Augmented Reality Visualizations using Imitation Learning for Collaborative Warehouse Robots

Extended Abstract

Kishan Chandan
SUNY Binghamton
Binghamton, New York, USA
kchanda2.binghamton.edu

Jack Albertson
SUNY Binghamton
Binghamton, New York, USA
jalbert5@binghamton.edu

Shiqi Zhang
SUNY Binghamton
Binghamton, New York, USA
zhangs@binghamton.edu

ABSTRACT

Augmented reality (AR) technologies have been applied to human-robot collaboration (HRC) domains to enable people to visualize the state of the robots. Current AR-based visualization strategies are manually designed. This design process requires a lot of human efforts, and domain knowledge. When too little information is visualized, human users find the AR interface not useful; when too much is visualized, they find it difficult to process the visualized information. In this paper, we develop an intelligent AR agent that learns visualization policies (what to visualize, when, and how) from demonstrations. We developed a Unity-based platform for simulating warehouse environments where human-robot teammates work on collaborative delivery tasks. We have collected a dataset that includes 6000 demonstrations of visualizing robots' current and planned behaviors. Our results from experiments with real human participants show that, compared with competitive baselines from the literature, our learned visualization strategy significantly increases the efficiency of human-robot teams in delivery tasks.

KEYWORDS

Augmented Reality; Multi-agent systems; Imitation Learning

ACM Reference Format:
Kishan Chandan, Jack Albertson, and Shiqi Zhang. 2022. Augmented Reality Visualizations using Imitation Learning for Collaborative Warehouse Robots: Extended Abstract. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 3 pages.

1 INTRODUCTION

Augmented Reality (AR) has been employed for human-robot collaboration (HRC) to provide an alternative communication mechanism with high bandwidth and low ambiguity [1, 5, 6, 8, 9, 11, 12, 15]. For instance, researchers have developed AR systems that allow humans to visualize the state of the robots [10], as well as the robot intentions [2, 14]. Existing AR-based HRC systems employ a static visualization strategy where the visualizations are always displayed to the user. As a result, those systems are not able to take important runtime factors into account (e.g., the number of robots, the current status of robots, the future intentions of robots, and the current status of human) to adapt the visualizations. The consequence is that AR interfaces can alter attentional focus and result in *inattentional blindness* [7], if there is visual clutter caused by too many unwanted visualizations. Such concerns motivate our research on enabling our AR agent to learn a policy for dynamically selecting visualization actions given the state of the human-multi-robot system. We design a framework called, ***Visualizations for Augmented Reality using Imitation Learning*** **(VARIL)**, for human-multi-robot collaboration in a shared environment.

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

2 VARIL FRAMEWORK

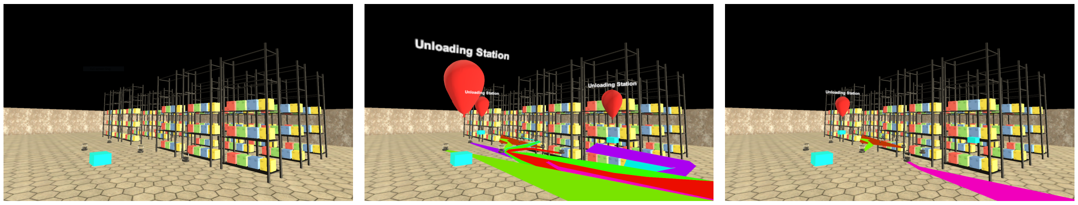
In this section, we describe our VARIL framework, and how it enables human-multi-robot teams to collaborate in a shared environment. VARIL allows the human to track the status of a team of robots using an AR interface. Moreover, VARIL supports learning an AR visualization policy using expert demonstrations, where the visualization agent uses the learned policy to dynamically select a visualization action based on the state of the human-multi-robot teams. Consider a team of robots working with a human worker in a shared space. The team of robots constantly share their state and plans using the AR interface, which is used by the human worker to track the status of the team of robots. The human worker simultaneously collaborates with the team of robots to complete the tasks. VARIL also consists of a human expert that gives demonstrations of AR visualizations at runtime, indicating what information should be visualized (or not) at specific times. Once a new policy is generated, the visualization agent updates the visualizations in the AR interface to mimic the expert demonstrator's suggested actions.

