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

Exam Solution Sheet

Robotics III - Sensors and Perception in Robotics

February 9, 2023, 17:30 – 18:30

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

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

Total:	45 out of 45 points

	Grade: **1,0**

Exercise 1 *Internal Sensors* (9 points)

1. Absolute Encoder

   (a) Transient States: 110, 101, 011, 100, 010, 001; Three possibilities for the first bit to flip, then 2 possibilities each for the second bit to flip before the last bit flips.

   (b) Alternative Method: Gray encoding. For two consecutive bit patterns on the ring always exactly one bit changes

   (c) Resolution: The encoder can assume $2^3 = 8$ different states. So the resolution is $360/8 = 45$ degrees.

2. Torque Measurement: First method: analog strain gauges are glued on a hollow shaft or spoke wheel. When the spokes/shaft deform, the resistance in the strain gauges changes and can be correlated to the torque. Second method: use a high-precision absolute encoder and a hollow shaft to measure the torsion of the shaft.

3. IMU: Components: 3-axis accelerometer, 3-axis gyroscope. The accelerometer can directly measure the orientation relative to the gravity vector but is susceptible to short term noise from linear acceleration of the IMU. The gyroscope measures the rate of change in angular position. Integration of this signal yields the angular position, but causes drift due to measurement errors being integrated. A complementary filter can be used to combine the high-pass-filtered gyroscope estimation and low-pass-filtered accelerometer estimation.

Exercise 2 *External Sensors* (9 points)

1. Difference:
   RADAR uses radio-waves, LiDAR uses light waves

2. LiDAR

   (a) Received signal

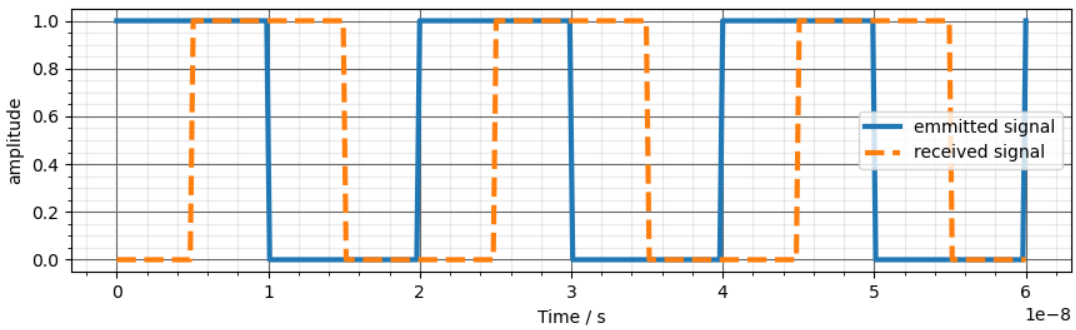

Figure 1: Received signal

   (b) Distance d:
   $d = \frac{1}{2} c \cdot t$

   $d = \frac{1}{2} \cdot 10 \times 10^{7}\,\mathrm{m\,s^{-1}} \cdot 5 \times 10^{-9}\,\mathrm{s} = 0.25\,\mathrm{m}$

   (c) Aliasing Effect
   Distance calculations will lead to ambiguous results if the light travels back and forth during a longer period of time than the period of the modulated light

3. Frame rate
   $f_{frame} = f/N = 5000\,\mathrm{Mbit\,s^{-1}}/1000 \cdot 333 \cdot 3 \cdot 10\,\mathrm{bit} = 500\,\mathrm{s^{-1}}$

4. (a) Projection schemes
      - Single light beam
      - Laser line (light section)
      - Projection of an encoded pattern (structured light)

   (b) Pattern decoding:
   The process of matching an image region with its corresponding pattern region is known as *pattern decoding* $\rightarrow$ similar to searching correspondences

   (c) Error sources
      - Variations in the propagation speed
      - Uncertainties in determining the exact time of arrival of the reflected pulse
      - Inaccuracies in the timing circuit used to measure the round-trip time of flight
      - Interaction of the incident wave with the target surface.

