Solution cv2 endterm

Computer Vision II: Multiple view geometry (Technische Universität München)

Chair of Computer Vision & Artificial Intelligence
School of Computation, Information and Technology
Technical University of Munich

Esolution

Place student sticker here

Note:

- During the attendance check a sticker containing a unique code will be put on this exam.
- This code contains a unique number that associates this exam with your registration number.
- This number is printed both next to the code and to the signature field in the attendance check list.

Computer Vision II: Multiple View Geometry

Exam:	IN2228 / Endterm	**Date:**	Friday 4th August, 2023
Examiner:	Dr. Haoang Li	**Time:**	08:00 – 10:00

	P 1	P 2	P 3	P 4	P 5	P 6
I						

Working instructions

- This exam consists of **12 pages** with a total of **6 problems**.
  Please make sure now that you received a complete copy of the exam.
- The total amount of achievable credits in this exam is 110 credits.
- Detaching pages from the exam is prohibited.
- Allowed resources:
  - one **pocket calculator** for basic algebraic operations and matrix multiplication
  - one **analog dictionary** English ↔ native language
- Do not write with red or green colors nor use pencils.
- Physically turn off all electronic devices, put them into your bag and close the bag.
- In all the exam questions, the ordinary coordinates of an image point are denoted by $\mathbf{p} = [u, v]^\top$ where u and v represent the column ID and row ID, respectively. The homogeneous coordinates of an image point are denoted by $\overline{\mathbf{p}} = [u, v, 1]^\top$. Similarly, the ordinary coordinates of a 3D point are denoted by $\mathbf{P} = [x, y, z]^\top$. The homogeneous coordinates of a 3D point are denoted by $\overline{\mathbf{P}} = [x, y, z, 1]^\top$.
- The unit is pixels, unless otherwise specified.
- For calculation questions, please round a decimal number to three decimal places (e.g., 1.234 and 0.123).

Left room from ________ to ________ / Early submission at ________

Problem 1 **Multiple Choice (22 credits)**

For each question, any number of answers (including one, two, three, and four) can be correct. You get full credits if all boxes are correct, and 0 otherwise.

Mark correct answers with a cross ☒
To undo a cross, completely fill out the answer option ■
To re-mark an option, use a human-readable marking ×■

a) Which statement(s) about Lie group and Lie algebra is/are correct?

- [x] Special Euclidean group SE(3) is commonly used to describe the 3D transformation with both rotation and translation.
- [x] The exponential map from the Lie algebra $\mathfrak{so}(3)$ to the Lie group SO(3) corresponds to Rodrigues' formula.
- [ ] The set of rotation matrices and the operation of addition constitute a special orthogonal group.
- [ ] Special orthogonal group SO(n) is a Lie Group, but special Euclidean group SE(n) is not a Lie Group.

b) Which statement(s) about image distortion is/are wrong?

- [x] Tangential distortion can be classified into barrel distortion and pincushion distortion.
- [ ] Tangential distortion is mainly caused by the misalignment of the lens.
- [x] For an ordinary lens with limited radial distortion, we typically consider a quartic polynomial function for distortion modeling.
- [x] The amount of radial distortion is typically a linear function of its distance from the principal point.

c) Which statement(s) about Zhang's method for camera calibration is/are correct?

- [ ] This method first determines the extrinsic parameters, followed by intrinsic parameters.
- [ ] This method involves QR/RQ decomposition to obtain extrinsic and intrinsic parameters.
- [x] After rotation computation, this method needs a post-processing step to enforce the orthogonality constraint of rotation.
- [ ] This method uses a single image of a chessboard.

d) Which statement(s) about Laplacian of Gaussian (LoG) kernel for blob detection is/are correct?

- [x] At the optimal scale of a blob, the convolution result between LoG kernel and an image patch reaches an extremum.
- [x] The parameter of LoG kernel is associated with the scale of blob.
- [ ] LoG kernel is more efficient than the difference of Gaussian (DoG) kernel.
- [x] To discard too close detected blobs, we typically exploit non-maximum suppression.

e) Which statement(s) about the eight-point algorithm for relative pose estimation is/are correct?

