Exam in
02504 Computer Vision
Spring 2023

General information

- The exam consists of 21 questions. All questions have equal weight: the correct answer gives 1 point, incorrect or missing answer gives 0 points. You only need to submit the answers to the questions. You should not upload any notes or calculations.
- There is *one* correct answer for each question. Some of the numeric results have been rounded, and may deviate slightly from your result. This should not prevent you from being able to pick the correct answer.
- Each page contains one question. If there are illustrations and images, those refer to the question on that page.
- The notation in the questions is the same as used in the course slides.
- You can load files for a specific exercise using
  - `np.load("filename.npy", allow_pickle=True).item()`

Relevant links

Resources are available for a subset of questions in form of `.npy` files. Filenames typeset in `typewriter font` indicate that you can find files in the `materials` folder.

Question 1

A camera has focal length 1200, principal point (400, 350), $\alpha = 1$, $\beta = 0$, radial distortion parameters $k_3 = 0.01$ and $k_5 = 0.04$.

What is the camera matrix?

a) $\begin{bmatrix} 1200 & 0 & 400 \\ 0 & 1212 & 350 \\ 0 & 0 & 1 \end{bmatrix}$

b) $\begin{bmatrix} 1200 & 12 & 400 \\ 0 & 1200 & 350 \\ 0 & 0 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 1200 & 0 & 400 \\ 0 & 1200 & 350 \\ 0 & 0 & 1 \end{bmatrix}$

d) $\begin{bmatrix} 1200 & 48 & 400 \\ 0 & 1200 & 350 \\ 0 & 0 & 0 \end{bmatrix}$

e) $\begin{bmatrix} 1200 & 0 & 400 \\ 0 & 1200 & 350 \\ 0 & 0 & 0 \end{bmatrix}$

f) $\begin{bmatrix} 1200 & 0 & 400 \\ 0 & 1200 & 350 \\ 0 & 0 & 1200 \end{bmatrix}$

g) $\begin{bmatrix} 1200 & 0 & 400 \\ 0 & 1248 & 350 \\ 0 & 0 & 1 \end{bmatrix}$

Question 2

Consider a camera that takes pictures of resolution 800×600 (width$\times$height) that has the following camera matrix:

$$\boldsymbol{K} = \begin{bmatrix} 1000 & 0 & 400 \\ 0 & 1000 & 300 \\ 0 & 0 & 1 \end{bmatrix}$$

You bilinearly resize an image from this camera so it has size 400×300 (width$\times$height).
What is the camera matrix for the resized image?

a) $\begin{bmatrix} 500 & 0 & 200 \\ 0 & 500 & 150 \\ 0 & 0 & 1 \end{bmatrix}$

b) $\begin{bmatrix} 1000 & 0 & 800 \\ 0 & 1000 & 600 \\ 0 & 0 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 500 & 0 & 400 \\ 0 & 500 & 300 \\ 0 & 0 & 1 \end{bmatrix}$

d) $\begin{bmatrix} 2000 & 0 & 800 \\ 0 & 2000 & 600 \\ 0 & 0 & 1 \end{bmatrix}$

e) $\begin{bmatrix} 250 & 0 & 200 \\ 0 & 250 & 200 \\ 0 & 0 & 1 \end{bmatrix}$

f) $\begin{bmatrix} 1000 & 0 & 200 \\ 0 & 1000 & 150 \\ 0 & 0 & 1 \end{bmatrix}$

g) $\begin{bmatrix} 1000 & 0 & 400 \\ 0 & 1000 & 300 \\ 0 & 0 & 1 \end{bmatrix}$

h) $\begin{bmatrix} 2000 & 0 & 400 \\ 0 & 2000 & 300 \\ 0 & 0 & 1 \end{bmatrix}$

i) $\begin{bmatrix} 500 & 0 & 800 \\ 0 & 500 & 600 \\ 0 & 0 & 1 \end{bmatrix}$

j) $\begin{bmatrix} 2000 & 0 & 200 \\ 0 & 2000 & 150 \\ 0 & 0 & 1 \end{bmatrix}$

Question 3

A camera has focal length 1720, principal point $(680, 610.0)$, $\alpha = 1$, and $\beta = 0$. The camera has the rotation:

```
cv2.Rodrigues(np.array([-0.1, 0.1, -0.2]))[0]
```

and the translation:

```
np.array([[0.09], [0.05], [0.05]])
```

A 3D point in the world coordinate system has coordinates:

```
np.array([-0.03, 0.01, 0.59])
```

What is the projection of this point to the camera's image plane?

a) $[1033.72, 927.31]^{\mathrm{T}}$

b) $[592.69, 639.97]^{\mathrm{T}}$

c) $[592.54, 639.15]^{\mathrm{T}}$

d) $[650.65, 591.40]^{\mathrm{T}}$

e) $[592.59, 640.30]^{\mathrm{T}}$

f) $[349.60, 377.10]^{\mathrm{T}}$

g) $[1023.50, 930.30]^{\mathrm{T}}$

Question 4

Sometimes we can multiply a matrix by an arbitrary non-zero real number and the matrix is functionally equivalent to the matrix before we multiplied it with the scalar. That is you can multiply the matrix with, for example, 7 and it would still represent the same matrix for computer vision purposes.

