas, and J. M. Zöllner, "The sleepwalker framework: Verification and validation of autonomous vehicles by mixed reality lidar stimulation," in *IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR)*, pp. 151–157, 2018.

[11] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International Conference on Machine Learning*, pp. 1126–1135, PMLR, 2017.

[12] F. Rose, E. Attree, B. Brooks, D. Parslow, P. Penn, and N. Ambihaipahan, "Training in virtual environments: Transfer to real world tasks and equivalence to real task training," *Ergonomics*, vol. 43, pp. 494–511, 05 2000.

[13] J. K. Haas, "A history of the unity game engine," 2014.

[14] "Microsoft Xbox." `https://www.xbox.com`. Last Accessed: 2021-8-12.

[15] "MSI VR One Backpack - MR READY." `https://mr.msi.com/Backpacks/vrone`. Last Accessed: 2021-8-12.

[16] "VIVE Pro." `https://www.vive.com`. Last Accessed: 2021-8-12.

[17] "Rokoko - Motion Capture System - Smartsuit Pro." `https://www.rokoko.com/products/smartsuit-pro`. Last Accessed: 2021-8-12.

[18] "Triviso SXGA61." `https://www.trivisio.com/hmd-nte`. Last Accessed: 2021-8-12.

[19] T. Oskiper, S. Samarasekera, and R. Kumar, "Multi-sensor navigation algorithm using monocular camera, imu and gps for large scale augmented reality," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 71–80, 2012.

[20] B. C. Matei, H. S. Sawhney, S. Samarasekera, J. Kim, and R. Kumar, "Building segmentation for densely built urban regions using aerial lidar data," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008.

[21] R. Hadsell, B. Matei, G. Salgian, A. Das, T. Oskiper, and S. Samarasekera, "Complex terrain mapping with multi-camera visual odometry and realtime drift correction," in *International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, pp. 493–500, 2012.