- [x] Eight-point algorithm can be applied to the estimation of both essential and fundamental matrices.
- [ ] Each point correspondence provides two linear constraints on the essential matrix.
- [x] Eight-point algorithm assumes that the entries of the essential matrix are independent, i.e., it neglects some inherent properties of the essential matrix.
- [x] When solving the linear system with respect to the elements of the essential matrix, we typically enforce a constraint on the norm of the unknown vector to avoid a zero solution.

f) Which statement(s) about triangulation based on a 2D-2D point correspondence is/are correct?

- [ ] Geometrically, a reconstructed 3D point is the midpoint of an arbitrary 3D line segment connecting two projection rays.
- [x] Both known poses and known intrinsic parameters of two cameras are needed by triangulation.
- [x] Based on the absolute poses of two cameras with respect to the world frame, a triangulated point is in the world frame.
- [ ] Mathematically, we generate a well-determined linear system with respect to an unknown 3D point.

g) Which statement(s) about 3D-2D geometry is/are wrong?

- [x] Perspective-n-points algorithms typically estimate the extrinsic and intrinsic parameters at the same time.
- [x] Perspective-n-points algorithms take the 3D points in the camera frame and corresponding 2D points as input.
- [x] Perspective-n-points algorithms can only determine 5 degrees of freedom of the camera pose.
- [x] The minimal case of perspective-n-points algorithms is five point correspondences.

h) Which statement(s) about perspective-n-point algorithm is/are wrong?

- [x] EPnP algorithm introduces three control points to estimate the absolute camera pose.
- [x] P3P algorithm relies on the circumference angles computed by the input 3D points.
- [x] Perspective-n-lines algorithms require that an endpoint of a 2D line segment correspond to an endpoint of a 3D line segment.
- [ ] In the presence of outliers, P3P algorithm is more suitable than EPnP algorithm to integrate with RANSAC algorithm in terms of valid sampling.

i) Which statement(s) about photometric error for 2D-2D geometry is/are correct?

- [x] Given a pair of calibrated RGB images (with the known intrinsic parameters), the parameters to optimize in the photometric error are only the depths of pixels and the relative camera pose.
- [x] Photometric error typically relies on the brightness consistency assumption.
- [ ] Minimization of photometric error is not sensitive to the initial values of parameters.
- [x] Inverse depth parametrization can improve the numerical stability.

j) Which statement(s) about direct SLAM methods is/are wrong?

- [x] Direct SLAM method is a two-step method. It first tracks the features, and then determines the camera movement based on the tracked features.
- [x] Semi-dense direct method typically tracks the pixels with small gradients.
- [ ] Dense direct method is typically slower than the semi-dense direct method and sparse direct method.
- [x] Sparse direct method is equivalent to the feature-based SLAM method such as ORB-SLAM.

k) Which statement(s) about non-linear optimization is/are wrong?

- [ ] The steepest method lets the objective function descend along the negative gradient direction at each iteration.
- [x] Newton's method is more prone to result in a zig-zag descending trajectory than the steepest method.
- [x] Gauss-Newton method is more time-consuming than Newton's method since Gauss-Newton method needs to compute the Hessian matrix.
- [x] Newton's method and Gauss-Newton method both use the second-order Taylor expansion to rewrite objective functions.

Problem 2 Correspondence Establishment (12 credits)

Let us consider the KLT tracker in a simplified case that the motion of an image patch is a pure 2D translation. A point moves from the position $\mathbf{p} = [x, y]^\top$ in the image $\mathcal{I}_0$ to the position $\mathbf{p}' = [x + u, y + v]^\top$ in the image $\mathcal{I}_1$. At the position $\mathbf{p} = [x, y]^\top$ in the image $\mathcal{I}_0$, the directional derivatives along the x- and y-axes are denoted by I_x and I_y, respectively. At the position $\mathbf{p} = [x, y]^\top$ in the image $\mathcal{I}_1$, the directional derivatives along the x- and y-axes are denoted by I'_x and I'_y, respectively. We denote the intensity at the position $\mathbf{p} = [x, y]^\top$ of the image I_0 by $I_0(x, y)$, and denote the intensity at the position $\mathbf{p} = [x, y]^\top$ of the image I_1 by $I_1(x, y)$. We denote the intensity difference between $I_0(x, y)$ and $I_1(x, y)$ by ΔI.