Which matrices are *not* functionally equivalent after multiplication with an arbitrary real number?

a) Camera matrix, Essential matrix, Homography matrix, Projection matrix, and Rotation matrix

b) Camera matrix, Fundamental matrix, and Projection matrix

c) Camera matrix, Homography matrix, and Projection matrix

d) Camera matrix, Homography matrix, and Rotation matrix

e) Fundamental matrix and Homography matrix

f) Camera matrix

g) Fundamental matrix, Homography matrix, and Rotation matrix

h) Camera matrix, Projection matrix, and Rotation matrix

i) Rotation matrix

j) Camera matrix and Projection matrix

k) Camera matrix and Rotation matrix

l) Camera matrix, Essential matrix, Fundamental matrix, Homography matrix, Projection matrix, and Rotation matrix

Question 5

You are given three cameras (1, 2 and 3) that share the same camera matrix `K` and have the following extrinsics. You can copy and paste the following into Python:

```
K = np.array([[900, 0, 1070], [0, 900, 610.0], [0, 0, 1]], float)
R1 = cv2.Rodrigues(np.array([-1.6, 0.3, -2.1]))[0]
t1 = np.array([[0.0], [1.0], [3.0]], float)
R2 = cv2.Rodrigues(np.array([-0.4, -1.3, -1.6]))[0]
t2 = np.array([[0.0], [1.0], [6.0]], float)
R3 = cv2.Rodrigues(np.array([2.5, 1.7, -0.4]))[0]
t3 = np.array([[2.0], [-7.0], [25.0]], float)
```

You observe the same point in all three cameras, but with some noise. The observed points for cameras 1 to 3 are respectively:

```
p1 = np.array([[1046.0], [453.0]])
p2 = np.array([[1126.0], [671.0]])
p3 = np.array([[1165.0], [453.0]])
```

How far is `p2` from the epipolar line in camera 2 corresponding to `p1`?

a) 121.9

b) 104.4

c) 0.09316

d) 4.5475e-13

e) 29.05

f) 0.05686

g) 1.029e+27

h) 1.241e+03

i) 7.519e+06

j) 1.057e+03

k) 13.27

l) 0.5821

Question 6

Use all three observations of the point from the previous question to triangulate the point with the linear algorithm from the slides. Do not normalize the points.

What is the triangulated point?

a) $[4.0011, 1.2936, 1.6394]^{\mathrm{T}}$

b) $[3.4977, -0.0931, 0.6278]^{\mathrm{T}}$

c) $[1.3231, 1.1011, -0.0974]^{\mathrm{T}}$

d) $[3.1006, 0.7432, 0.4649]^{\mathrm{T}}$

e) $[1.2001, 1.1861, -0.1542]^{\mathrm{T}}$

f) $[1.0923, 1.1802, -0.1207]^{\mathrm{T}}$

g) $[1.1408, 1.1566, -0.0946]^{\mathrm{T}}$

h) $[1.1537, 0.9317, -0.2053]^{\mathrm{T}}$

Question 7

You now want to solve the triangulation from the previous exercise with non-linear triangulation. We can use Levenberg-Marquardt to do this, by minimizing the squared euclidean norm of a function. What is the best choice of this function for the task of non-linear triangulation?

a) A function $f := \mathbb{R}^4 \to \mathbb{R}^3$ that takes a homogeneous 3D point and returns the distance between the reprojected point and the observed point in each image.

b) A function $f := \mathbb{R}^3 \to \mathbb{R}^{3 \cdot 3}$ that takes a 3D point and returns the distance of the reprojection to all epipolar lines in each image.

c) A function $f := \mathbb{R}^3 \to \mathbb{R}^3$ that takes a 3D point and returns the distance between the reprojected and the observed point in each image.

d) A function $f := \mathbb{R}^3 \to \mathbb{R}^{3 \cdot 3}$ that takes a 3D point and returns the reprojected point in homogeneous coordinates minus the observed point in homogeneous coordinates in each image.

e) A function $f := \mathbb{R}^3 \to \mathbb{R}^{3 \cdot 2}$ that takes a 3D point and returns the reprojected point minus the observed point in each image.

f) Something else

Question 8

What is the minimum number of point correspondences you need to estimate an essential matrix?

