CSE483-Mobile Robotics

Mid-semester exam

Monsoon 2019

September 21st

Maximum points: 30 Duration: 90 minutes

Instructions

- This is an **open-book** exam. You are allowed to use any paper notes or textbooks that you have brought with you.
- Laptops, tablets, or smartphones are NOT allowed. You also cannot collaborate with other students.
- Your answers must be concise and to-the-point. Verbosity will NOT fetch you additional marks.
- Sufficient space has been provided for each question. Using additional sheets are discouraged, if you need them you're probably doing something wrong.
- You do NOT get credit for replicating whatever is present in the textbook or your notes. Please do not fill your answer scripts with excerpts from such sources.
- Use the last page for rough work or for any of your answers, if necessary.
- State your assumptions clearly if there is any ambiguity with the question(s).

Roll number: TA

Seat:

Invigilator sign:

Q1	Q2	Q3	Q4	Q5	Q6	Total

Q1) *Warm-up:* Fill up the following table by indicating the quantities that are known, to be estimated, or unknown, and the type of measurements that are needed. **(5 points)**

Problem	Structure (Scene geometry)	Motion (Camera parameters)	Measurements
F-matrix estimation	Unknown	Estimate	2D - 2D features
Camera calibration	Known	Estimate	2D-3D features
Triangulation	Estimate	Known	2D-2D features
Stereo rectification	Unknown	Known & Estimate	-
PnP	Known	Estimate	2D-3D features
Bundle adjustment	Estimate	Estimate	2D-3D features

Q2) *Transformations:*

(i) Derive the expression for T_W^C if $T_C^W = \begin{bmatrix} R_C^W & P_{CORG}^W \\ \mathbf{0}_{3\times1} & 1 \end{bmatrix}$. **(2 points)**

We know $(T_c^w)(T_w^c) = I$

$$\Rightarrow \begin{pmatrix} R_c^w & P_{c,org}^w \\ 0_{1\times3} & 1 \end{pmatrix}\begin{pmatrix} R_w^c & P_{w,org}^c \\ 0_{1\times3} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\Rightarrow R_c^w R_w^c = I_{3\times3}$$

$$\Rightarrow R_w^c = (R_c^w)^{-1}$$

$$= (R_c^w)^T$$

And, $R_c^w P_{w,org}^c + P_{c,org}^w = 0_{3\times1}$

$$\Rightarrow P_{w,org}^c = (-R_c^w)^{-1} P_{c,org}^w$$

$$= -(R_c^w)^T P_{c,org}^w$$

$$\therefore T_w^c = \begin{pmatrix} (R_c^w)^T & -(R_c^w)^T P_{c,org}^w \\ 0_{1\times3} & 1 \end{pmatrix}$$

(ii) Consider the following figure and answer questions (a) to (c).

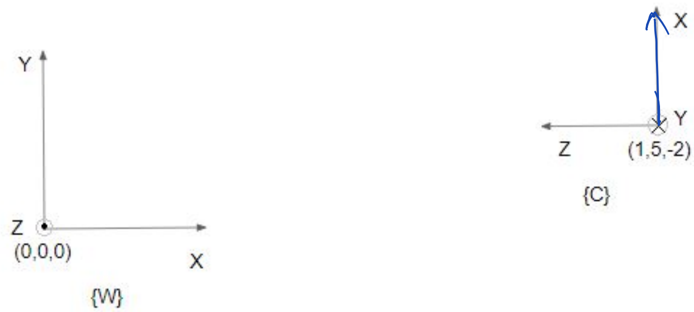


Figure 1: World Frame and Camera Frame

{W} represents the world frame and {C} represents the camera frame. The Z axis of {W} is coming out of the plane. Whereas the Y axis of {C} is going into the plane.

(a) Find R_C^W. ~~(2 points)~~ (1 point)

Expressing the unit coordinate axes vectors of {C} in terms of {W},

$$X_c^w = (0, 1, 0)^T$$

$$Y_c^w = (0, 0, -1)^T$$

$$Z_c^w = (-1, 0, 0)^T$$

Stacking them column-wise we get,

$$R_c^w = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

(b) Find the YXZ-Euler angles representation for R^W_C. ~~(2 points)~~ (1 point)

$$R_c^w = R_y(-\pi/2)\, R_x(0)\, R_z(\pi/2)$$

$$= \begin{pmatrix} c(-\pi/2) & 0 & s(-\pi/2) \\ 0 & 1 & 0 \\ -s(-\pi/2) & 0 & c(-\pi/2) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c(\pi/2) & -s(\pi/2) & 0 \\ s(\pi/2) & c(\pi/2) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

This is only one possible combination. Euler angles are not unique.