0 1 2 3 4 5 6 7 8

a) How to determine whether the pixel $\mathbf{p}$ is a good feature to track, and how to compute 2D motion $[u, v]$ in the image? Note: Please provide a linear system with respect to the 2D motion vector. The derivation of this linear system is not required. **(8 credit(s))**

(4 Marks.) We should track the pixels whose associated M-matrices have large eigenvalues (it is also acceptable if students give M'-matrices).

$$M = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}, \quad M' = \begin{bmatrix} \sum I'_x I'_x & \sum I'_x I'_y \\ \sum I'_x I'_y & \sum I'_y I'_y \end{bmatrix}.$$

(4 Marks. If I'_x and I'_y is written as I_x and I_y, student can get two marks.) We generate a linear system for motion computation, i.e.,

$$\underbrace{\begin{bmatrix} \sum I'_x I'_x & \sum I'_x I'_y \\ \sum I'_x I'_y & \sum I'_y I'_y \end{bmatrix}}_{M'} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum I'_x \Delta I \\ \sum I'_y \Delta I \end{bmatrix}.$$

To estimate the motion $[u, v]^\top$ of a feature, we solve the above linear system, i.e.,

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum I'_x I'_x & \sum I'_x I'_y \\ \sum I'_x I'_y & \sum I'_y I'_y \end{bmatrix}^{-1} \begin{bmatrix} \sum I'_x \Delta I \\ \sum I'_y \Delta I \end{bmatrix}.$$

0 1 2 3 4

b) As shown in Figure 2.1, we have three pixels a, b, and c. In theory, which pixel can be reliably tracked? Please give your conclusion and explain the reason (both intuitive and mathematical explanations are acceptable). **(4 credit(s))**

Figure 2.1: Three candidate pixels a, b, c (i.e., the centers of squares) to track.

(2 Mark) We should track the pixel b.
(2 Marks. Both intuitively explanation based on textures and mathematical explanation based on eigenvalues are acceptable.) The pixel b lies at an image area with sufficient textures. Its associated M-matrix has large eigenvalues.
Given that pixel b is not very distinguishable in the printed exam papers, we also give full marks to a student who chooses the pixel c and gives reasonable explanations.

Problem 3 2D-2D Geometry (24 credits)

We use a monocular camera to obtain two images $\mathcal{I}_1$ and $\mathcal{I}_2$ from different viewpoints. This camera has been calibrated beforehand.

a) The image height is 480 pixels, the image width is 640 pixels, the focal length is 800 pixels, and the principal point lies at the image center. We do not consider the image distortion and skew factor. Assume that we have known the relative rotation $\mathbf{R}_{1\to2}$ and translation $\mathbf{t}_{1\to2}$ from the first camera frame to the second camera frame:

$$\mathbf{R}_{1\to2} = \begin{bmatrix} -0.83 & 0.56 & 0 \\ -0.56 & -0.83 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{t}_{1\to2} = \begin{bmatrix} 0.1 \\ 0.5 \\ 0.1 \end{bmatrix}.$$

In the image $\mathcal{I}_1$, we extract a point $\mathbf{p} = [393, 167]^\top$. In the image $\mathcal{I}_2$, we extract two points $\mathbf{q}_a = [37, 369]^\top$ and $\mathbf{q}_b = [235, 331]^\top$. Based on the knowledge of epipolar geometry, which of the points $\mathbf{q}_a$ and $\mathbf{q}_b$ is more likely to be the corresponding point to the point $\mathbf{p}$? **(13 credit(s))**

0 1 2 3 4 5 6 7 8 9 10 11 12 13

For each step, student can only get full or zero mark, unless otherwise specified.
(3 Marks. Each normalization result corresponds to 1 Mark.) We define the intrinsic matrix $\mathbf{K}$ by

$$\mathbf{K} = \begin{bmatrix} 800 & 0 & 320 \\ 0 & 800 & 240 \\ 0 & 0 & 1 \end{bmatrix}.$$

