KIT-Department of Informatics
Prof. Dr.-Ing. Tamim Asfour

Exam Solution Sheet

Robotics III - Sensors and Perception in Robotics

September 2, 2021, 11:00 – 12:00

Family name:	Given name:	Matriculation number:
Bond	James	007

Exercise 1	10 out of 10 points
Exercise 2	9 out of 9 points
Exercise 3	8 out of 8 points
Exercise 4	10 out of 10 points
Exercise 5	8 out of 8 points

Total:	45 out of 45 points

	Grade: **1,0**

Exercise 1 *Internal Sensors* (10 points)

1. Internal and external sensors: Internal (proprioceptive) Sensors: Measure the internal state of the robot. [1 p.]
   Examples: Gyroscope, Accelerometer, IMU, Joint angle encoder, Joint torques, Temperature of robot components,...
   External (exteroceptive) Sensors: Gain information about the environment of the robot.
   Examples: Cameras, Distance sensors, Radar, tactile sensors, position sensors.

2. Formula and computation for R_2 [2 p.]

   - Formula: $R_2 = \frac{R_1 \cdot U_2}{U_1}$
   - Value: $R_2 = \frac{50\Omega \cdot 4V}{2V} = 100\Omega$

3. Advantage Wheatstone bridge: [1 p.]
   The differential measurement of voltage in combination with an amplifier allows to use the entire range of the ADC.

4. Encoder

   (a) Angular resolution: An encoder disk with three rings has $2^3 = 8$ distinct states so the angular resolution is $\frac{360°}{8} = 45°$. [1 p.]

   (b) Gray encoding: On a disk with Gray-encoding, only one track changes state at a time. This eliminates the problem of erroneous transient states. [1 p.]

5. Inertial measurement units

   (a) To infer the angular position, the angular rate needs to be numerically integrated, which induces drift over time that can not be compensated. [1 p.]

   (b) The gyroscope-based orientation estimate is high-pass filtered to eliminate the drift while the accelerometer estimate is low-pass filtered to eliminate errors induced by movements of the sensor. Both values are then fused together. [2 p.]

   (c) Rotation around the gravity vector is only picked up by the gyroscope which drifts. The magnetometer can directly infer the rotation around this axis. [1 p.]

Exercise 2 *Tactile and Visual Sensing* (9 points)

1. Tactile Sensors:
   - Resistive
   - Capacitive
   - Optical
   - Magnetic (Hall-effect)
   - Pressure: barometer as MEMS sensor
   - Sensing orientation: IMU

2. *Deformable material(gel)* in front of *camera*, illuminated by from below by *different light colors*

   Measurement principle:

   -Camera captures RGB image of the gel from below

   -Calculate deformations from reflected RGB light

3. (a)

$$z = \frac{b \cdot f}{d} \Rightarrow d = \frac{b \cdot f}{z}$$

$$d = \frac{0.15\,\mathrm{m} \cdot 4\,\mathrm{mm}}{0.75\,\mathrm{m}} = 0.8\,\mathrm{mm}$$

   (b) Advantage: Higher depth resolution

   Disadvantage: Not suited for objects close to the camera

4. One camera provides the RGB image, the other one captures the projected pattern (infrared) for depth sensing. One camera captures visible light, the other one infrared.

   A further required component is the IR-pattern projector.

Exercise 3 *Feature Extraction* (8 points)

1. Correlation Functions: 2 p.

   (a) Normalization:
   Invariant with respect to constant additive or multiplicative brightness differences

   (b) Applications:
   - Solving of correspondence problem in stereo vision
   - Object recognition
   - Image-based localization
   - SIFT (DoG)
   - Moravec Operator / Corner Detection
   - Pattern Matching

2. Harris Corner Detector:

   (a) Approximation: 1 p.
   The image function is approximated by a Taylor Series Expansion

   (b) Image Regions: 1 p.
   Harris & Stephens proposed the corner response measure:

$$C(u, v) = \det(M) - \kappa \cdot \text{trace}(M)^2$$

   No eigenvalue decomposition is needed, only determinant and trace.

3. Features: 1 p.
   A Feature Detector only extracts the position of the feature on the image, while a descriptor extracts a vector of properties for a (known) feature.

4. SIFT: 1 p.
   - Object recognition/detection
   - Motion tracking
   - Stereo calibration
   - Image indexing and retrieval
   - Robot navigation
   - Gaze Stabilization
   - Solving the Correspondence Problem
   - ...

5. Pose Estimation:

   (a) Name: 1 p.
   Perspective n-Point (PNP)

   (b) Advantage: 1 p.
   - Robust, since stereo triangulation is used

- Better accuracy (especially depth), depending on setup

Disadvantage:

- Stereo calibration is needed
- Inaccuracy with strong lens distortion
- Correspondence Problem needs to be solved+
- Limited Distance
- Needs Good Lightning

Exercise 4 *Scene Understanding* (10 points)

1. Scene Representations

   (a) Difference: **1 p.**
   Object classification gives one label for the whole image, while object segmentation gives pixel-wise labels.

   (b) Approaches: **1 p.**
   - Sliding window
   - Region-proposals / Region-proposal networks

   (c) Reason: **1 p.**
   A network architecture for images usually expects a fixed regular grid of values (pixels), while a point cloud usually is an unordered set of values (points) .

