Last name	First name	Matriculation number

Mock Exam of Vision Algorithms for Mobile Robotics (UZH-DINF2039/ETH-151-0632-00L, HS21)

09.12.2021

The maximum number of points that you can get in the exam is 90.

Conventions

Please follow the conventions below:

- Pose transformations between frames A and B are denoted with rotation matrix and translation vector R_{AB} and t_{AB} such that the origin of B expressed in A is at t_{AB} and the (x, y, z) unit vectors of frame B expressed in frame A are the columns of R_{AB}.
- W denotes the world or global frame and C the camera frame.
- The camera looks in the positive z direction, x points to the right in the direction of the image width and y down in the direction of the image height.

1 Multiple Choice (7 P.)

You get $+0.5$ point for every correct answer, -0.5 point for every wrong answer, and 0 points for unanswered questions. The total sum of all points for this question is at least 0.
Mark the correct choice.

1. What calibration object is used in Zhang's method?
   (a) One planar grid
   (b) Two planar grids, which are perpendicular to each other
   (c) Three planar grids, which are perpendicular to each other and form a corner
   (d) 3D cube

2. For a calibrated camera, what is the minimum number of points required by the PnP algorithm to obtain one unique solution?
   (a) 6
   (b) 4
   (c) 3
   (d) 1

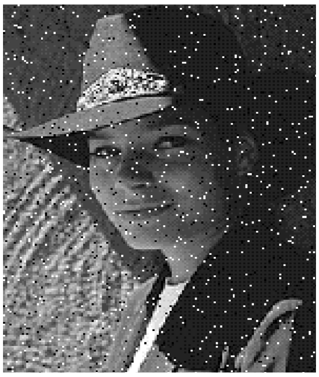
3. Mark the correct output c of the convolution between the 1D image a and the 1D filter b using zero padding. The output should have the same size as the input image.

$$a = [3, 1, 2] \qquad b = [2, 1, 2]$$

$$c = a * b$$

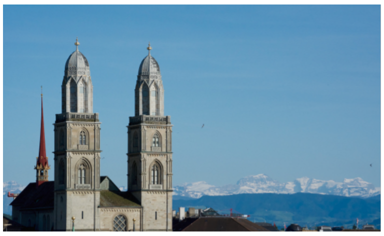
   (a) $c = [6, 5, 11]$
   (b) $c = [11, 5, 6]$
   (c) $c = [4, 11, 5]$
   (d) $c = [5, 11, 4]$

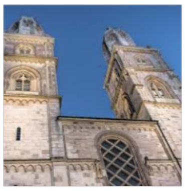
4. Which filter is best suitable to reduce the noise of the following image?



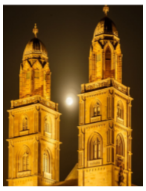
(a) Gaussian filter
(b) Mean filter
(c) Median filter
(d) Sobel filter

5. For the image shown below, which of the templates (a-d) would best detect the two towers using normalized cross correlation?

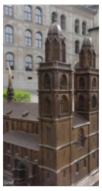




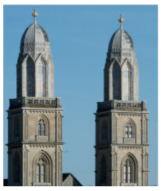
(a)



(b)



(c)



(d)

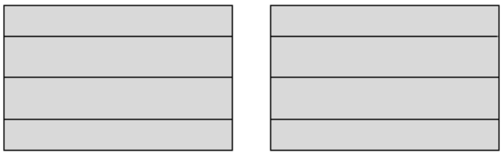
6. Which statement is correct? $\subseteq$ means "is a particular case of".
(a) VO $\subseteq$ SfM $\subseteq$ SLAM
(b) SfM $\subseteq$ SLAM $\subseteq$ VO
(c) VO $\subseteq$ SLAM $\subseteq$ SfM

7. Which of the following feature detectors is scale invariant?
(a) Harris
(b) SIFT
(c) Shi-Tomasi
(d) None of the above

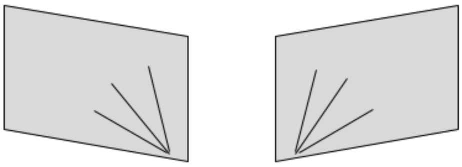
8. What would be the closest depth measured by a stereo camera with baseline b and focal length f and resolution of $W \times H$ (W = width and H = height)? Assume a simplified and rectified stereo setup in 2D, i.e., both cameras have identical intrinsic parameters and both image planes are coplanar and aligned with the baseline.
(a) $\frac{bf}{H}$
(b) $2W$
(c) $\frac{bf}{W}$
(d) $2H$

9. Which of the following sketches shows the epipolar lines corresponding to the setting of a monocular camera moving sideways on a straight rail parallel to the width axis?