We use $\mathbf{K}$ to normalize the image points $\mathbf{q}_a$ and $\mathbf{q}_b$ by

$$\hat{\mathbf{p}} = \mathbf{K}^{-1}\overline{\mathbf{p}} = \begin{bmatrix} 0.091 \\ -0.091 \\ 1 \end{bmatrix}, \quad \hat{\mathbf{q}}_a = \mathbf{K}^{-1}\overline{\mathbf{q}}_a = \begin{bmatrix} -0.354 \\ 0.161 \\ 1 \end{bmatrix}, \quad \hat{\mathbf{q}}_b = \mathbf{K}^{-1}\overline{\mathbf{q}}_b = \begin{bmatrix} -0.106 \\ 0.114 \\ 1 \end{bmatrix}.$$

(4 Marks) Based on the known rotation $\mathbf{R}_{1\to2}$ and translation $\mathbf{t}_{1\to2}$, we compute the essential matrix $\mathbf{E}$ by

$$\mathbf{E} = [\mathbf{t}_{1\to2}]_\times \mathbf{R}_{1\to2} = \begin{bmatrix} 0.056 & 0.083 & 0.500 \\ -0.083 & 0.056 & -0.100 \\ 0.359 & -0.363 & 0 \end{bmatrix}.$$

(6 Marks. Each epipolar constraint corresponds to 2 Marks. Selection of $\mathbf{q}_b$ corresponds to 2 Mark.) We use the essential matrix to validate point pairs $(\hat{\mathbf{p}}, \hat{\mathbf{q}}_a)$ and $(\hat{\mathbf{p}}, \hat{\mathbf{q}}_b)$ as

$$\hat{\mathbf{q}}_a^\top \mathbf{E}\,\hat{\mathbf{p}} \neq 0, \quad \hat{\mathbf{q}}_b^\top \mathbf{E}\,\hat{\mathbf{p}} \approx 0.$$

Given that $(\hat{\mathbf{p}}, \hat{\mathbf{q}}_b)$ basically satisfies the epipolar geometry constraint, the point $\mathbf{q}_b$ is more likely to be the corresponding point to the point $\mathbf{p}$.

0 1 2 3 4

b) We extracted some key points in the images $\mathcal{I}_1$ and $\mathcal{I}_2$ respectively, and then established N 2D-2D point correspondences $\{(\mathbf{p}_i, \mathbf{p}'_i)\}_{i=1}^{N}$. $\mathbf{p}_i$ is in the first image $\mathcal{I}_1$, and $\mathbf{p}'_i$ is in the second image $\mathcal{I}_2$. These correspondences are associated with N 3D points lying on the same 3D plane π.
Assume that we have known the coordinates $\mathbf{n}_2$ of the unit normal of the 3D plane π in the second camera frame, as well as the rotation $\mathbf{R}_{1\to 2}$ and translation $\mathbf{t}_{1\to 2}$ from the first camera frame to the second camera frame. Please express the coordinates $\mathbf{n}_1$ of the normal of the 3D plane π in the first camera frame. **(4 credit(s))**

(**2 Mark**) We first compute the rotation $\mathbf{R}_{2\to 1}$ from the second camera frame to the first camera frame by

$$\mathbf{R}_{2\to 1} = \mathbf{R}_{1\to 2}^{\top}.$$

(**2 Mark**) We compute the normal $\mathbf{n}_1$ of the plane in the first camera frame by

$$\mathbf{n}_1 = \mathbf{R}_{2\to 1}\mathbf{n}_2.$$

0 1 2 3

c) Each point correspondences $(\mathbf{p}_i, \mathbf{p}'_i)$ in the above subproblem b) can be fitted by a 3×3 homography as

$$\overline{\mathbf{p}}'_i = \mathbf{H}\overline{\mathbf{p}}_i.$$

Please express the homography $\mathbf{H}$ based on 1) the homogeneous coordinates of the 3D plane $\overline{\pi} = [a, b, c, d]^{\top}$ in the first camera frame, 2) the rotation $\mathbf{R}_{1\to 2}$ and translation $\mathbf{t}_{1\to 2}$ from the first camera frame to the second camera frame, and 3) the intrinsic matrix $\mathbf{K}$ of camera. No derivation is required. **(3 credit(s))**

(3 Marks) The homography can be expressed by

$$\mathbf{H} = \mathbf{K}\big(\mathbf{R}_{1\to 2} + \mathbf{t}_{1\to 2}\frac{[a, b, c]}{-d}\big)\mathbf{K}^{-1}.$$

0 1 2 3 4

d) Based on the direct linear transformation (DLT), what is the minimum number of 2D-2D point correspondences to compute the essential matrix and homography, respectively? No derivation is needed. You are only required to give your conclusion. **(4 credit(s))**

(2 Mark) Eignt point correspondences for the essential matrix.
(2 Mark) Four point correspondences for the homography.

