Time: 30 minutes Max Marks: 10 (UG) / 14 (PG)

Instructions:
- Do not plagiarize. Do not assist your classmates in plagiarism.
- Show your full solution for the questions to get full credit.
- Attempt all questions that you can. Q4 is for extra credit only for UG students. PG students should be solving all three questions.
- In the unlikely case a question is not clear, discuss it with an invigilator. Please ensure that you clearly include any assumptions you make, even after clarification from the invigilator.

1. (1+1+2=4 points) [a] Why is normalization required for fundamental matrix (**F**) estimation? [b] Give the transformation that you would apply to normalize the image points before estimating **F**? [c] What is the rank of the Essential matrix? Justify your answer?

Solution: [a] Normalization for Fundamental matrix estimation is required because the epipolar constraint ($\mathbf{x'}^\top \mathbf{F}\mathbf{x} = 0$ is written in terms of pixel coordinates, the range of which depends on the image resolution. Since there are both linear as well as quadratic terms involving pixel coordinates, the coefficient matrix can be severely ill-conditioned. See the screenshot in Fig. 1. To improve the conditioning of the coefficient matrix, normalization is essential, which can be done by centering the pixels independently in each of the two images and scaling the average length or scaling the standard deviation of the pixel's distance from the origin or scaling the farthest pixel's distance from the origin to be a constant (e.g., 1). See the slide screenshot in Fig. 2 showing the similarity transformation applied to each image.

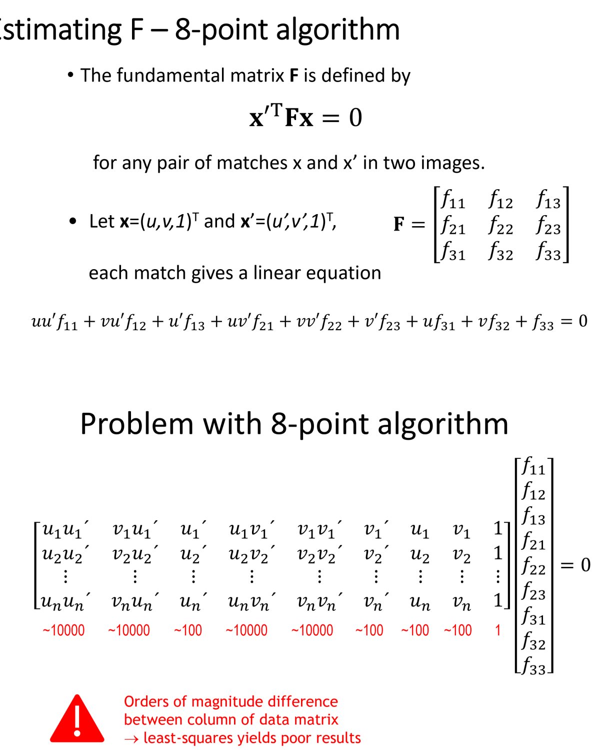

Fig 1.

[b] The transformation required for normalizing the data is a similarity transformation comprising of scaling and centering. Depending on the scaling statistic, i.e., unit average length / standard deviation / maximum value, the diagonal elements will change and accordingly the translation (3^{rd} column) will

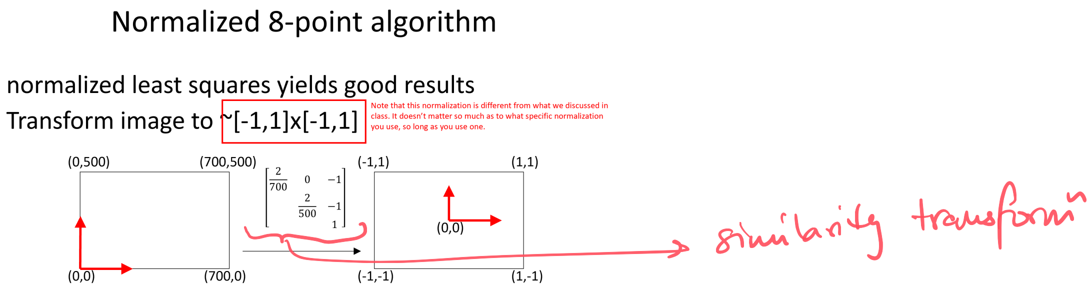

Fig. 2

change. See the slide screenshot in Fig. 2 for an example showing the similarity transformation applied to each image.

[c] The rank of the Essential matrix is strictly 2. It is the product of a skew-symmetric matrix (the cross-product operator of the translation vector) and a rotation matrix. The former has a 1-dimensional null-space, while the latter is full-rank. The product of the two matrices makes the Essential Matrix ($\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$) rank 2.

2. (1+1+3+1=6 points) [a] Write the epipolar constraint using the Essential matrix. Define each element in the expression. [b] Write the expression of the Essential matrix in the form of the rotation and translation between the two cameras. [c] Write the expression of the Essential matrix in the case of a pure translation (i.e., $\mathbf{R} = \mathbf{I}$) along the Y-axis of the first camera. What can you say about the epipolar lines? What can you say about the epipoles in this special case? (Show derivations for full credit) [d] How do the fundamental and the essential matrices relate to each other? Assume different intrinsic matrices for both cameras.

Solution: [a] The epipoloar constraint using the Essential matrix is given as:

$$\widetilde{\mathbf{x}}'^\top \mathbf{E} \widetilde{\mathbf{x}} = 0$$

where $\mathbf{E}$ is the Essential matrix, and $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{x}}'$ are the corresponding points in the normalized pixel homogeneous coordinates, i.e., $\widetilde{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{x}$, where $\mathbf{x}$ is the homogeneous representation of the pixel coordinates and $\mathbf{K}$ is the intrinsic camera matrix. The variable $\widetilde{\mathbf{x}}'$ is defined similarly for the second image with $\widetilde{\mathbf{x}}' = \mathbf{K}'^{-1}\mathbf{x}'$.