2.1 State-Action Space of Visualization Agent

In our implementation of VARIL, we learn a visualization strategy for two types of visualizations, one for the visualization of robots, and another one for the visualization of shared tasks. The state-space (S_R) of our visualization agent for the robot consists of the following:

- $humanState$: $\{close, moderate, far\}$
- $robotTaskState$: $\{picking, dropping\}$
- $robotRemainingTasks$: $\{few, many\}$
- $robotWaitingTime$: $\{short, medium, long\}$
- $nearbyRobots$: $\{few, many\}$
- $nearbyRobotVizStatus$: $\{few, many\}$

The above state representation consists of all the key states that represent the entire human-multi-robot system from each agent's perspective. The action space of the visualization agent depends on the number of different visualizations. In our case for robots, we have three different visualizations, which are, the live location, planned trajectory, and transparent avatar to visualize the status of the robots. Each of the visualizations can be turned on or off for



(a) No AR visualization (b) Full AR visualization (c) Learned AR visualization (ours)

Figure 1: Three AR visualization strategies in a virtual warehouse environment, where a virtual human works with a team of mobile robots on collaborative delivery tasks. (a) No AR visualization, where the human worker does not know where some robots are, and what the robots plan to do. (b) Full AR visualization, where the human worker can be overwhelmed by the visual indicators. (c) Our learned AR visualization, where the AR agent uses a learned policy to dynamically determine a visualization strategy based on the current world state of both human and robots.

each agent. The state-space (S_D) of our visualization agent for each drop station is a Cartesian product of the following sets:

- $humanState$: $\{close, moderate, far\}$
- $nRobotsAtDropStation$: $\{few, many\}$
- $robotsWaitTimeAtDropStation$: $\{short, medium, long\}$

The action space for each drop station consists of enabling or disabling the balloon on the drop station. The state and action space can vary based on the implementation of the visualization agent.

Data Collection and Policy Learning: We collected a dataset of 6000 demonstrations provided by a human expert at runtime during which a human worker was working with a team of robots to collaborate on a delivery task. The human worker and the human expert both were provided with an AR interface to track the robots. The human worker used the AR interface to collaborate with the robots, whereas the role of the human expert was to provide feedback on the visualizations seen using the AR interface.

3 EXPERIMENTS

We conducted experiment with 25 participants in a simulated warehouse environment that we developed using Unity (Figure 1). With this experiment, we aim to evaluate the following hypothesis: I) VARIL improves the overall efficiency in human-multi-robot team task completion in comparison to other AR-based methods from the literature that employ static visualization policies.

3.1 Experiment setup

We deployed the simulated warehouse environment to a web server to facilitate the online experiment. We have replicated the visualizations of two different systems from the literature, called ARROCH [4], and CRMIAR [13] [1] to make comparisons of their visualization strategies with VARIL. In our experiment, human participants were able to teleoperate the virtual human around the environment. Each robot is assigned three boxes for each trial and once the last box has been dropped off by the robot, it will begin navigating to its starting position. The trial is completed once all twelve robots have delivered their three boxes, and reached their starting position.

[1]We create the acronym of CRMIAR that indicates "Communicating Robot Motion Intent with Augmented Reality" for the simplicity of referring to the method.

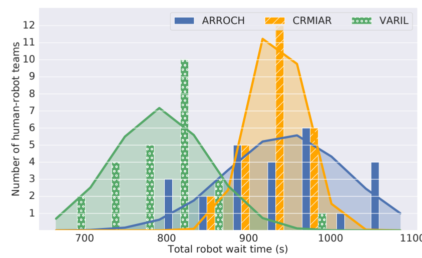


Figure 2: A histogram to show the distribution of total waiting time (second) of all robots;

3.2 Results

Figure 2 shows the histogram, where the x-axis represents the total wait time of all the robots in seconds, whereas the y-axis represents the number of human-robot teams with the corresponding robot wait times. From the figure, it can be observed that most human-robot teams, VARIL had the shortest robot wait times as compared to ARROCH and CRMIAR. Also, we plotted the Gaussian curves to clearly show the distribution of the data points for all the methods. We also analyzed the statistical significance, in every trial, we sum up the task completion time of all the robots. We found that VARIL performed **significantly better** than both ARROCH and CRMIAR, where $0.01 < p - value < 0.05$. Additionally, by comparing the total robot wait times, we observed that the wait times for robots in VARIL were significantly shorter than the baselines. All of these results support Hypothesis-I that states VARIL improves the human-robot team's task completion efficiency.