   (d) Difference and examples: **1 p.**
   - Static spatial relations only depend on the *current* object poses, while dynamic spatial relations also depend on their *past* poses .
   - Examples for static spatial relations:
     above, below, right of, left of, behind, in front of, contained/inside, touching/in contact
   - Examples for dynamic spatial relations:
     moving together, getting closer / closer to, moving apart / further from, on other side of

2. Graph Networks

   (a) Types: **1 p.**
   - Learned: The update functions can be learned from data.
   - Purpose: (of *update* functions) Computing the updated node/edge/global attributes.
     (Note: The purpose of *aggregation* functions is reducing a variable number of attributes to a single element.)

   (b) Properties: **2 p.**
   Aggregation functions must be
   - order-invariant (i. e. invariant to the order of its inputs) and
   - take an arbitrary amount of elements .

   This is necessary so graph network blocks can be applied to graphs with arbitrary and various number of nodes and edges.

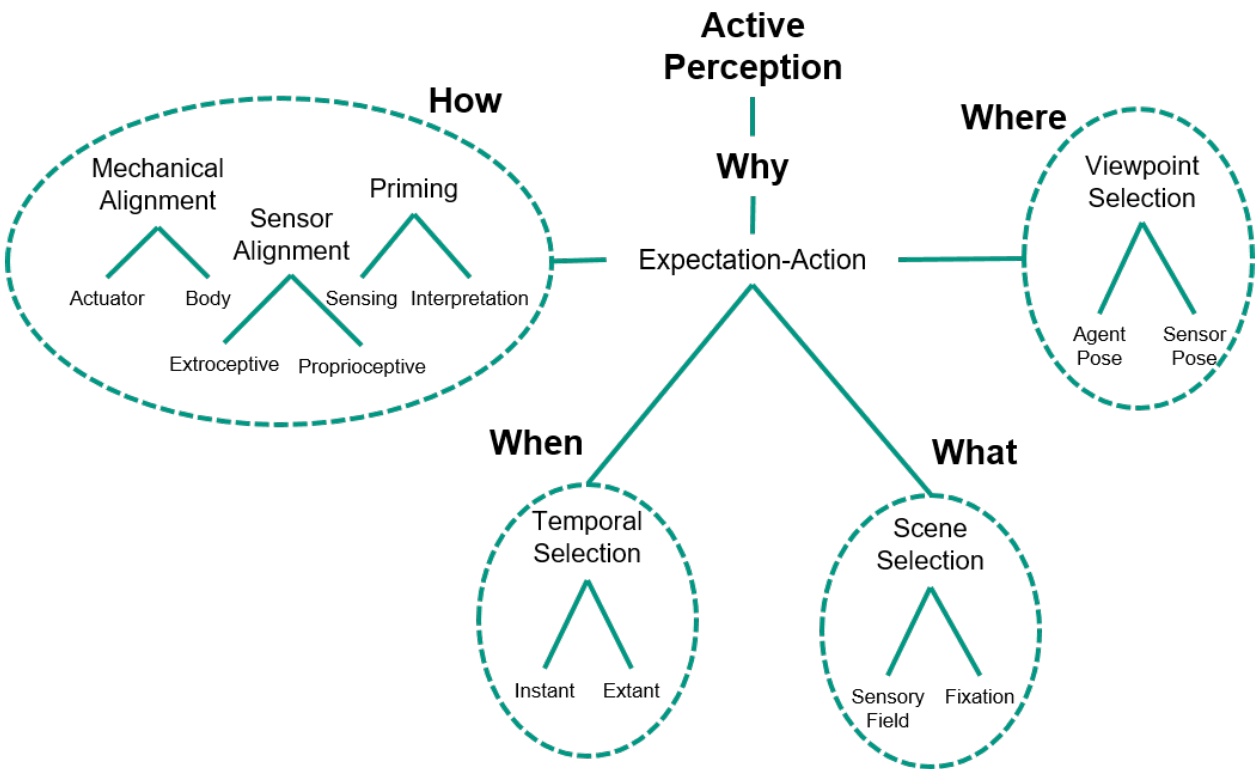
   (c) Functions: **3 p.**
   - Edge update function Φ^e: takes $\mathbf{e}_k,\ \mathbf{v}'_i,\ \mathbf{v}_{s_k},\ \mathbf{u}$
   - Node update function Φ^v: takes $\mathbf{v}_i,\ \mathbf{u},\ \bar{\mathbf{e}}'_i$
   - Global update function Φ^u: takes $\mathbf{u},\ \bar{\mathbf{v}}',\ \bar{\mathbf{e}}'$

Exercise 5 *Active Vision and Gaze Stabilization* (8 points)

1. Yarbus key observation: [1 p.]

   - Patterns of eye movements are similar (but not identical) when different people view the same painting or when a single individual was shown the same painting a number of times, with between one and two days separating the recording
   - Different pattern of eye movements depending on the task

2. Active perception: [2.0 p.]



3. Gaze stabilization [3.0 p.]

   (a)
   - The goal of the *Vestibulo Ocular Reflex* is to generate eye velocities to counteract head movements.
   - Inertial Measurement Unit (IMU)
   - control law:

   $$\dot{q}_{eyes} = -k_{vor} \cdot \begin{bmatrix} \omega_{yaw} \\ \omega_{pitch} \end{bmatrix}$$

   - q_{eyes} velocities for the eye joints
   - ω rotational velocity of the head
   - k_{vor} control parameter

   (b) The OKR compensates the retinal slip and thus is able to counteract any kind of perturbation (internal, external) while the VOR can counteract only internal perturbations.

4. Reafference 2 p.

- Terms and Relations:
  - afference = measured sensor values
  - reafference = predicted sensor values
  - exafference = afference - reafference
- Example: acceleration values of an IMU without self-induced motion.