(c) Find P^C_{WORG} and T^C_W. **(1 points)**

$$P^c_{w,org} = -R^c_w \; P^w_{c,org}$$

$$= -(R^w_c)^T P^w_{c,org}$$

$$= -\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 5 \\ -2 \end{pmatrix} = \begin{pmatrix} -5 \\ -2 \\ +1 \end{pmatrix}$$

$$T^c_w = \begin{pmatrix} R^c_w & P^c_{w,org} \\ 0_{1\times 3} & 1 \end{pmatrix} = \begin{pmatrix} (R^w_c)^T & P^c_{w,org} \\ 0_{1\times 3} & 1 \end{pmatrix}$$

Q3.1) *Single-view geometry:* Given a camera matrix P, detail how you can obtain the camera center and the rotation matrix R without knowing the intrinsic parameter matrix K. **(2 points)**.

$$P = K[R \mid t]$$

$$= [KR \mid Kt]$$

$$KR = [p_1 \ p_2 \ p_3]$$

Taking $(KR)(KR)^T$

$$\Rightarrow KK^T = [p_1 \ p_2 \ p_3] \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix}$$

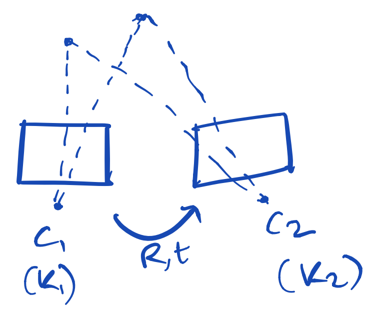
From this we can solve for K.

Once we find K,

$$R = K^{-1} [p_1 \ p_2 \ p_3]$$

$$t = K^{-1} p_4$$

Q3.2) *Reconstruction:* State and justify the cases when the 3D reconstruction obtained from two views is (a) Unambiguous (b) Up to an unknown scaling factor (c) Up to an unknown projective transformation. **(3 points)**



(a) When K_1 & K_2, R&t are known the reconstruction is unambiguous.

When all these quantities are known, corresponding points can be back-projected to their respective rays in 3D and their intersections can be found.

The reconstruction is valid upto an Euclidean transformation depending on the choice of the arbitrary world reference frame.

(b) When K_1 & K_2 are known, but R & t is unknown the reconstruction is valid only up to scale.

Since the R & t is unknown we can estimate it only by estimating the essential matrix relating the two views using point correspondences. And since the essential matrix is homogeneous i.e. it is valid only up to scale, the recovered translation t will also be valid only up to scale. Hence the triangulated 3D points will also only be valid upto scale.

(c) When both K_1 & K_2, R_1 & R_2 are unknown the reconstruction is valid only upto a projective transformation M.

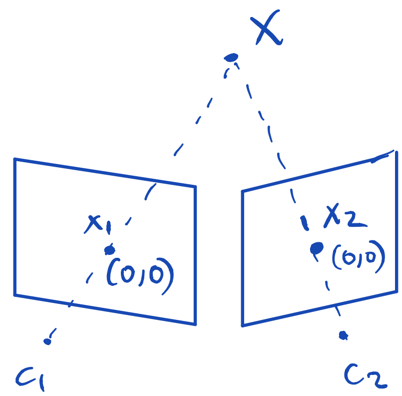
In this case we can only estimate the fundamental matrix F relating the two views using point correspondences. And since F is uniquely determined only by point correspondences, there can be multiple choices of the camera matrix P & world point X that give rise to the same image points.

$$\text{ie } \mathbf{x} = PX = (PM^{-1})(MX)$$

for any 4x4 projective transformation M.

Hence we say the reconstruction can at best be recovered only upto a projective transformation.

Q4) *Essential matrix:* Two cameras fixate on a point P in 3D space such that their *optical axes* intersect at this point. Show that the E_{33} element of their associated Essential matrix E is zero. **(5 points)**



We know the optical axis passes through the center of the image plane i.e. the principal point.

This means the 3D point is imaged by both the cameras onto their respective principal points. These principal points are $(0,0,\lambda)$ in normalized coordinates.

