Human-Robot Communication with VR/AR Technology

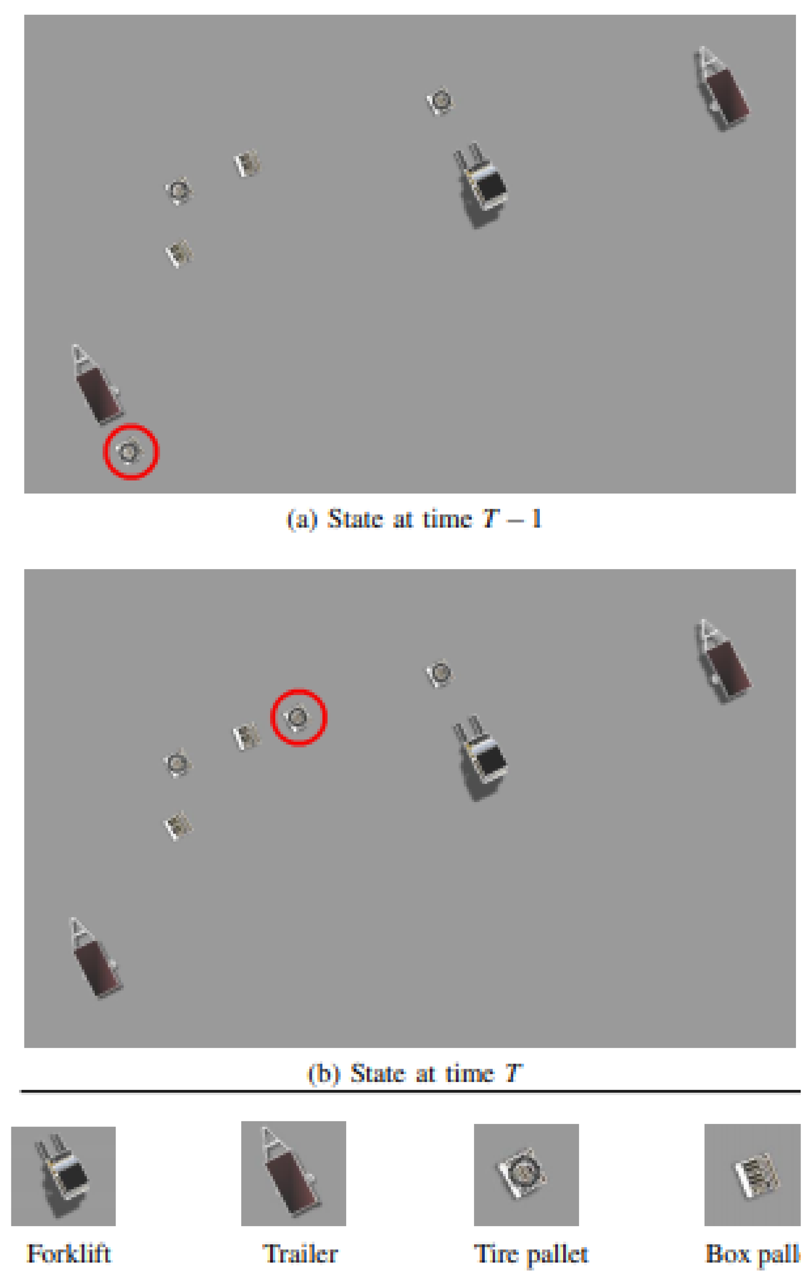
Michael Chung, Computer Science and Data Science

