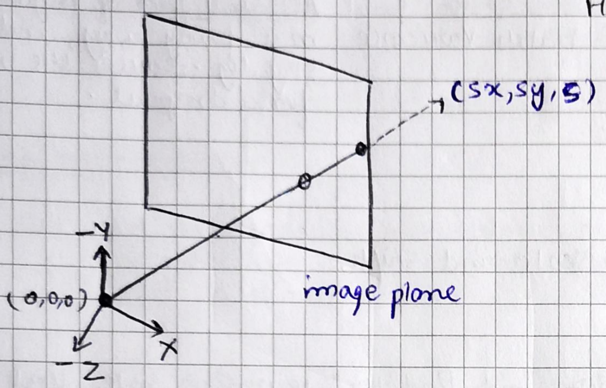
Epipolar Geometry.

→ The projective plane: why do we need homogeneous coordinates?
(a) Represents points at infinity.
(b) Homographies
(c) Perspective projection
(d) multi-view relationship.

The geometric intuition? A point in the image is a ray in the projective space.



Homogeneous Coordinates in P^2

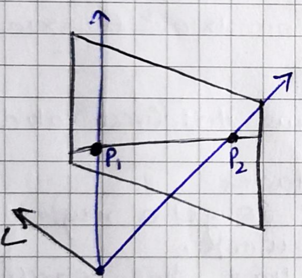
$(sx, sy, s) = (x, y, 1) \longleftrightarrow (x, y)$ $\quad s \neq 0$

Eucleidian Coordinates in R^2 $(s=1)$

$(x, y, 0) \longrightarrow (\infty, \infty)$ i.e. direction.

Point at ∞ $(s=0)$

→ what does a line in the image correspond to in image space?



(a) A line is a plane of rays through origin
→ all rays (x,y,z) satisfy equation:
$ax + by + cz = 0$

→ in vector notation $0 = \underbrace{[a \; b \; c]}_{L^T} \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{P}$

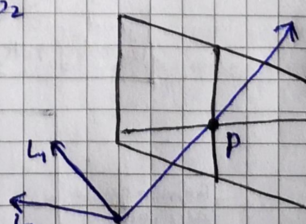
(b) A line is also represented as a homogeneous 3-vector L
→ It is perpendicular to every point (ray) p on the line $L^T P = 0$

(c) what is the line L spanned by rays P_1 and P_2?
→ L is ⊥r to p_1 and p_2, L is the plane normal.
$L = P_1 \times P_2$
→ cross product of p_1 and p_2 will give me L. which will define a vector normal to a plane.

(d) what is the intersection of two lines L_1 and L_2?

→ As we had L being perpendicular to p_1 and p_2 when we have two lines intersecting at the same point, it can be represented as

$p = L_1 \times L_2$

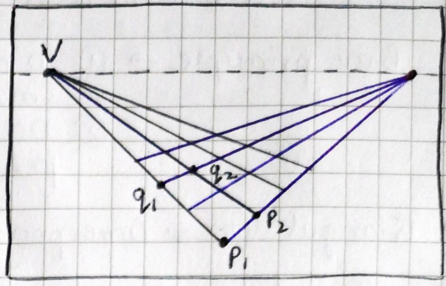


→ p is ⊥r to L_1 and L_2

→ Duality behaviour (lines and points)

(c) Points and lines are dual in projective space
→ given any formula, can switch the meaning of points and lines to get another formula.

(d) Eg. computing the vanishing point(s)



→ Intersect $p_1 q_1$ with $p_2 q_2$

$$V = (p_1 \times q_1) \times (p_2 \times q_2)$$

→ In practise, least squares version
→ it is better to use more than 2 lines and compute the "closest" point of intersection.

→ Cross-product as matrix operation →

(a) from a 3 element vector, a skew-symmetric matrix is defined as

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \qquad [a]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

(b) This allows to formulate the cross product of two vectors as matrix multiplication.

$$a \times b = [a]_\times \cdot b$$
$$= \left(a^T \cdot [b]_\times\right)^T$$

→ Intersection of parallel lines
let us suppose, we have 2 lines (L, L') which are parallel. If we compute the intersection of these 2 lines we will get a vector which will represent a point at infinity.

$L = (a, b, c)^T \qquad L' = (a, b, c')^T \qquad L \times L' = (b, -a, 0)^T$

$$\underbrace{[L]_\times}_{\text{skew matrix of } L} L' = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c' \end{bmatrix} = \begin{bmatrix} b(c'-c) \\ a(c-c') \\ 0 \end{bmatrix} \approx \begin{bmatrix} b \\ -a \\ 0 \end{bmatrix}$$

skew matrix of L

$(b, -a)$ → Tangent vector
(a, b) → Normal direction

→ Epipolar Geometry - Scenarios

(a) A stereo rig consisting of two cameras
→ 2 images are acquired simultaneously.