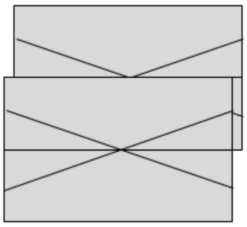
(a)



(b)



(c)



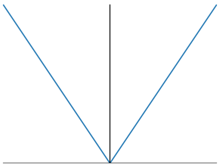
(d) None

10. What is the minimum number of correspondences required for general (i.e. unconstrained) structure from motion with a calibrated camera?
(a) 1
(b) 5
(c) 2
(d) 8

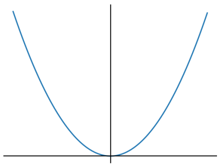
11. In which of the following cases can you recover the metric scale?
(a) Stereo
(b) Calibrated structure from motion
(c) Uncalibrated structure from motion
(d) None of the above

12. For the same set of inputs, RANSAC always provides the same result.
(a) True
(b) False

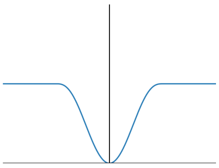
13. Which plot corresponds to the Tukey norm?



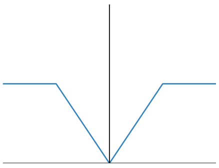
(e)



(f)



(g)



(h)

14. What is the minimum number of 2D point correspondences necessary to determine the 2D transformation between two images of the same planar object?
(a) 4
(b) 5
(c) 6
(d) 8
(e) None of the above

2 Application Question (20 P.)

Suppose that you want to write an algorithm that runs in real time on a smartphone to obtain a sparse and reasonably accurate 3D point cloud of a statue using only your phone onboard sensors. Thus, you take pictures in a way that the statue is not far from your phone and is always completely visible in the image. Assume a calibrated setup and a camera motion all around the statue.

1. Describe which sensors of the smartphone you would like to use and motivate their choice.
2. List the building blocks of the pipeline.
3. For each building block, detail the individual steps and design choices by keeping in mind that you want to obtain the most efficient pipeline possible.
4. Since the camera motion is around the statue, the images contain different parts of the statue. How do you deal with this problem?
5. List three possible failure modes of the algorithm and possible solutions.

3 Theory Questions (63 P.)

Please answer each question in detail.

1. (6 P.)What is the reprojection error and how is it used for refining the intrinsic calibration parameters of a camera? State also the formula of the reprojection error and name each variable in it.

2. a) (3 P.) Define the PnP problem for a calibrated camera by stating the known and unknown entities.
   b) (2 P.) How many solutions will the PnP algorithm give if you have one, two, three, and four 3D-2D correspondences?

3. Derive the Harris cornerness response function.
   a) (3 P.) Start with the problem formulation of the Harris corner detector and motivate the problem formulation.
   b) (7 P.) Based on the problem formulation, derive the cornerness response function R of the Harris detector.
   c) (2 P.) What are the differences in the cornerness response function of Harris and Shi-Tomasi?

4. (6 P.) Describe how the HOG descriptor can be constructed from an image patch.

5. (2P.) Derive the depth Z of a 3D point based on the corresponding image points u_l in the left camera and u_r in the right camera. Assume a simplified and rectified stereo setup in 2D, i.e., both cameras have identical calibration parameters and both image planes are coplanar and aligned with the baseline.
6. (5 P.) State and describe two possible ways to estimate depth with metric scale?
7. (3 P.) State and describe the relation between essential and fundamental matrix.
8. (3 P.) What is the benefit of RANSAC in comparison to an EM algorithm?
9. (2 P.) What is the probability of success p if n_{it} iterations are used for N datapoints containing n_{out} outlier datapoints and a minimum of s datapoints is required for estimating the model.
10. (7 P.) Provide two methods for local optimization in a VO-pipeline, write their cost functions and describe each term. Also state which one is more precise and how to reduce the computational complexity for the more complex optimization method.
11. Describe the Bag-Of-Words approach used for place recognition.
    a) (2 P.) What is a visual word? How to extract visual words from descriptors
    b) (2 P.) What is an image vocabulary? How to build it?
    c) (5 P.) How to perform image retrieval?

12. (3 P.) What is the IMU measurement model? Write the formula and describe the terms.