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

Discussed in Tutorial on 07.07.2026.

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

1. Visual Servoing

a. In many manipulation pipelines, the robot first detects the object and then executes the planned motion without further visual correction. Explain why this open-loop approach can fail in practice. Give at least four concrete reasons.
b. Visual servoing compares the current visual state with a desired visual state using the error

$$e(t) = s(t) - s^*$$

Explain what s(t), s*, and e(t) represent.

Give at least three examples of visual features that could be used.

c. Compare image-based visual servoing (IBVS) and position-based visual servoing (PBVS). Where is the error defined in each case, and what is the main practical difference between the two approaches?
d. Visual servoing relies on continuous visual feedback. Name and briefly explain at least three possible failure cases of visual servoing.
e. Could visual servoing be applied to moving targets? If yes, explain why continuous visual feedback can help, and mention at least one limitation.

2. Target Driven Action Selection

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

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**)