Question 1 a)

~~The epipole of any point e as~~

For any point x, the epipolar line $l' = Fx$ contains the epipole e'.

$$\Rightarrow \quad (e'^T) l' = 0$$

$$(e'^T) Fx = 0 \qquad \forall x$$

$$\Rightarrow \quad (e'^T F) x = 0 \quad \Rightarrow \quad e'^T F = 0$$

This shows that e' is the left null space of F.
Similarly it can be shown that e is the right null space of e.

$$\Rightarrow \quad \boxed{Fe = 0}$$

Question 1 b)

$$x^T F x' = 0$$

$$(x^T F x')^T = 0^T$$

$$x'^T F^T x = 0$$

$\Rightarrow F^T$ is the fundamental matrix between I_2 and I_1.

Reason: F contains information about relative orientation of I_2 wrt I_1 and since orientations are relative, so $F_{12} \neq F_{21}$.

a) Question 1 c)

Fundamental matrix is used for uncaliberated camera pairs while essential matrix is used for caliberated camera pair.

$$x^T F x' = 0$$

$$\underbrace{x^T (K)^{-T} (R)^{-T} S_b (R')^{-1} (K')^{-1} x'}_{F} = 0$$

K, K' is the caliberation matrix for camera 1 and 2.

$$x^T K^{-T} \underbrace{R^{-T} S_b R'^{-1}}_{E} K'^{-1} x' = 0$$

$$\therefore \quad E = R^{-T} S_b R'^{-1}$$

$$\Rightarrow \boxed{F = K^{-T} E K'^{-1}}$$

PAGE No.

DATE / /

Question 2 a)

Consider a world coordinate X and its image coordinate in I_1 is x and I_2 is x'.

$$x = PX \quad , \quad x' = P'X$$

Considering camera 1 as the base frame. $\Rightarrow R = I$ and $t = 0$ (by definition)

$$x = K[I|0]\begin{bmatrix} X \\ 1 \end{bmatrix}$$

$$\Rightarrow \quad x = KX$$

Camera 2's frame has only a rotation R from base frame

$$\therefore \quad x' = K'[R|0]\begin{bmatrix} X \\ 1 \end{bmatrix}$$

$$\Rightarrow x' = K'RX$$

$$x' = K'R\underbrace{K^{-1}x}_{X}$$

$$x' = (K'RK^{-1})x$$

Or $\quad x = (K'RK^{-1})^{-1}x' = KR^{-1}K'^{-1}x'$

PAGE No.	
DATE	/ /

Thus, the homography relation is given as:

$$\boxed{x = [KR^{-1}K'^{-1}]x'} \Rightarrow \underbrace{x = H^{-1}x'}$$

homography matrix.

* If translation is involved, such a homography relation won't be involved.

$$x = K[I|0]\begin{bmatrix} X \\ 1 \end{bmatrix}$$

$$x' = K[R|-X_0]\begin{bmatrix} X \\ 1 \end{bmatrix}$$

Translation.

$$\Rightarrow x' = K'R(X - X_0)$$

$$x' = K'R(K^{-1}x - X_0)$$

Clearly, this doesn't match the homography relation.

In this case, there is an ambiguous scaling factor which leads to subtraction of X_0 from $K^{-1}x$ and this leads to multiple possible x which relates to x' due to the unknown depth factor. Thus we get a line (epipolar) instead of a direct point x.

PAGE No.	
DATE	/ /

Question 2 b)

There are 3 different methods to map pixels in image 1 to image 2.

(i) Fundamental matrix
(ii) Essential matrix
(iii) Homography matrix

1) Fundamental matrix is used for a pair of uncaliberated cameras.
~~lle~~ It has 11 parameters which describe the mapping.

$$\boxed{x^T F x' = 0}$$

$\uparrow$
fundamental matrix.

Assumption: The rotation ~~betwe~~ and translation between the two viewing positions must be small.
This is because the smaller the translations and rotations are, the more number of common world points can be found.

PAGE No.

DATE / /

2) Essential matrix :

We use essential matrix when we know the caliberation matrix for the cameras, ie K and K'.

$$\cancel{x^T K^{-T} E K' x' = 0} \qquad \boxed{x^T K^{-T} E K' x' = 0}$$

Essential matrix.

Assumptions : (i) The cameras are caliberated and we know the Caliberation matrices K and K'.