Mentor: Dr. Zhang, Assistant Professor

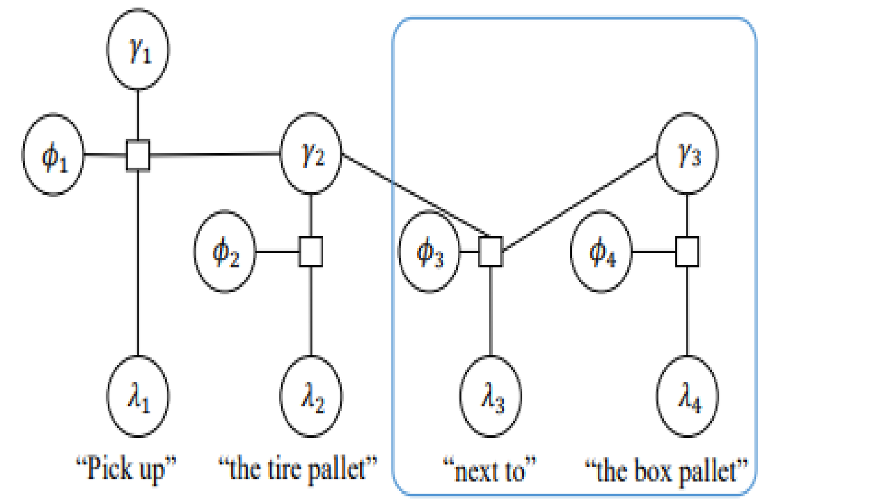
Arizona State University

Introduction

The current issue of the communication between humans and robots is how words are grounded, or derive their meanings from a robot's perspective. Grounding frameworks are heavily dependent in the environment a robot is in, and as a result, the natural language process must take in multiple contextual definition for words in order to have effective communication. This project aims to create an environment where it takes in both

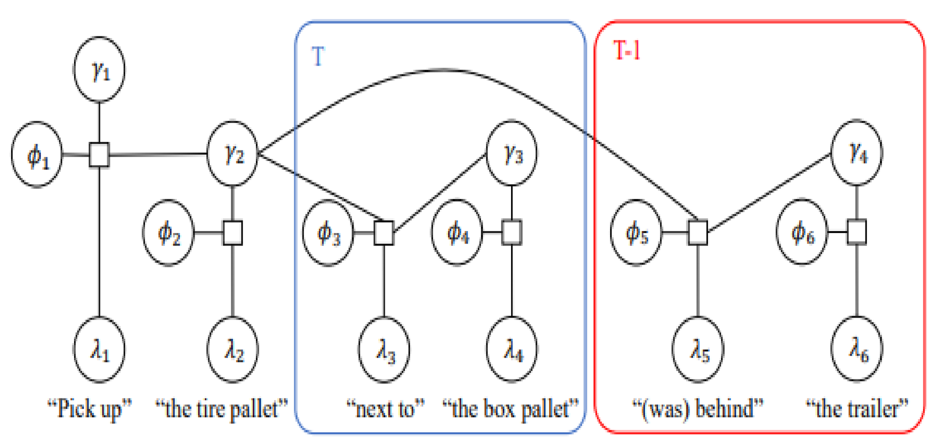


An example of referring objects based on what time state they are in. (a) demonstrates an object that has been loaded/unloaded by the forklift before (b).



γ_i represents groundings in the environment. λ_i indicates constituent phrases from natural language commands. ϕ_i is a correspondence factor indicating the correctness of the mapping between groundings and constituent phrases.

This figure depicts the G^3 framework where it considers the spatial context of one temporal state.

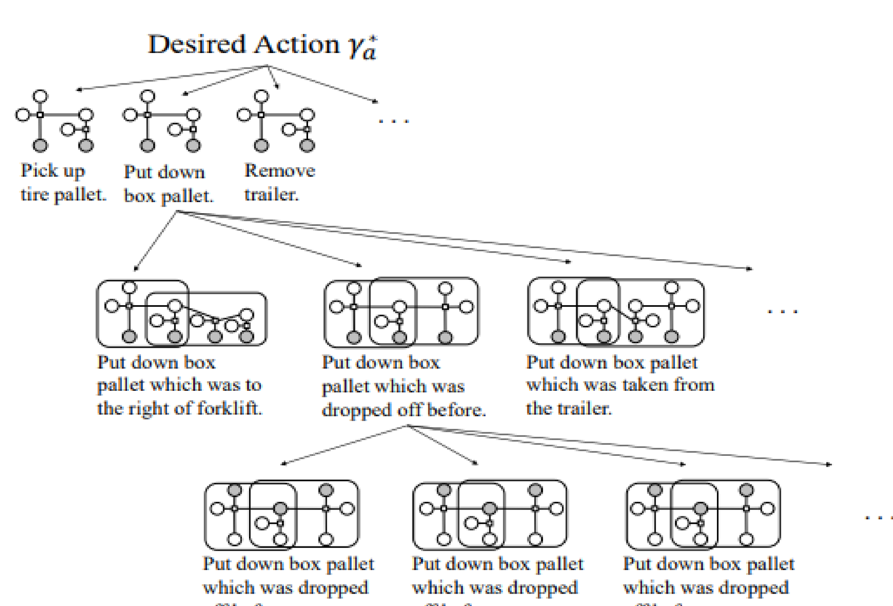


The G^3 framework where it considers the spatial context of one temporal state (present) and the prior temporal state (past).