Problem 4 3D-2D Geometry (17 credits)

a) For perspective-n-point problem, please list two algorithms that involve computing the coordinates of 3D points in both world and camera frames. **(4 credit(s))**

0 1 2 3 4

(**4 Marks. Each algorithm corresponds to 2 Marks.**) P3P and EPnP.

b) Given the rotation $\mathbf{R}_{\mathcal{C}\to\mathcal{W}}$ and translation $\mathbf{t}_{\mathcal{C}\to\mathcal{W}}$ from the camera frame $\mathcal{C}$ to the world frame $\mathcal{W}$, please express the rotation $\mathbf{R}_{\mathcal{W}\to\mathcal{C}}$ and translation $\mathbf{t}_{\mathcal{W}\to\mathcal{C}}$ from the world frame $\mathcal{W}$ to the camera frame $\mathcal{C}$. **(4 credit(s))**

0 1 2 3 4

(4 Marks. Each formula corresponds to 2 Marks. If there is a small error in a formlular 1 Mark.)

$$\mathbf{R}_{\mathcal{W}\to\mathcal{C}} = \mathbf{R}_{\mathcal{C}\to\mathcal{W}}^\top, \quad \mathbf{t}_{\mathcal{W}\to\mathcal{C}} = -\mathbf{R}_{\mathcal{C}\to\mathcal{W}}^\top \mathbf{t}_{\mathcal{C}\to\mathcal{W}}.$$

c) There is a 3D point $\mathbf{P}$. We know its coordinates in the world frame $\mathcal{W}$, i.e., $\mathbf{P}_\mathcal{W} = [-0.5, 1, 5.7]^\top$. We use a monocular camera to obtain an image $\mathcal{I}$. The camera has been calibrated beforehand. The image height is 480 pixels, the image width is 640 pixels, the focal length is 800 pixels, and the principal point lies at the image center. We do not consider the image distortion.
The above 3D point $\mathbf{P}$ corresponds to a 2D point $\mathbf{p} = [393, 167]^\top$ in the image $\mathcal{I}$. We know the rotation $\mathbf{R}_{\mathcal{W}\to\mathcal{C}}$ and translation $\mathbf{t}_{\mathcal{W}\to\mathcal{C}}$ from the world frame $\mathcal{W}$ to the camera frame $\mathcal{C}$:

0 1 2 3 4 5 6 7 8 9

$$\mathbf{R}_{\mathcal{W}\to\mathcal{C}} = \begin{bmatrix} -0.83 & -0.56 & 0 \\ 0.56 & -0.83 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{t}_{\mathcal{W}\to\mathcal{C}} = \begin{bmatrix} 0.70 \\ 0.69 \\ -0.19 \end{bmatrix}.$$

Please compute the reprojection error with respect to the 3D point $\mathbf{P}$ and the image point $\mathbf{p}$. **(9 credit(s))**

(2 Marks) We transform the 3D point $\mathbf{P}_\mathcal{W}$ from the world frame to the camera frame as

$$\mathbf{P}_\mathcal{C} = \mathbf{R}_{\mathcal{W}\to\mathcal{C}}\mathbf{P}_\mathcal{W} + \mathbf{t}_{\mathcal{W}\to\mathcal{C}} = \begin{bmatrix} 0.555 \\ -0.420 \\ 5.510 \end{bmatrix}.$$

(2 Mark) We define the intrinsic matrix $\mathbf{K}$ by

$$\mathbf{K} = \begin{bmatrix} 800 & 0 & 320 \\ 0 & 800 & 240 \\ 0 & 0 & 1 \end{bmatrix}.$$