(ii) The rotation angle and translation distance must be small.

3) Homography matrix :

This is used when there is only rotation involved and no translation.

$$\boxed{x = Hx'}$$

Assumptions : (i) Must be captured from same camera.

(ii) No translation involved.

(iii) Only rotation is involved.

(iv) Both images capture same planar surface.

PAGE No.
DATE / /

Question 2 c)

Stereo rectification has 2 homographies involved.
(i) Right camera
(ii) Left camera.

* We use homography of left camera to make epipolar lines parallel and bring epipoles to infinity.

* We use homography of right camera to align right camera with respect to left and then make epipolar lines // and bring epipoles to ∞.

⇒ These are called as homographies because we are dealing with images captured from same cameras and only rotations involved and no translation.

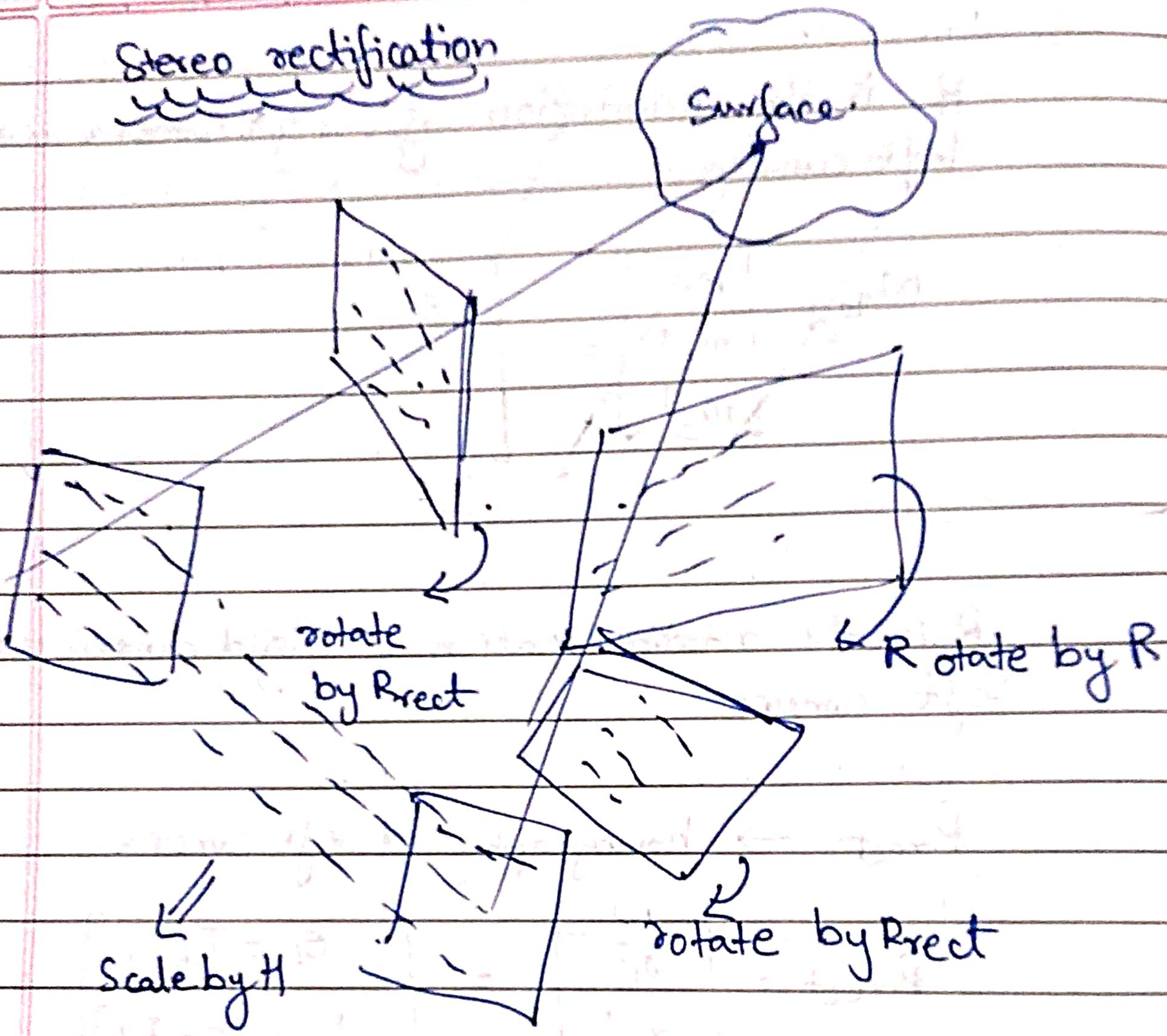
R is the rotation matrix of right camera wrt left camera.

