Rheinische
Friedrich-Wilhelms-
Universität Bonn
Prof. Dr. Maren Bennewitz

Institut für Informatik
Abteilung VI
Humanoid Robots Lab
Adresse:
Friedrich-Hirzebruch-Allee 8
53115 Bonn

Humanoid Robotics

Assignment 8

Due Thursday, June 26th, before lecture.

Perception for Grasping, Pushing, Interactive Perception and Foundation Models:

1. Which learning based grasp model for which scenario (Total: 3 points)

In the lecture we discussed **VGN**, **AnyGrasp** and **GPD** as three examples of learning-based grasping model for parallel jaw gripper. For each of the following scenarios, choose the most suitable model. Justify your choice based on what is presented in the lecture slides: consider factors such as sensor modality, real-time capabilities, object specificity, and task goals.

a. A mobile manipulator with a wrist-mounted RGB-D camera approaches an object that is moving on a conveyor belt. The robot must track and refine the grasp online as it aligns with the object.
b. A mobile manipulator with a wrist mounted RGB camera that is used for semantic segmentation for generating selective object masks, and LiDAR sensor that provides point clouds navigates home environments to grasp and bring specific objects to the user. The user may change the task mid-way and the robot must adapt and change its target location to grasp the new object.
c. A 7 dof manipulator arm with a depth only sensor is fixed in front of a bin with cluttered objects of the same type in a factory. The goal is to clear the bin by removing the object from the bin and placing it on a fixture as fast as possible.

2. Target Driven Action Selection (Total: 3 points)

The user has instructed via an LLM based agent to a general-purpose service robot (GPSR) equipped with a head mounted RGB-D camera and a 2-finger gripper to retrieve a box of soup can from the shelf quickly and bring it to the kitchen so that they can cook. The GPSR perceives part of a red soup can behind several occluding items on a pantry shelf. It maintains a voxel-based semantic belief map and estimates.

The robot can perform one of the following actions:

- **Reposition Whole Body (Can be for active viewing, better pushing or better grasping)**
- **Push object**
- **Grasp object**

The robot can also perform the action **Reposition Head** in conjunction with reposition body, push and grasp actions as well. The different actions have the following typical information gain relationship:

IGVb ~ IGMb >> IGMh >> IGVh

- **IGVb** (Information Gain from Reposition Whole Body and View, action length: 1)
- **IGMb** (Information Gain from Pushing an Occluding Box and Reposition Whole Body and View, action length: 2)

- **IGMh** (Information Gain from Pushing an Occluding Box and Reposition Head and View, action length: 1)
- **IGVh** (Information Gain from Repositioning Head for Better View, action length, 1 when independently done but can also be combined with push and grasp action)

Consider three different scenarios where robot has an action sequence of length 1, 2, 3 and 4. How would you sequence the actions for the task of object retrieval? Please note the entire task is considered a **failure** if the robot fails to **grasp** the object. Explain the difference in choices if the task goal was **semantic mapping** of the shelf.

3. Foundation Models (Total: 2 points)

For each task below, the robot fails despite using the listed model(s). Identify which one model is most likely responsible, and give one short reason based on lecture concepts.

- “Put the book on the top shelf” The robot lifts the book but drops it before placing it. (Models used: **LLM+VFM+VLA**)
- "Give me the beer mug." The robot hands over a white tea cup instead. (Models used: **VLM+VLA**)
- "The blue basket has oranges and apples freshly harvested from the glasshouse. Sort all good apples into the red basket." The robot moves all apples into the red basket, including bad apples. (Models used: **LLM** + **VFM+ VLA**)
- "Place the egg into the carton." The robot crushes the egg during grasping. (Models used: **LMM+VFM+VLA**)

4. Analytical Planar Pushing (Total: 7 points)

In this assignment, you will implement and test the analytical model for planar pushing based on the quasi-static framework introduced by Kloss et al. (2017). You will compute motion cones, detect sticking vs. sliding contacts, and predict object motion based on pusher velocity. The assignment also includes a small visualization component to better understand pushing interactions.

a. Compute the friction cone boundaries given a surface normal and coefficient of friction. (1 point)

b. Compute the torque exerted around the object center given contact point and boundary forces. (1 point)

c. Compute the motion cone directions using friction cone and torque values. (1 point)

d. Classify whether a given pusher velocity results in sticking or sliding contact. (1 point)

e. Compute the resulting object velocity from a pushing action under the correct contact condition. (1 point)

f. Run and explain two test scenarios: (2 points)

i. Push at the center vs. push at an offset: how does torque affect motion?

ii. Show and explain a sliding vs. sticking contact case.