a) 1

b) 2

c) 3

d) 4

e) 5

f) 6

g) 7

h) 8

i) 9

j) 10

k) 11

l) 12

Question 9

Which of the following statements about Zhang's algorithm for camera calibration is true?

a) Zhang's algorithm only works for flat calibration objects.

b) The calibration process iteratively adjusts the focus of the camera lens to improve image quality.

c) Zhang's algorithm can only be used to calibrate cameras with a small field of view, because it relies on accurately detecting corners.

d) The camera is calibrated using a series of homographies. Each homography imposes linear constraints on the fundamental matrix, which is then solved using SVD.

e) The camera intrinsics are obtained by decomposing the projection matrix $\boldsymbol{P}$ into the intrinsics and extrinsics.

f) Corners are detected in images of checkerboards which are then directly used to minimize the reprojection error.

g) Zhang's algorithm requires knowing the size of the checkerboard and the distance to it.

Question 10

When doing camera calibration we find an $\boldsymbol{R}$ and a $\boldsymbol{t}$ for each calibration object.

For the i^{th} calibration object, what do these describe?

a) The estimated $\boldsymbol{R}$ and $\boldsymbol{t}$ describe the coordinate transformation from the first calibration object to the i^{th} calibration object.

b) The estimated $\boldsymbol{R}$ and $\boldsymbol{t}$ describe the coordinate transformation from the i^{th} calibration object to the first calibration object.

c) The estimated $\boldsymbol{R}$ and $\boldsymbol{t}$ describe the coordinate transformation from the camera to the i^{th} calibration object.

d) The estimated $\boldsymbol{R}$ and $\boldsymbol{t}$ describe the coordinate transformation from the i^{th} calibration object to the camera.

Question 11

You are given the following:

$$\boldsymbol{q} = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} \quad \boldsymbol{l} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

What is the distance from 2D point $\boldsymbol{q}$ to the line $\boldsymbol{l}$?

a) 0.33

b) 16

c) -4

d) 5.33

e) 0.74

f) 8

g) 2.67

h) 7.16

i) 2.39

j) 0

Question 12

Harris corner detector. For a small region of a larger image, we have computed the elements of the smoothed Hessian matrix/structure tensor. They are available in `harris.npy` and also presented here:

$g * (I_x^2)$

	0	1	2	3	4
0	16.8	18.5	20.0	20.8	20.6
1	21.0	23.4	25.5	26.7	26.5
2	25.8	29.0	31.8	33.4	33.2
3	30.4	34.4	37.9	39.9	39.7
4	33.9	38.6	42.7	45.1	44.9

$g * (I_y^2)$

	0	1	2	3	4
0	35.2	31.8	27.1	22.0	17.6
1	33.3	30.2	25.9	21.4	17.4
2	29.3	26.8	23.3	19.5	16.2
3	24.4	22.5	19.8	16.9	14.6
4	19.5	18.2	16.3	14.4	12.8

$g * (I_x I_y)$

	0	1	2	3	4
0	-6.5	-6.3	-5.2	-3.3	-1.0
1	-6.7	-6.9	-6.0	-4.1	-1.6
2	-6.5	-7.1	-6.4	-4.7	-2.3
3	-5.9	-6.7	-6.3	-4.9	-2.8
4	-4.8	-5.7	-5.7	-4.8	-3.2

Let $k = 0.06$ and set the threshold to $\tau = 516$.

Does the Harris corner detector find any corners in the image? Corners are specified as (row index, column index).

a) There is a corner at (2, 2).

b) There is a corner at (3, 3).

c) There is a corner at (3, 1).

d) There is a corner at (1, 2).

e) There is a corner at (1, 1).

f) There is a corner at (2, 3).

g) There is a corner at (2, 1).

h) There is a corner at (1, 3).

i) There is a corner at (3, 2).

j) There is no corner in the image.

Question 13

RANSAC. Load the supplied file `RANSAC.npy`.

You are fitting a straight line to a set of 2D points (`points`) with RANSAC. In the current iteration you fit the line through `x1` and `x2`. You set the threshold to $\tau = 0.2$

How many points are inliers in the current iteration?

a) 95

b) 30

c) 51

d) 100

e) 6

f) 96

g) 34

h) 74

i) 41

j) 90

k) 0

l) 46

Question 14

We are using RANSAC to estimate a homography matrix. At iteration number 191 we find a model where 103 out of 404 point matches are inliers, which is the highest number of inliers we have observed so far.

Given the current information, what is the smallest number of iterations we need to run in total in order to be 95% sure that we will have fitted at least one model to only inliers?

a) 167824

b) 708

c) 323

d) 516

e) 24285

f) 13

g) 9

h) 31

i) 1088

Question 15

SIFT features are found using Differences of Gaussians (DoGs).