Applying the epipolar constraint,

$$x_2^T E x_1 = 0$$

$$\Rightarrow \begin{bmatrix} 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \lambda \end{bmatrix} = 0$$

$$\Rightarrow \lambda \begin{bmatrix} E_{31} & E_{32} & E_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \lambda \end{bmatrix} = 0$$

$$\Rightarrow \qquad \lambda^2 E_{33} = 0$$

Since $\lambda \neq 0$

$$\Rightarrow \quad E_{33} = 0.$$

Q5) *Homography*: Suppose a camera, with intrinsic matrix K, rotates about its optical centre by a rotation matrix R. (a) Show that its two views are related by a homography H such that $x_2 = Hx_1$ where x_1 is a point in the first image and x_2 is its corresponding point in the second image. **(2.5 points)** (b) Also show that if θ is the rotation between the two views, then the angle 2θ corresponds to the homography H^2. **(2.5 points)**

(a) Consider a point $X \in \mathbb{R}^3$.

Its location on the first camera's image plane is given as,

$$\lambda_1 x_1 = K[I|0]X$$

$$\Rightarrow \lambda_1 K^{-1} x_1 = X$$

After rotation by R, the same point X is imaged as

$$\lambda_2 x_2 = K[R|0]X$$

$$\Rightarrow \lambda_2 X_2 = K[R|0]\lambda_1 K^{-1} x_1$$

$$\Rightarrow \lambda_2 X_2 = \lambda_1 K[R|0]K^{-1}x_1$$

$$\Rightarrow x_2 = KRK^{-1}x_1 \quad \text{(dropping the constants)}$$

$$\Rightarrow x_2 = Hx_1$$

where $H \in \mathbb{R}^{3\times3}$ is a homography that maps $\mathbb{P}^2 \to \mathbb{P}^2$.

(b) If the camera is rotated by 2θ ie $[R|0][R|0]$

then

$$x_2 = K[R|0][R|0]K^{-1}x_1$$

$$= K[R|0]K^{-1}K[R|0]K^{-1}x_1$$

$$= H\,H\,x_1$$

$$= H^2 x_1$$

(optional)

$$R(\theta)R(\theta) = \begin{bmatrix} c\theta & -s\theta & 0 \\ s\theta & c\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} c\theta & -s\theta & 0 \\ s\theta & c\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(works for any repeated axis)

$$= \begin{bmatrix} c^2\theta - s^2\theta & -2c\theta s\theta & 0 \\ 2c\theta s\theta & -s^2\theta + c^2\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} c(2\theta) & -s(2\theta) & 0 \\ s(2\theta) & c(2\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= R(2\theta)$$

Q6) *Dense-VO*: Dense-VO is one other type of visual odometry where the camera motion is estimated by aligning consecutive image frames and then finding the transformation that best minimizes the *photometric error* between them. Suppose there is a camera C with known intrinsics K, and it captures two images $\mathrm{I}_1, \mathrm{I}_2$ from two views separated by a rotation R and translation t. The photometric error between these two views is given as $\sum_{\mathrm{x} \in \mathrm{I}_1} \|\mathrm{I}_1(\mathrm{x}) - \mathrm{I}_2(w(\mathrm{x}, (\mathrm{R}|\mathrm{t}))\|^2$ where $w(\mathrm{x}, (\mathrm{R}|\mathrm{t}))$ is a function that maps a point x in the first image to a point in the second image given the camera motion R, t. (a) Assuming d is the depth of the point x in I_1 relative to the first view, describe the steps involved to map this point to the second image, and hence provide a mathematical expression for $w(\mathrm{x}, (\mathrm{R}|\mathrm{t}))$. **(3 points)** (b) What is the nature of this photometric error? Very briefly in words mention how it can be solved for to find the best camera motion. **(2 points)**

(a) Since we know the depth of the point x in I_1, we can back-project it to $X \in \mathbb{R}^3$ by inverting the pinhole projection equations.

$$\lambda x = KX$$

$$\Rightarrow K^{-1}(\lambda x) = X$$

$$\Rightarrow \lambda (K^{-1}x) = X$$

$$\Rightarrow d\,(K^{-1}x) = X$$

We then transform the coordinates of X into the second camera's frame and re-project it onto its image.

$$x' = K[R|t]X$$

$$= dK[R|t]K^{-1}x$$

$$= w(x, [R|t])$$

(b) The photometric error

$$I_1(x) - I_2(w(x, [R|t]))$$

is the difference in the intensity value between $x \in I_1$ and its warped location in I_2. Its nature is non-linear/non-cvx in $[R|t]$.

We can try to recover $[R|t]$ by finding the value of $[R|t]$ that best minimizes the total photometric error. This minimization is carried out using an iterative estimation method like Gauss-Newton or Levenberg-Marquardt, where we start with an initial estimate for $[R|t]$ and iteratively refine it under the assumption that the function is locally linear at each step.

Extra space

Extra space