$R_{rect} \rightarrow$ homography of left camera.

$$R_{rect} = \begin{bmatrix} r_1^T \\ r_2^T \\ r_3^T \end{bmatrix} = \begin{bmatrix} e_1 = T/\|T\| \\ \frac{1}{\sqrt{T_x^2 + T_y^2}} [-T_y \; T_x \; 0] \\ r_1 \times r_2 \end{bmatrix}$$

homography of right camera = $R R_{rect}$

Stereo rectification



PAGE No.
DATE / /

Question 3)

We use DLT for estimating the projection matrix P from known correspondances between the image coordinates and world points.

$$x = PX$$

image coordinates — x; Projection matrix — P; World points — X.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

{In homogeneous coordinates}

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix}$$

$$x = \frac{P_{11}X + P_{12}Y + P_{13}Z + P_{14}}{P_{31}X + P_{32}Y + P_{33}Z + P_{34}}$$

$$y = \frac{P_{21}X + P_{22}Y + P_{23}Z + P_{24}}{P_{31}X + P_{32}Y + P_{33}Z + P_{34}}$$

$$P = KR\left[I_3 \mid -X_0\right]$$

K: c, s, m, x_H, y_H — 5 parameters

R: 3 rotation parameters.

X_0: 3 translation parameters.

Since there are 11 unknown parameters for uncaliberated cameras, we need atleast 6 points to estimate parameters.

⇒ For each corresponding pair $\{x_i, X_i\}$ we have

$$x_i = P X_i \qquad i = 1, 2, \dots 6 \{\text{or more}\}$$

$$x_i = \begin{array}{c} \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} \end{array} \begin{array}{l} \rightarrow A^T \\ \rightarrow B^T \; X_i \\ \rightarrow C^T \end{array}$$

$$A = \begin{bmatrix} P_{11} \\ P_{12} \\ P_{13} \\ P_{14} \end{bmatrix} \qquad B = \begin{bmatrix} P_{21} \\ P_{22} \\ P_{23} \\ P_{24} \end{bmatrix} \qquad C = \begin{bmatrix} P_{31} \\ P_{32} \\ P_{33} \\ P_{34} \end{bmatrix}$$

$$\Rightarrow \quad x_i = \frac{A^T X_i}{C^T X_i} \qquad y_i = \frac{B^T X_i}{C^T X_i}$$

$$x_i C^T X_i - A^T X_i = 0 \quad \dots\dots \text{eq 1}$$

$$y_i C^T X_i - B^T X_i = 0 \quad \dots\dots \text{eq 2}$$

We can write eq. 1, 2 as linear combination of A, B, C vectors, by unrolling them into a long vector p.

$$p = \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} p_{11} \\ p_{12} \\ \vdots \\ p_{21} \\ \vdots \\ p_{31} \\ \vdots \\ p_{34} \end{bmatrix}_{12\times 1}$$

$$\Rightarrow \text{eq } 1 = [-X_i, -Y_i, -Z_i, -1, 0, 0, 0, 0, x_iX_i, x_iY_i, x_iZ_i, x_i]\, p$$

$$\text{eq } 2 = [0, 0, 0, 0, -X_i, -Y_i, -Z_i, -1, y_iX_i, y_iY_i, y_iZ_i, y_i]\, p$$

Let eq 1 $= a^T_{x_i} p$ and eq 2 $= a^T_{y_i} p$

We can stack all the 6 points into 12×12 matrix.

$$\begin{bmatrix} a_{x_1}^T \\ a_{y_1}^T \\ a_{x_2}^T \\ \vdots \\ a_{x_6}^T \\ a_{y_6}^T \end{bmatrix}_{12\times 12} p = 0 = M_{12\times 12}\, p = 0$$

PAGE No.
DATE / /

To solve a system of linear equations of form

$Ax = 0$, is equivalent to find null space of A. which can be found by solving for SVD of A and, choosing x as singular vector belonging to singular value 0.