(b) A single moving camera (static scene)
→ Two images are acquired sequentially

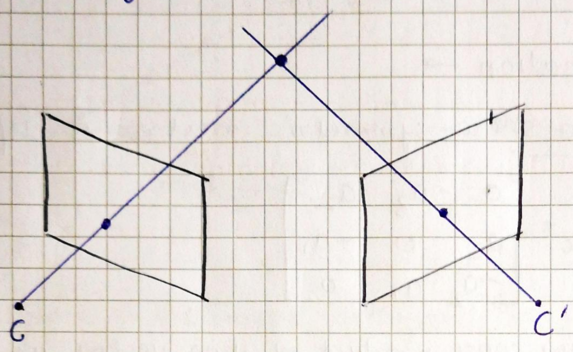
(c) The above 2 scenarios are geometrically equivalent.

→ Epipolor Geometry → objective
The objective of Epipolor geometry is "given 2 images of a scene captured by known cameras, compute the 3D position of the scene (structure recovery)

Basic principle → ① 3D points are triangulated from corresponding image points.
② Determine 3D points at intersection of 2-back-projected rays.

→ Triangulation – Correspondences

① Corresponding points are images of the same scene point.

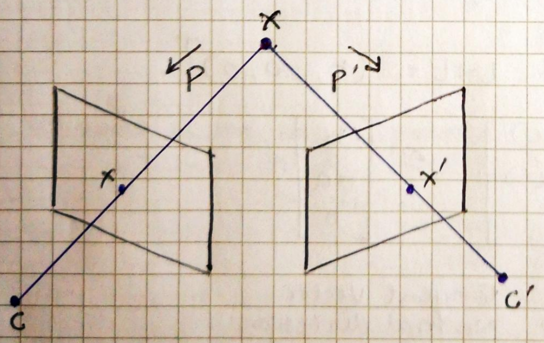


* The back projected points generate rays which intersect at 3D scene point.

② For each point in the 1st image, determine the corresponding point in the 2nd image. ⇒ This is a search problem.

③ For each pair of matched points, determine the 3D point by triangulation. ⇒ This is an estimation problem.

→ Notation ① The two cameras with projection matrices P and P' and a 3D point X with the corresponding 2D points in the images (x and x')



$x = PX$
$x' = P'X$

$P = 3 \times 4$ Camera matrix
X = 4D vector
x = 3D Vector

→ Epipolor line → Given an image point in 1 view, where is the corresponding point in the other view?

① If we have a point in our 1st image and if i assume that i know the "projection matrix" (P) and (P') on the other side. we have to see that how is this point mapped in our 2nd image.

② The point in the 1st image is basically a representation of a 3d point which can be represented by a ray (in the 1st image) called "back projection".

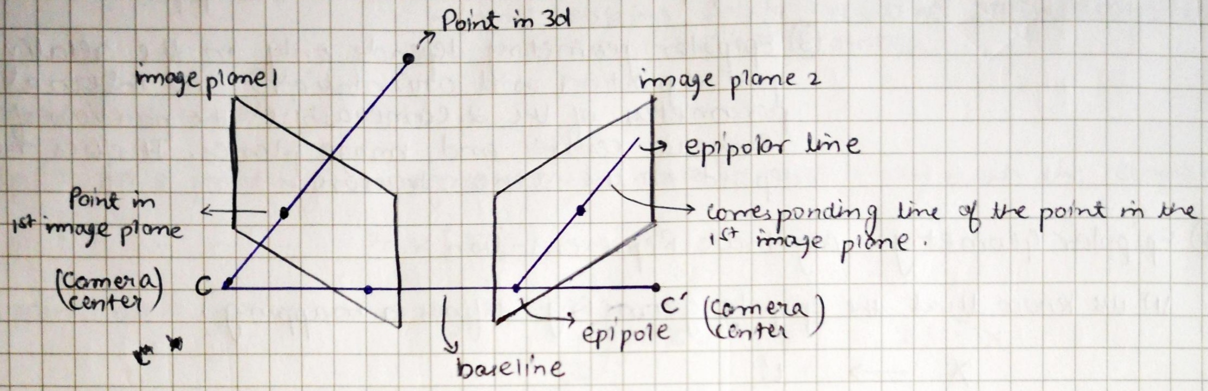
③ If we reproject this in our 2nd image, we get a line called "Epipolor line".

④ Knowing the projection matrix of 2 cameras, the point maps to a line. (epipolar line) in the other view.

⑤ A Baseline is a line b/w our 2 camera centres. The baselines cuts the 2nd camera image plane at the epipolar line. called the epipole.

⑥ If we move our 3d point or change it, the baseline will still cut the image plane at epipole.

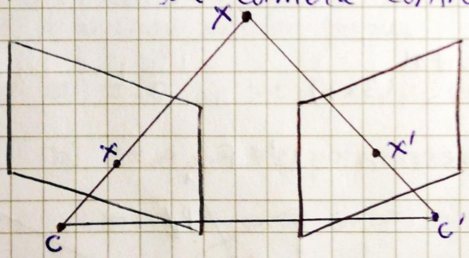
⑦



⑧ Our point in the 1st image planes reduces to a line in the corresponding image.

