UNIVERSITÄT BONN
RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN
HUMANOID ROBOTS LAB
Prof. Dr. Maren Bennewitz
Postanschrift:
Friedrich-Hirzebruch-Alle 8, 53115 Bonn

Humanoid Robotics
Discussion in Tutorial on 12.05.2026.
Important: Solution should be **type-written** and handed in as pdf, handwritten solutions are **not** accepted.

Assignment 3 (Graded) (Total Points: 25)

1. Transformations and Projection (4 points)

A 3D point in the world frame is

$$P_w = (2.0,\ -1.0,\ 4.0) \text{ m}.$$

A camera has intrinsics:

$$f_x = 200, \quad f_y = 150, \quad c_x = 160, \quad c_y = 120.$$

The camera pose in the world frame is given by:

- Rotation: $R = R_z(90°)$
- Translation: $t_{cw} = (0,\ 1,\ 0)$ m

(a) Compute the coordinates of P_w in the camera frame. (1 point)

(b) Project the point onto the image plane and compute pixel coordinates (u, v). (1 point)

(c) Explain briefly under what condition a 3D point becomes non-observable in the pinhole camera model. (1 point)

(d) If the point lies behind the camera, propose one modification (pose or intrinsics) to make it observable. (1 point)

2. Depth Measurement and Projective Projection (2 points)

Humanoids H1 and H2 observe the same household scene containing a nearby cup on a table and a farther cabinet in the background. H1 uses its time-of-flight camera C1, while H2 uses its stereo camera C2.

(a) Briefly explain how C1 and C2 obtain depth information differently. (1 point)

(b) In C2's camera image, the parallel edges of the table appear to converge, and the farther cabinet appears smaller than the nearby cup. Explain briefly whether this camera is damaged or this is expected. (1 point)

3. 3D World Representation – TSDF and Mapping (9 points)

A voxel stores a signed distance value D and an accumulated weight w. For this exercise, use the simplified confidence model $w_i = 1/z_i$. The TSDF update rule is

$$D_{\text{new}} = \frac{w_{\text{old}} D_{\text{old}} + w_i d_i}{w_{\text{old}} + w_i}, \qquad w_{\text{new}} = w_{\text{old}} + w_i.$$

Initial state:

$$D_0 = 0, \qquad w_0 = 0.$$

Measurements:

i	z_i (m)	d_i (m)
1	1	+0.5
2	2	+0.2

(a) Compute the TSDF value and weight after integrating the first two measurements. (2 points)

(b) Interpret shortly the physical meaning of the sign of a TSDF value. What does $D(x) > 0$, $D(x) = 0$, and $D(x) < 0$ indicate? (1 point)

(c) Compare TSDF and occupancy grid representations with respect to

- surface reconstruction,
- free-space reasoning.

Provide two key differences. (2 points)

(d) Explain briefly why TSDF maps can represent surfaces with sub-voxel accuracy, while occupancy grids usually cannot. (1 point)

(e) The following TSDF values are stored for voxel centers along one camera ray after fusion. The voxel centers are ordered from the camera outward.

x (m)	$D(x)$ (m)	w
0.00	+0.08	2
0.05	+0.02	3
0.10	−0.03	4
0.15	−0.07	2
0.20	+0.00	0

Assume that a voxel with $w = 0$ is unknown. Assume that a surface or occupied voxel is detected if $w > 0$ and either $|D(x)| \le \epsilon$ (with $\epsilon = 0.03\,\text{m}$) or a sign change occurs between neighboring observed voxels. Estimate the surface position using linear interpolation between the two neighboring observed voxels where the sign changes. Classify each voxel as observed free space, surface/occupied, inside/behind surface, or unknown. Briefly explain why this conversion from TSDF values to occupancy labels is only a heuristic. (3 points)

4. Active Perception and Planning (10 points)

Consider a robot exploring a 2D occupancy grid. Each cell has an occupancy probability p. Unknown cells have $p = 0.5$, free cells have $p = 0$, and occupied cells have $p = 1$. Use the binary Shannon entropy formulation to compute information gain.

Assume that observing a cell completely resolves its occupancy state. The information gain I of a viewpoint is the entropy reduction of the cells observed from that viewpoint. A cell that has already been observed from a previous viewpoint should not be counted again.

The robot starts at position

$$R = (1, 1).$$

There are four candidate viewpoints:

$$V_1 = (2, 2), \qquad V_2 = (3, 2), \qquad V_3 = (9, 9), \qquad V_4 = (9, 10).$$

Assume:

- Motion cost is the **Manhattan distance**
- The utility of visiting a viewpoint is

$$U = \frac{I}{C},$$

  where I is the information gain from that visit and C is the motion cost required to reach that viewpoint from the current robot position.
- The information gain of a viewpoint depends on the order in which viewpoints are visited, because different viewpoints may observe overlapping cells.

The cells visible from each viewpoint are given below.

Viewpoint	Visible unknown cells	Visible free cells	Visible occupied cells
V_1	4	8	3
V_2	5	7	2
V_3	31	5	6
V_4	30	4	5

Some cells are visible from more than one viewpoint. The pairwise overlaps between viewpoints are given below. These overlaps are symmetric, so each pair is listed only once.

Overlapping viewpoints	Overlapping unknown cells	Overlapping free cells	Overlapping occupied cells
V_1 and V_2	1	3	1
V_1 and V_3	1	2	1
V_1 and V_4	0	1	1
V_2 and V_3	2	2	1
V_2 and V_4	1	1	1
V_3 and V_4	10	2	3

There is no cell that is visible from more than two viewpoints.

(a) Using binary Shannon entropy, compute the information gain of each viewpoint when selected as the first view. Determine the next-best-view under a pure information-gain criterion. (2 points)

(b) Compute the motion cost from R to each viewpoint using Manhattan distance. Based on your answer to part (a), state the motion cost incurred by the pure information-gain choice. (1 points)

(c) Compute the utility $U = I/C$ for each viewpoint when chosen directly from R. Determine the utility-based next-best-view. (2 points)

(d) Starting from the utility-based next-best-view from part (c), choose one additional viewpoint greedily. That is, at the second step, choose the remaining viewpoint with the highest immediate utility $U = I/C$ from the robot's current position. Use the overlap information to avoid counting already observed cells again. Compute the resulting two-view sequence, its total information gain, its total motion cost, and its sequence utility

$$U_{\text{seq}} = \frac{I_{\text{total}}}{C_{\text{total}}}.$$

(2 points)

(e) Now consider all possible ordered two-view sequences that start at R and visit exactly two distinct viewpoints from $\{V_1, V_2, V_3, V_4\}$. For each sequence, compute:

- the total information gain,
- the total motion cost,
- the sequence utility

$$U_{\text{seq}} = \frac{I_{\text{total}}}{C_{\text{total}}}.$$

Determine the optimal two-view sequence under sequence utility. Compare it with the greedy sequence from part (d). (3 points)