[b] $\mathbf{E} = [\mathbf{t}_\times]\mathbf{R}$ where the matrix $[\mathbf{t}_\times]$ is the skew-symmetric cross product matrix for the translation vector $\mathbf{t}$. The matrix $\mathbf{R}$ is the rotation matrix between the two camera positions.

$$[\mathbf{t}_\times] = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

[c]

$$\mathbf{E} = \begin{bmatrix} 0 & 0 & t_y \\ 0 & 0 & 0 \\ -t_y & 0 & 0 \end{bmatrix} \mathbf{I}$$

An epipolar line ~~are~~ is given by $\ell' = \mathbf{E}\mathbf{x}$, where $\mathbf{x}$ is any pixel in the first image and the line ℓ' is the corresponding epipolar line in the second image.

$$\ell' = \mathbf{E}\mathbf{x} = \begin{bmatrix} 0 & 0 & t_y \\ 0 & 0 & 0 \\ -t_y & 0 & 0 \end{bmatrix} \mathbf{I} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} t_y \\ 0 \\ -u \cdot t_y \end{bmatrix}$$

Here u and v are the pixels in normalized camera coordinates. The line ℓ' has its Y component zeros. This implies that the normal vector is in the XZ plane[1], i.e., the normal to the 2D plane only moves in the XZ plane implying that this 2D plane intersects with the Z=1 plane to always form a vertical line. This further implies that the epipolar lines are always vertical. If all epipolar lines are vertical, regardless of the specific values of u and v, it means that they are parallel to each other, and the epipoles (which is also the point of intersection of all epipolar lines) are at infinity.

[d] $\mathbf{F} = (\mathbf{K}'^{-1})^\top \mathbf{E}\mathbf{K}^{-1}$, where $\mathbf{K}'$ and $\mathbf{K}$ are the intrinsic matrices of the two cameras.

3. (*Extra credit only for UG*: 2+1+1=4 points) For a planar calibration pattern, all the 3D points (e.g., corresponding to chessboard corners) lie on the X-Y plane of the world coordinate frame (defined by the chessboard surface). [a] Use the image formation pipeline and show that the resulting relationships between the chessboard corner points and their corresponding image points in image $\mathbf{I}_1$ are related by a 3×3 homography matrix $\mathbf{H}_1$. [b] Given that the intrinsic matrix $\mathbf{K}$ is an upper triangular matrix, justify why this homography is a projective transformation. [c] Suppose we obtain another image $\mathbf{I}_2$ from a different viewpoint and the corresponding homography between the chessboard corners and their respective image points is $\mathbf{H}_2$. What is the homography that would map the corner points from $\mathbf{I}_2$ to those of $\mathbf{I}_1$ in terms of $\mathbf{H}_1$ and $\mathbf{H}_2$?

Solution:

[a] $P_i = [X_i, Y_i, 0]^T$ $i=1,\dots,N$, $P_i \in \mathbb{R}^3$

Corresponding image points $\tilde{P}_i = [u_i, v_i, w_i]^T$ in homogeneous co-ordinates

From the image formation pipeline, we have the following relationship

$$\tilde{P}_i = K[R \mid t] P_i$$

$K \to$ intrinsic camera matrix (3×3)

$R \to$ rotation matrix (3×3)

$t \to$ translation vector (3×1)

$$\begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ 0 \\ 1 \end{bmatrix}$$

→ homogeneous representation of world 3D points

$$\begin{bmatrix} u_i \\ v_i \\ w_i \end{bmatrix} = \underbrace{\begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{bmatrix}}_{\text{This product leads to a } 3\times3 \text{ matrix that has the last row as the vector}} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix}$$

$\boxed{3\times3 \text{ matrix}}$ ⇒ planar homography

[b] $[r_{31} \; r_{32} \; t_3] \neq [0 \; 0 \; 1]$ ⇒ projective transformation [see Lec-05 slides #24-28]

[1] See Lec-05 slides 8-11 for visualizing the 2D projective space and the representation of lines in homogeneous coordinates as the normal of 2D plane passing through the origin.

[c] $\tilde{p}^{(1)} = K^{(1)} [R^{(1)} \; t^{(1)}] P$ ← Image formation applied to image I_1

($K^{(1)}, R^{(1)}$ & $t^{(1)}$ are intrinsic & extrinsic parameters)

Since P has the form $\begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}$

we have

$$\tilde{p}^{(1)} = K^{(1)} \begin{bmatrix} \vec{r}_1^{(1)} & \vec{r}_2^{(1)} & t^{(1)} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

$$\tilde{p}^{(1)} = H^{(1)} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

{ $\vec{r}_1^{(1)}$ $\vec{r}_2^{(1)}$ are the 1st & 2nd columns of the rotation matrix $R^{(1)}$. These combined with $t^{(1)}$ & $K^{(1)}$ gives the homography $H^{(1)} = H_1$

Similarly

$$\tilde{p}^{(2)} = H^{(2)} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

{ for camera 2 $H^{(2)} = H_2$

Since Homographies are invertible mappings and they map 2D points on a plane to 2D points on another plane, we have

$$[H^{(2)}]^{-1} \tilde{p}^{(2)} = \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

$$\Rightarrow \quad \tilde{p}^{(1)} = H^{(1)} [H^{(2)}]^{-1} \tilde{p}^{(2)}$$

$$\Rightarrow \quad \tilde{p}^{(2)} = H^{(2)} [H^{(1)}]^{-1} \tilde{p}^{(1)}$$

The above eq^n relates the corner points in I_1 & I_2 via the 2 homographies. H_1 & H_2