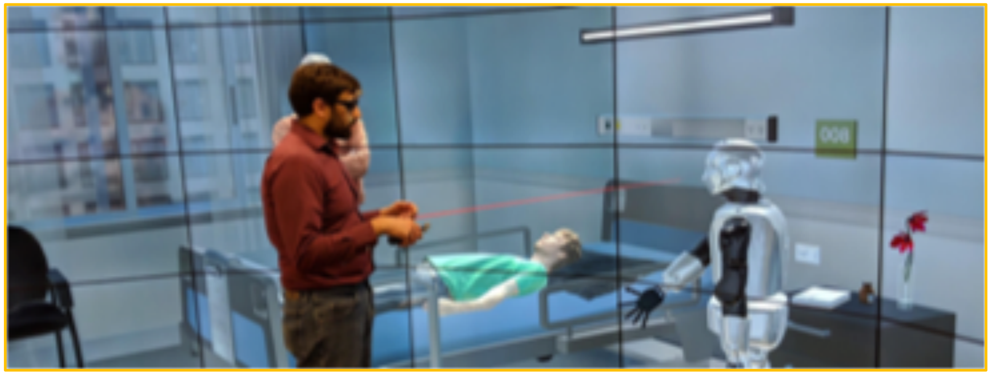
EAGER: Learning Language in Simulation for Real Robot Interaction

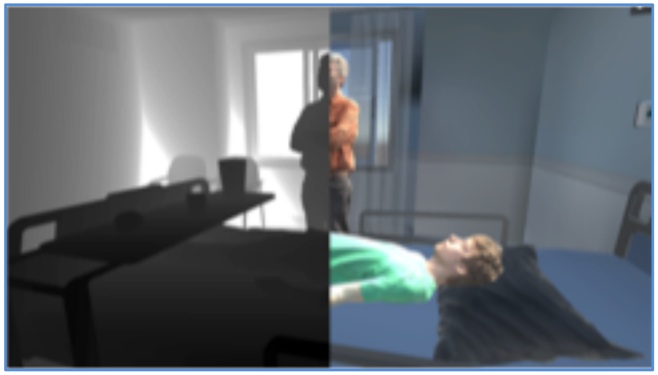
Cynthia Matuszek, Francis Ferraro, Don Engel (*cmat, ferraro, donengel@umbc.edu*)

Grounded language interactions are key to deploying robots in human environments, but gathering data from people in a variety of settings is a bottleneck. We collect simulated sensor data from the robot's side, including human interaction, in sophisticated simulations.



Human data: Using a monitor wall to talk to a virtual robot, using a combination of Unity, ROS, and Gazebo.

Robot sensors: A split view of simulated robot sensor data. Right is RGB, left is from a depth sensor.



Approach & System Design

- The Unity game engine lets us quickly build rich, varied worlds and scenarios
- ROS manages simulated sensing and actuation
- The Gazebo simulation environment renders robots from URDF spec files and provides simulated environment sensor readings
- A video wall with head- and hand-tracking provides an immersive environment
- Includes Vicon tracking system and Kinect
- Content can appear on a video wall and/or in inexpensive VR headsets

1. *Virtual Reality and Photogrammetry for Improved Reproducibility of Human-Robot Interaction Studies*. IEEE VR (poster). Mark Murnane, Max Breitmeyer, Cynthia Matuszek, Don Engel. 2019
2. Learning from Human-Robot Interactions in Modeled Scenes. ACM SIGGraph (short). Mark Murnane, Max Breitmeyer, Francis Ferraro, Cynthia Matuszek, Don Engel. 2019.

Goals and Broader Impacts

Targeted scenes are used to collect:

- Gesture, gaze, and speech by a human
- Sensor data from robot(s)

By capturing the human model rather than using video, we create a replayable corpus for learning.

This will support human-robot interaction research in the community, but also **lower the barrier to entry for robotics research** by providing a rich, inexpensive testbed.

2020 National Robotics Initiative (NRI) Principal Investigators' Meeting
FEBRUARY 27 - 28, 2020 | ARLINGTON, VIRGINIA

Award ID#:1940931