4 CONCLUSIONS AND FUTURE WORK

In this paper, we present our framework, *Visualizations for Augmented Reality using Imitation Learning* (VARIL), that introduces a learning-based Augmented Reality (AR) visualization strategy for human-multi-robot collaboration. We have designed a system that learns a visualization policy using imitation learning, where the visualization agent dynamically selects a visualization action based on the state of the human-multi-robot system. We compared our framework with competitive baselines from the literature, and the results suggest that VARIL significantly increases the efficiency of human-robot collaboration. In the future, we will evaluate the user-experience performance of VARIL through questionnaires and explore VR-based human-robot collaboration [3].

REFERENCES

[1] Connor Brooks and Daniel Szafir. [n.d.]. Visualization of intended assistance for acceptance of shared control. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 11425–11430.

[2] Ravi Teja Chadalavada, Henrik Andreasson, Robert Krug, and Achim J Lilienthal. 2015. That's on my mind! robot to human intention communication through on-board projection on shared floor space. In *Mobile Robots (ECMR), 2015 European Conference on*. IEEE, 1–6.

[3] Kishan Chandan, Jack Albertson, Xiaohan Zhang, Xiaoyang Zhang, Yao Liu, and Shiqi Zhang. [n.d.]. Learning to Guide Human Attention on Mobile Telepresence Robots with 360° Vision. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 5297–5304.

[4] Kishan Chandan, Vidisha Kudalkar, Xiang Li, and Shiqi Zhang. 2021. ARROCH Augmented Reality for Robots Collaborating with a Human. *In 2021 International Conference on Robotics and Automation (ICRA)* (2021).

[5] Jeffrey Delmerico, Roi Poranne, Federica Bogo, Helen Oleynikova, Eric Vollenweider, Stelian Coros, Juan Nieto, and Marc Pollefeys. 2022. Spatial Computing and Intuitive Interaction: Bringing Mixed Reality and Robotics Together. *IEEE Robotics & Automation Magazine* (2022).

[6] Morteza Dianatfar, Jyrki Latokartano, and Minna Lanz. 2021. Review on existing VR/AR solutions in human–robot collaboration. *Procedia CIRP* 97 (2021), 407–411.

[7] Benjamin J Dixon, Michael J Daly, Harley Chan, Allan D Vescan, Ian J Witterick, and Jonathan C Irish. 2013. Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surgical endoscopy* 27, 2 (2013), 454–461.

[8] 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 (2008), 1.

[9] Terran Mott, Thomas Williams, Hao Zhang, and Christopher Reardon. 2021. You Have Time to Explore Over Here!: Augmented Reality for Enhanced Situation Awareness in Human-Robot Collaborative Exploration. (2021).

[10] Faizan Muhammad, Amel Hassan, Andre Cleaver, and Jivko Sinapov. 2019. Creating a Shared Reality with Robots. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*.

[11] James F Mullen, Josh Mosier, Sounak Chakrabarti, Anqi Chen, Tyler White, and Dylan P Losey. 2021. Communicating inferred goals with passive augmented reality and active haptic feedback. *IEEE Robotics and Automation Letters* 6, 4 (2021), 8522–8529.

[12] M Ostanin, R Yagfarov, D Devitt, A Akhmetzyanov, and A Klimchik. 2021. Multi robots interactive control using mixed reality. *International Journal of Production Research* 59, 23 (2021), 7126–7138.

[13] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating Robot Motion Intent with Augmented Reality. In *Proceedings of the International Conference on Human-Robot Interaction*.

[14] A. Watanabe, T. Ikeda, Y. Morales, K. Shinozawa, T. Miyashita, and N. Hagita. 2015. Communicating robotic navigational intentions. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

[15] Ziyi Zhang, Sara Willner-Giwerc, Jivko Sinapov, Jennifer Cross, and Chris Rogers. 2021. An Interactive Robot Platform for Introducing Reinforcement Learning to K-12 Students. In *International Conference on Robotics in Education (RiE)*. Springer, 288–301.