Which of the following statements about Differences of Gaussians in SIFT is true?

a) The DoG is a combination of two images with different levels of contrast and brightness, resulting in an image that highlights the edges and details of the original image.

b) Since the Gaussian is separable we only do subtraction and convolutions in the x-direction when computing the DoG.

c) To compute the DoG at scale σ we need to blur the image with separable Gaussians at scales $\frac{1}{k}\sigma$, σ, and $k\sigma$.

d) The DoG is obtained by subtracting two 1D Gaussians with scales σ and $k\sigma$ from each other, and convolving the image with this in both the x- and y-direction.

e) Because the image is 2D we convolve it with a single 2D Gaussian in order obtain the images needed to compute the DoG.

f) The DoG is obtained by subtracting two images that have been convolved with 1D Gaussians in both x- and y- directions.

Question 16

Regarding the invariances of SIFT, which of the following statements are true?

a) SIFT is invariant to scale because the keypoints are detected in the differences of Gaussians pyramid.

b) SIFT is invariant to perspective distortions because it uses affine transformations.

c) SIFT is invariant to rotation because it uses separable Gaussian filters.

d) SIFT is invariant to lighting because it uses learned features.

e) SIFT is invariant to noise because it uses median filters.

f) SIFT is invariant to rotation because the descriptor is a circular histogram of oriented gradients.

g) SIFT is invariant to translation because it uses SVD to find the keypoint.

Question 17

You are tasked with implementing a visual odometry algorithm and you have already calibrated the camera.

You need to find the pose of the second camera relative to the first, and choose to use RANSAC. What is the best thing to estimate with RANSAC in order to find the relative pose between the first two cameras?

a) The homography matrix because it requires the fewest correspondences to estimate and can be decomposed into the relative pose.

b) The essential matrix because it requires fewer correspondences to estimate.

c) The relative pose, because it only requires 3+1 correspondences to estimate with `cv2.solvePnP`

d) The essential matrix because it has more degrees of freedom.

e) The fundamental matrix because it has more degrees of freedom.

f) The fundamental matrix because it has more epipolar constraints.

Question 18

The pose of a camera is aligned with the world coordinate system, and it has camera matrix

$$\boldsymbol{K} = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}.$$

What is the projection matrix of this camera?

a) $\begin{bmatrix} k_{11} & k_{12} & k_{13} & 0 \\ k_{21} & k_{22} & k_{23} & 0 \\ k_{31} & k_{32} & k_{33} & 0 \end{bmatrix}$

b) $\begin{bmatrix} k_{11} & k_{12} & k_{13} & 0 \\ k_{21} & k_{22} & k_{23} & 0 \\ k_{31} & k_{32} & k_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

d) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

f) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

g) $\boldsymbol{K}^{-1}$

h) $\begin{bmatrix} k_{11} & k_{12} & k_{13} & 1 \\ k_{21} & k_{22} & k_{23} & 1 \\ k_{31} & k_{32} & k_{33} & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

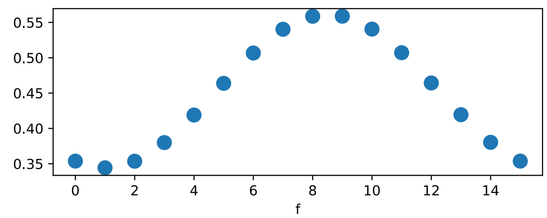
i) $\begin{bmatrix} k_{11} & k_{12} & k_{13} & 1 \\ k_{21} & k_{22} & k_{23} & 1 \\ k_{31} & k_{32} & k_{33} & 1 \end{bmatrix}$

j) $\boldsymbol{K}$

Question 20

A projector doing phase shifting, has projected a sequence of sinusoidal patterns with the same number of periods in all of them onto a scene. For each projection a camera captured an image.

You plot the intensity of the camera pixel at (40, 16) (row, column) in the camera for all captured images and get the following plot:



What can we infer about the length of the sequence (s) and the number of periods in the pattern (n)?

a) $s = 40$, $n = 8$.

b) $s = 40$ n cannot be determined.

c) s cannot be determined, $n = 16$

d) $s = 16$ n cannot be determined.

e) $s = 8$ n cannot be determined.

f) s cannot be determined, $n = 40$

g) s cannot be determined, n cannot be determined.

h) $s = 40$ n cannot be determined.

i) $s = 8$, $n = 8$.

j) $s = 16$, $n = 16$.

k) $s = 16$, $n = 40$.

l) s cannot be determined, $n = 40$

m) $s = 8$, $n = 40$.

n) s cannot be determined, $n = 8$

o) $s = 40$, $n = 16$.

Question 21

This question refers to Exercise 13 in the course:

What is the value of the estimated disparity map at (360, 400) (row, column; zero indexing) without doing subpixel refinement?

a) 14

b) 63

c) 0

d) 7

e) 4

f) -21

g) -2