Exercise 3 *Feature Extraction* (9 points)

1. Points of Interest

   (a) Moravec Operator: 2 p.
   Definition:

$$
\begin{aligned}
D(u,v,s,t) &= \sum_{(u_i,v_j)\in W(u,v)} \left(I(u_i+s,v_j+t) - I(u_i,v_j)\right)^2 \\
D(2,2,1,0) &= 12 \\
D(2,2,0,1) &= 6 \\
D(2,2,1,1) &= 14 \\
D(2,2,-1,1) &= 11
\end{aligned}
$$

   (b) Image Structure Tensor: 2 p.
   Definition:

$$
M(u,v) = \begin{pmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{pmatrix}
$$

$$
M(u,v) = \begin{pmatrix} 40 & 0 \\ 0 & 0 \end{pmatrix}
$$

   (c) Calculation of Image Gradients: 1 p.
   The image gradients I_x and I_y can be calculated using 2D convolutions with Prewitt and Sobel filters.

2. Scale Invariant Feature Transform

   (a) SIFT algorithm: 2 p.

   (I) Scale-space extreme detection
   (II) Keypoint localization
   (III) Orientation assignment
   (IV) Generation of keypoint descriptors

   (b) 6D Pose Estimation: 2 p.
   Extract SIFT features for known object and store them in a database . For the estimation of the pose of a known object, identify correspondences between the image features and the stored objects features and filter them (e.g. RANSAC). Determine the homography between matched features to calculate the pose of the object.

Exercise 4 *Scene Understanding* (9 points)

1. With a Region Proposal Network (RPN), an object *detection* method (which classifies whether an object exists in an image) can be applied to ... (one of the following is sufficient) 1 p.
   - detect *multiple* objects in an image
   - *localize* objects in an image
2. Outputs: 1 p.
   - Semantic segmentation: Class (or equivalent) per pixel
   - Instance segmentation: Instance ID per pixel
3. Machine Learning Approaches
   (a) Benefit: Point clouds provide explicit 3D information. For example, an object overlaps others in an image but there are large gaps between them in the 3D point cloud. 1 p.
   (b) Reasons: There is no clear order between points in a point cloud or nodes/edges in a graph, and the number of points or nodes/edges is variable. (Referring to either point clouds or scene graphs is sufficient.) 2 p.
   (c) Solution: These methods leverage operations that are invariant of the order of inputs and can be applied to any number of elements , such as min, max, sum, or average. 1 p.
4. Object Relations
   (a) Difference: Spatial relations encode the relative positions/locations of objects, while support relations encode which objects physically support each other / cause other objects to fall when removed. 1 p.
   (b) Difference: Static relations consider only one frame / point in time, while dynamic relations consider changes over time / multiple frames / multiple points in time. 1 p.
   (c) Task: Action / activity recognition 1 p.

Exercise 5 *Robot Vision* (9 points)

1. *Fill in the table using a checkmark (✓)* where appropriate
 2.5 p.

	classical computer vision	**active vision**	**active perception**
Multi-modal sensory input			✓
Changing the environment			✓
Image processing	✓	✓	✓
Changing agent's state		✓	✓
Viewpoint selection		✓	✓

2. Segmentation of Unknown Objects

   (a) Heuristics
 3 p.

heuristic	**object type**
Planes, cylinders and spheres amongst SIFT features (RANSAC)	textured objects
Unicolored regions of promising size: color MSERs (Maximally stable extremal regions)	single-colored objects
Visually salient regions (Difference of Gaussians filter)	objects that are neither textured nor unicolored

   (b) Match two point clouds A and B using ICP:
 1.5 p.

   1. For each $\mathbf{a} \in \mathbf{A}$ find closest point in $\mathbf{B}$
   2. Calculate transformation $\mathbf{T}$ that minimizes the mean squared distance of the correspondences
   3. Apply $\mathbf{T}$ to all $\mathbf{a} \in \mathbf{A}$

   Iterate until convergence

(c) Modification to ICP: 1 p.
Colored point cloud registration ("Colored ICP"): Use weighted cartesian + color distance in step 1 of ICP

(d) Reasons: 1 p.

i. complete segmentation
ii. more pushes reveal different sides → generate a multi-view descriptor