(2 Marks) We use the intrinsic matrix $\mathbf{K}$ to project the 3D point $\mathbf{P}_\mathcal{C}$ into a 2D point $\mathbf{p}'$.
(1 Marks) and scale the point:

$$\overline{\mathbf{p}}' = \mathbf{K}\mathbf{P}_\mathcal{C} = \begin{bmatrix} 400.581 \\ 179.020 \\ 1 \end{bmatrix}.$$

Step 3. (2 Marks)
We compute the distance d between the extracted point $\mathbf{p}$ and the projection $\mathbf{p}'$ as

$$d = \|\mathbf{p} - \mathbf{p}'\| = 14.211.$$

Problem 5 3D-3D Geometry (18 credits)

There are N static 3D points in a 3D space. We use a monocular camera to observe these points from different viewpoints. The coordinates of these points in the first and second camera frames are denoted by $\{\mathbf{P}_i\}_{i=1}^N$ and $\{\mathbf{Q}_i\}_{i=1}^N$. Assume that we have established N 3D-3D point correspondences $\{(\mathbf{P}_i, \mathbf{Q}_i)\}_{i=1}^N$. For example, the 3D point $\mathbf{P}_1$ corresponds to the 3D point $\mathbf{Q}_1$.

0 1 2 3

a) Let us use $\mathbf{R}_{2\to1}$ and $\mathbf{t}_{2\to1}$ to denote the rotation and translation from the second camera frame to the first camera frame, respectively. Please define the objective function with respect to $\mathbf{R}_{2\to1}$ and $\mathbf{t}_{2\to1}$ to align the above corresponding points. **(3 credit(s))**

(3 Marks)

$$\min_{\mathbf{R}_{2\to1},\mathbf{t}_{2\to1}} \sum_{i=1}^{N} \|\mathbf{R}_{2\to1}\mathbf{Q}_i + \mathbf{t}_{2\to1} - \mathbf{P}_i\|^2,$$

where $\|\cdot\|$ represents the L_2-norm.

0 1 2 3 4 5 6 7 8 9 10 11 12

b) Given the above 3D-3D point correspondences, please state the main steps to compute the rotation $\mathbf{R}_{2\to1}$ and translation $\mathbf{t}_{2\to1}$ from the second camera frame to the first camera frame. **(12 credit(s))**

For each step, student can only get full or zero mark, unless otherwise specified.
(2 Marks. Each formula corresponds to 1 Mark.) We compute the centers of mass of two point sets in different camera frames, respectively:

$$\mu_{\mathcal{P}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{P}_i, \quad \mu_{\mathcal{Q}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Q}_i.$$

(2 Marks. Each formula corresponds to 1 Mark.) We normalize two point sets using respective centers of mass, i.e.,

$$\hat{\mathbf{P}}_i = \mathbf{P}_i - \mu_{\mathcal{P}}, \quad \hat{\mathbf{Q}}_i = \mathbf{Q}_i - \mu_{\mathcal{Q}}.$$

(2 Marks) Based on the normalized point sets, we compute a matrix $\mathbf{W}$ by

$$\mathbf{W} = \sum_{i=1}^{N}\hat{\mathbf{P}}_i\hat{\mathbf{Q}}_i^\top.$$

(2 Marks) We perform the singular value decomposition (SVD) of the matrix $\mathbf{W}$:

$$\mathbf{W} = \mathbf{U}\begin{bmatrix}\sigma_1 & 0 & 0\\ 0 & \sigma_2 & 0\\ 0 & 0 & \sigma_3\end{bmatrix}\mathbf{V}^\top.$$

(2 Mark) We use the above matrices $\mathbf{U}$ and $\mathbf{V}$ to compute the rotation $\mathbf{R}_{2\to1}$ from the second camera frame to the first camera frame:

$$\mathbf{R}_{2\to1} = \mathbf{U}\mathbf{V}^\top.$$

(2 Mark) Then we use the estimated rotation $\mathbf{R}_{2\to1}$ and the above centers of mass to compute the translation $\mathbf{t}_{2\to1}$ by

$$\mathbf{t}_{2\to1} = \mu_{\mathcal{P}} - \mathbf{R}_{2\to1}\mu_{\mathcal{Q}}.$$

c) In practice, outliers of point correspondences are inevitable. Please list three methods to remove outliers. **(3 credit(s))**

0
1
2
3

(3 Marks. Each method corresponds to 1 Mark.) Expectation-maximization, RANSAC, and robust kernel function (M-estimator) such as Huber loss.