∴ We calculate the SVD of $M_{12\times12}$ and choose p as the last column in V^T.

$$M_{12\times12} = U_{12\times12} \; S_{12\times12} \; V^T_{12\times12}$$

Last column in V^T ie v_{12}

$$p = v_{12} = \begin{bmatrix} p_{11} \\ p_{12} \\ \vdots \\ p_{34} \end{bmatrix}$$

we convert p to projection matrix P by reshaping it into 3x4 matrix.

$$p = \begin{bmatrix} p_{11} \\ p_{12} \\ \vdots \\ p_{34} \end{bmatrix} \longrightarrow P = \begin{bmatrix} p_{11} & p_{12} \cdots & p_{14} \\ p_{21} & \cdots & p_{24} \\ p_{31} & \cdots & p_{34} \end{bmatrix}$$

FAILING CONDITION

This algorithm would fail if all points are on a plane.

Consider all points lying on some $Z = z$

$$\text{Consider} \begin{bmatrix} a^T_{x_i} \\ a_{y_i} \end{bmatrix} = \begin{bmatrix} -X_i & -Y_i & -Z_i & -1 & 0 & 0 & 0 & 0 & x_iX_i & x_iY_i & zx_i & x_i \\ 0 & 0 & 0 & 0 & -X_i & -Y_i & -Z_i & -1 & y_iX_i & y_iY_i & zy_i & y_i \end{bmatrix}$$

These three columns can be easily converted into zero columns using some operations.

$$= \begin{bmatrix} -X_i & -Y_i & 0 & -1 & 0 & 0 & 0 & 0 & x_iX_i & x_iY_i & 0 & x_i \\ 0 & 0 & 0 & 0 & -X_i & -Y_i & 0 & 1 & y_iX_i & y_iY_i & 0 & y_i \end{bmatrix}$$

$M_{12\times 12}$ becomes a rank deficient matrix and thus no solution exists if all x_i lie on same plane.

PAGE No.
DATE / /

We found that $Mp = 0$
but $Mp \neq 0$ because of errors/redundant observations.

$$\rightarrow \quad Mp = \Delta K$$

we want p such that $\Delta K^T \Delta K$ is minimized.

$$\hat{p} = \underset{p}{\arg\min}\ \Delta K^T \Delta K$$

$$= \underset{p}{\arg\min}\ p^T M^T M p.$$

We can clearly see that this value will be minimized by a singular vector belonging to the smallest singular value.

$\therefore\ p = V_{12}$ {last column of V^T corresponding to singular vector corresponding to smallest singular value},

PAGE No.
DATE / /

Question 4 a)

m observations are given to us and we know the relative ~~pose~~ transformation between any two pose.

Let the poses of robot be: $\{X_0, X_1, \ldots X_{m-1}\}$

We also have the control signals between observation i and j.

u_{ij} to go from i to j and

u_{ji} to go from j to i

Each of the pose is a ~~3x1~~ ~~3x~~ 3x1 vector

$$X_i = \begin{bmatrix} X_{xi} \\ X_{yi} \\ X_{\theta_i} \end{bmatrix} \quad \{2D \text{ plane}\}$$

$$u_{ij} = \begin{bmatrix} \Delta\phi_{ij} & t_{ij} \end{bmatrix}^T$$

angle translation.

$$\hat{X}_j = f(X_i, u_{ij})$$

Our goal is to minimize the error between X_j and $\hat{X}_j$.

$\therefore$ Cost function

$$\sum_{i=0}^{m-1} \sum_{\substack{j=0 \\ j \neq i}}^{m-1} \| f(\hat{X}_i, u_{ij}) - X_j \|^2$$

$\because$ $f(\hat{X}_i, u_{ij})$ is a highly non linear function

we want to linearize it.

$$f(\hat{X}_i, v_{ij}) \approx \underbrace{f(X_i, v_{ij})}_{\text{initial estimate}} + \underset{\substack{\uparrow \\ \text{Jacobian}}}{F_{ij}} \delta \hat{X}_i$$

Cost function $=$

$$\sum_{i=0}^{m-1} \sum_{\substack{j=0 \\ j \neq i}}^{m-1} \| f(X_i, v_{ij}) + F_{ij} \delta \hat{X}_i - X_j \|^2$$

$\underbrace{f(X_i, v_{ij}) - X_j}_{a_{ij}}$ is the odometry error

$\therefore$ Cost function $= \displaystyle\sum_{i=0}^{m-1} \sum_{\substack{j=0 \\ j \neq i}}^{m-1} \| F_{ij} \delta \hat{X}_i - a_{ij} \|^2$

Minimizing this would give us optimized trajectory.

