Augmented Reality Visualizations using Imitation Learning for Collaborative Warehouse Robots

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

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.

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 ($S_q$) 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

ACM Reference Format:

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.

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-1 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