Problem 6 Single-view Geometry (17 credits)

We use a monocular camera to obtain two images $\mathcal{I}_1$ and $\mathcal{I}_2$ from different viewpoints in a man-made environment satisfying Manhattan world assumption. The intrinsic matrix $\mathbf{K}$ of the camera is known beforehand.

0 1 2 3 4 5 6 7 8 9 10

a) Assume that in the image $\mathcal{I}_1$, we have estimated two vanishing points $\mathbf{v}_1$ and $\mathbf{v}_2$. Please state how to compute the third vanishing point $\mathbf{v}_3$ in the image $\mathcal{I}_1$. **(10 credit(s))**

(4 Marks) We use the vanishing points $\bar{\mathbf{v}}_1$ and $\bar{\mathbf{v}}_2$ to compute the vanishing directions $\mathbf{d}_1$ and $\mathbf{d}_2$:

$$\mathbf{d}_1 = \mathbf{K}^{-1}\bar{\mathbf{v}}_1, \quad \mathbf{d}_2 = \mathbf{K}^{-1}\bar{\mathbf{v}}_2.$$

2 marks if a student says "unproject", "compute vanishing directions," or similar without formulas.
0 marks if a student treats v_i as vanishing directions, but the student can still get whole marks on the following points.
(3 Marks) Based on Manhattan world assumption, the third vanishing direction is orthogonal to vanishing directions $\mathbf{d}_1$ and $\mathbf{d}_2$. We thus compute $\mathbf{d}_3$ by

$$\mathbf{d}_3 = \mathbf{d}_1 \times \mathbf{d}_2.$$

1 mark if a student says that d_3 should be orthogonal to d_1 and d_2 without a formula.
(3 Marks) Given the estimated direction $\mathbf{d}_3$, we compute the vanishing point $\bar{\mathbf{v}}_3$ by

$$\bar{\mathbf{v}}_3 = \mathbf{K}\mathbf{d}_3.$$

1 mark if a student says "project", "intersect with an image plane" or similar without formulas.

Any sampling based method, which doesn't garantee to produce a correct result recieves maximum 5 marks if fully correct.

0 1 2 3 4 5 6 7

b) In the image $\mathcal{I}_2$, we have estimated three vanishing points $\{\mathbf{v}'_i\}_{i=1}^3$. Assume that we have prior information that $\mathbf{v}'_i$ is associated with the above vanishing point $\mathbf{v}_i$ $(1 \leqslant i \leqslant 3)$. Please state how you can use this information to identify whether a relative camera rotation $\mathbf{R}_{1\to 2}$ from the first camera frame to the second camera frame is correct. Here, we do not consider the effect of noise. **(7 credit(s))**

(2 Marks) We use the vanishing point $\bar{\mathbf{v}}'_i$ to compute the vanishing direction $\mathbf{d}'_i$ in the second camera frame:

$$\mathbf{d}'_i = \mathbf{K}^{-1}\bar{\mathbf{v}}'_i.$$

1 mark if a student says "unproject", "compute vanishing directions," or similar without formulas if they were not presented in (a).
0 marks if a student treats v_i as vanishing directions, but the student can still get whole marks on the following points.
(3 Marks) We transform the vanishing direction $\mathbf{d}_i$ in the subproblem a) from the first camera frame to the second camera frame based on the rotations $\mathbf{R}_{1\to 2}$:

$$\mathbf{d}''_i = \mathbf{R}_{1\to 2}\mathbf{d}_i.$$

1 mark if a student says "apply rotation", "rotate" or similar without formula. It should be clear from the phrasing that only rotation is applied.
(2 Marks) If $\{\mathbf{d}''_i\}_{i=1}^3$ are aligned to $\{\mathbf{d}'_i\}_{i=1}^3$ respectively (i.e., corresponding vectors are equal up to scale), $\mathbf{R}_{1\to 2}$ is a correct rotation.
1 mark if a student check for equality or says "compare", "should align" if the vanishing directions were not normalized.

Additional space for solutions–clearly mark the (sub)problem your answers are related to and strike out invalid solutions.