Technical Approach

1. Mapping natural language phrases and the groundings in the environment
2. Transitioning inverse semantics framework into temporal spatial inverse semantics
3. Training and testing temporal spatial inverse semantic framework on the experiment domain

$S \rightarrow VB\ NP$
$S \rightarrow VB\ NP\ PP$
$PP \rightarrow IN\ NP$
$NP \rightarrow NP\ VBD\ PP$
$NP \rightarrow NP\ VBD\ VB\ PP$
$VB \rightarrow$ pick up | load | unload | drive forward
$NP \rightarrow$ the trailer | the forklift | the tire pallet
the box pallet | the trailer | you
$IN \rightarrow$ on | to the left of | to the right of
behind | in front of | from | to
$VBD \rightarrow$ was | were



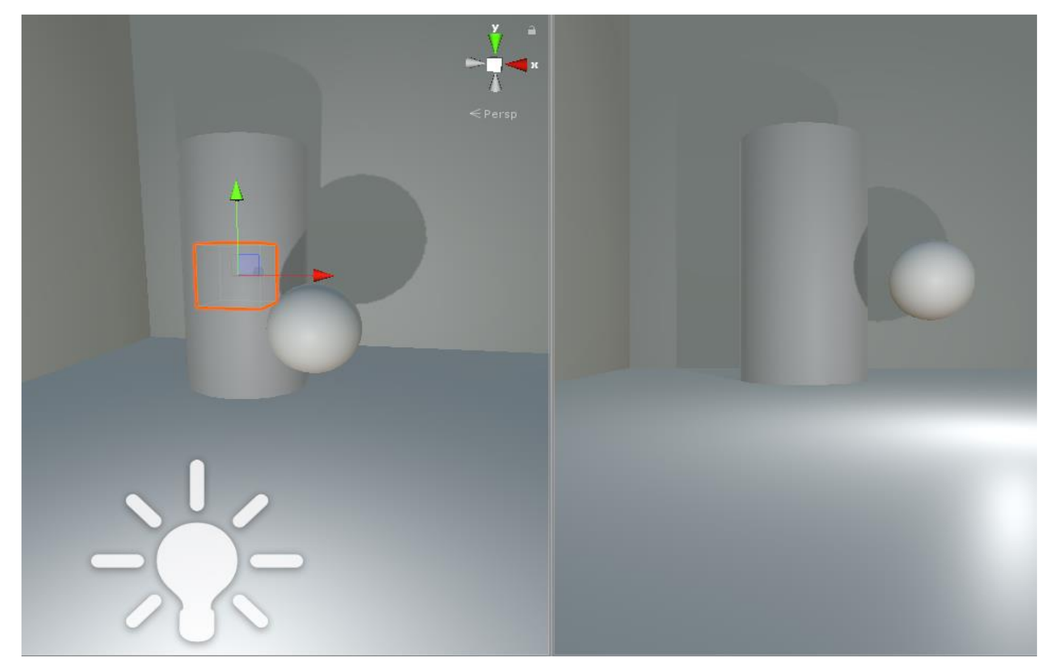
Results

- Setting up TeSIS (temporal spatial inverse semantics) code.