PAGE No.
DATE / /

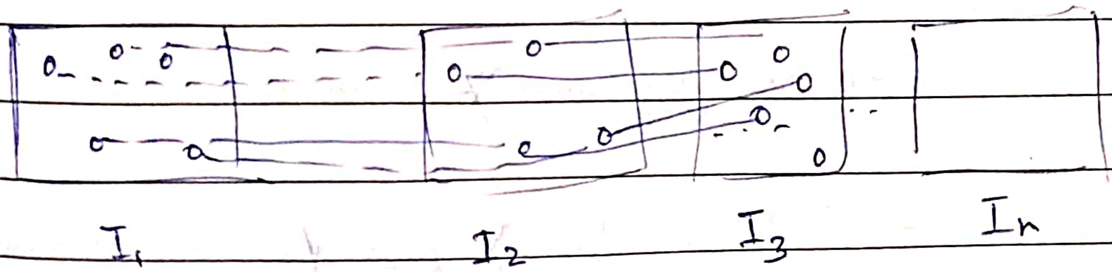
Question 4b)

Given a set of images $\{I_1, I_2, \dots I_n\}$, if we want to obtain the pose estimates and optimise the trajectory, this is Structure from motion problem.

1) The first step would be to compute features in the images which are likely to be invariant across all images (or atleast most of them)

This can be done using feature detectors such as
- SIFT
- ORB
- Hessian-Laplacian

Now we have features in the images.



2) Next step is to find the correspondances between these features as shown in the illustration above.

This can be done by matching the feature descriptors via approximate nearest neighbour.

PAGE No.
DATE / /

3) Now that we have correspondance we need to compute the 3D world points and the projection matrix.

Initial guess:

No additional information given so can't use triangulation ~~for~~ as it need Projection matrix.

Consider m scenes and n points in total (features we have extracted)

P_j is projection matrix of j^{th} scene.
X_{ij} is 3D word point of i^{th} feature in image in j^{th} view
x_{ij} is the i^{th} feature in j^{th} view.

$$\boxed{x_{ij} = P_j \cdot X_{ij}}$$

Estimate P_j and X_is to minimize:

$f(P_1, P_2, \ldots P_m, X_1, X_2, \ldots X_n) \Leftarrow$ loss function.

This will be minimized for.

$$\underset{P_i, X_j}{\arg\min} \sum_{i=1}^{M} \sum_{j=1}^{N} (w_{ij}) \| P_i X_j - x_{ij} \|^2$$

$$w_{ij} = \begin{cases} 1 & \text{if } j^{th} \text{ point seen in } i^{th} \text{ frame.} \\ 0 & \text{if not seen.} \end{cases}$$

PAGE NO.

DATE / /

$$\rightarrow \underset{P_i, X_j}{\operatorname{argmin}} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\left[\frac{P_{i11} X_j + P_{i12} Y_j + P_{i13} Z_j + P_{i14}}{P_{i31} X_j + P_{i32} Y_j + P_{i33} Z_j + P_{i34}} - x_{ij} \right]^2 + \left[\frac{P_{i21} X_j + P_{i22} Y_j + P_{i23} Z_j + P_{i24}}{P_{i31} X_j + P_{i32} Y_j + P_{i33} Z_j + P_{i34}} - y_{ij} \right]^2 \right)$$

This is reprojection error.

$P_i X_j = \hat{x}_{ij}$ {estimate of j^{th} point in i^{th} scene}

$$\underset{X_j, P_i}{\operatorname{argmin}} \sum_{i=1}^{M} \sum_{j=1}^{N} \| \hat{x}_{ij} - x_{ij} \|^2$$

~~Using this, we will have an initial estimate for 3D world points for Bundle adjustment.~~

~~We didn't have projection matrix so we can't use triangulation for initial estimates.~~

4) This can be solved using LM-algorithm.

$$(J^T J + \lambda I) \Delta k = -J^T r(k)$$