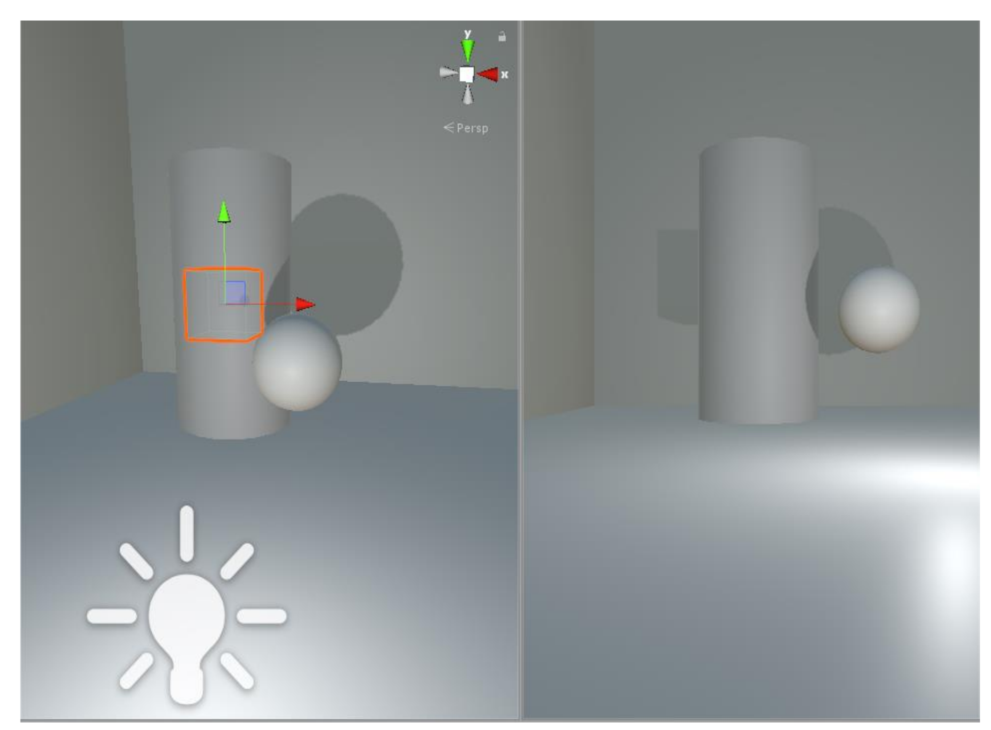
Replicated environment to run simulation on local machine. TeSIS code used a spatial language understanding (SLU) framework that required various packages including the SLU Core and Corpora.

- Debugging Unity Engine

Constructed an environment on Unity in order to simulate the goals of multi-modal communication. The limitations and boundaries of the Unity framework in regards to shadows turned out to be a product of the limitations of version 2018.4.17f1. In this version, the Unity engine does not allow for casting real-time shadows when you turn the object's shadow off, even when there is an object in front of it. This is due to how Unity treats an object with shadow casting off as a transparent object where light can pass through it.



Cube (highlighted orange) is located inside the cylinder and casts no shadow behind it. The behavior of shadows in all objects in this picture are behaving properly



When certain elements have the casting shadow off (in this case the cylinder), Unity turns off shadows by allowing light to pass through the object. As a result, the cube's shadow is now being shown on the right. Even though the cube is completely inside the cylinder and should not be casting a shadow onto the back wall, the cylinder is treated as a transparent object where light still manages to pass through and cast the cube's shadow.

Anticipated Outcomes and Impact

Following the completion of this project, we expect to find new methods for human-robot communication, create an integrated platform for testing, and develop additional research papers and results. The impact of this project is that it will increase the adaptability of a robot in its environment and have improved communication in a human-robot workspace using VR/AR technology. Additionally, this code and dataset will be available on Github so that other researchers and developers may apply this study to their own.

Future Research and Work

Because mixed-reality technologies show promise for more effective ways of providing information to a human and predicting what the robot must do to execute its planning, the need for research in applying such mixed-technologies to human-robot workspaces is a must. Thus, future research and work is needed to incorporate the theoretical simulations of the temporal spatial inverse semantic framework into hardware machines such as virtual and augmented reality devices. Utilizing VR/AR technology will allow us to address more complex interactions and to visualize a virtual agent's internal decision-making processes.

References

[1] Chakraborti, T., Sreedharan, S., Kulkarni, A., & Kambhampati, S., "Projection-Aware Task Planning and Execution for Human-in-the-Loop Operation of Robots in a Mixed-Reality Workspace," in IROS, 2018.

[2] Gong Z., & Zhang Y., "Temporal spatial inverse semantics for robots communicating with humans," in ICRA, 2018.

[3] Gong Z., & Y. Zhang., "Behavior Explanation as Intention Signaling in Human-Robot Teaming," in ICRA, 2018.

[4] Kulkarni, A., Zha, Y., Chakrabarti, T., Vadlamudi, S., Zhang, Y. & Kambhampati, S., "Explicability as Minimizing Distance from Expected Behavior," in AAMS, 2019.

[5] S. Sato and S. Sakane, "A human-robot interface using an interactive hand pointer that projects a mark in the real work space," in ICRA, vol. 1. IEEE, 2000, pp. 589–595.

Acknowledgments

Thank you to Yu Zhang , Ze Gong, and Andrew Boateng for extensively helping me throughout this project.