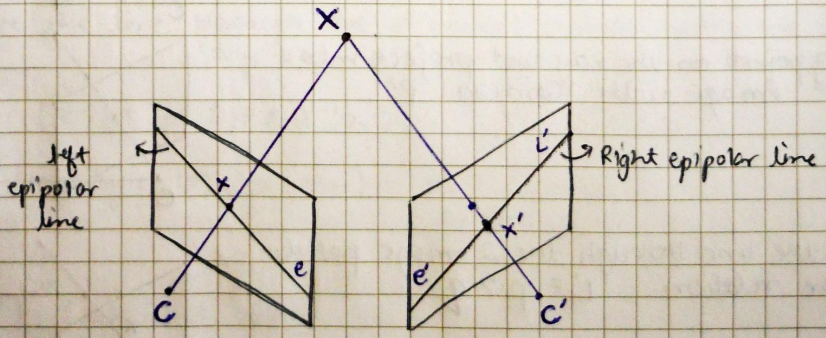
⑨ Epipolar constraint → (1) Reduces the correspondence problem to 1D search along an epipolar line.

⑩ Epipolar plane → Epipolar geometry is a consequence of the coplanarity of the camera centres and scene point.



The camera centres, corresponding points and scene point lie in a single plane, known as the "epipolar plane".

⑮ Nomenclature →



① The epipolar line l' is the image of the ray through x.
② The epipole "e" is the point of intersection of the line joining the camera centres with the image plane
③ → This line is the baseline for a stereo rig.
→ The translation vector for a moving camera.

③ The epipole is the image of the center of the other camera:

$$e = PC' \qquad e' = P'C$$

⑯ The epipolar pencil → ① As the position of the 3D point varies, the epipolar planes "rotate" about the baseline. This family of planes is known as "Epipolar pencil".

② All epipolar lines intersect at the epipole.

③ Epipolar geometry depends only on the relative pose (position and orientation) and internal parameters of the 2 cameras i.e. the position of the camera centres and image planes. It does not depend on the scene structure.

⑰ Epipolar Geometry - Algebraic Representation

(1) We know that the epipole geometry defines a mapping:

$x \longrightarrow l'$

x → point in the first image

l' → epipolar line in 2nd image.

(2) The map only depends on the cameras P, P' (not on structure of the scene)

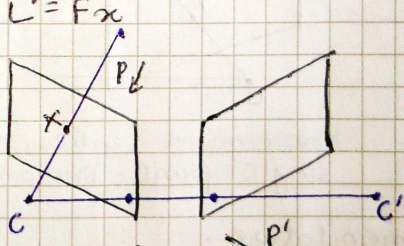
(3) It will be shown that the map is linear and can be written as

$$l' = Fx$$

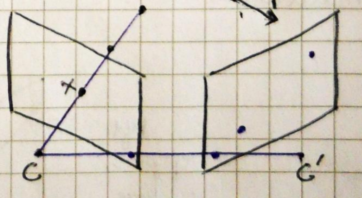
F → fundamental matrix (3x3)

(4) Derivation of the algebraic expression $l' = Fx$

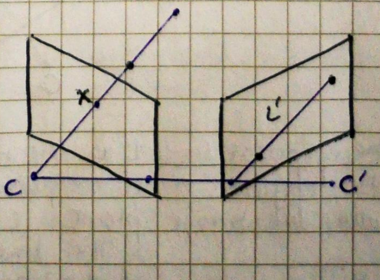
① For a point x in the first image, back project a ray with camera P



② Choose 2 points on the ray and project into the 2nd image with camera P'.



③ Compute the line through the 2 image points using the relation $l' = p \times q$



④ Recalling the camera matrices:

$$P = K[R|t]$$

Calibration matrix — K; Rotation — R; Translation — t: from world to camera coordinate frame

First camera $P: K[I|0]$

→ World coordinate frame aligned with 1st camera.
→ Now looking for the relative orientation b/w the 1st and the 2nd camera.

Second camera $P' = K'[R|t]$

⑤ Step 1: For a point x in the first image, back project a ray with the camera

$$P = K[I|0]$$

a point back projects to a ray: $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = ZK^{-1}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = ZK^{-1}x$

Point Depth (Z)

since $X(Z) = \begin{pmatrix} ZK^{-1}x \\ 1 \end{pmatrix}$

satisfies $PX(Z) = K[I|0]X(Z) = x$

Step 2: Choose 2 points on the ray and project onto the 2nd image with camera P'.

We can choose any points along the ray but we use a trick to identify the 2 already known points such as: (fully defined points)

(a) $Z=0$ is the camera center $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$

(b) $Z=\infty$ is the point at infinity $\begin{pmatrix} K^{-1}x \\ 0 \end{pmatrix}$

Projecting these 2 points onto the 2nd view

$$P'\begin{pmatrix} 0 \\ 1 \end{pmatrix} = K'[R|t]\begin{pmatrix} 0 \\ 1 \end{pmatrix} = K't \quad , \quad P'\begin{pmatrix} K^{-1}x \\ 0 \end{pmatrix} = K'[R|t]\begin{pmatrix} K^{-1}x \\ 0 \end{pmatrix} = K'RK^{-1}x$$

Step 3: Compute the line through the 2 image points using the relation

$$l' = p \times q$$

$$l' = (K't) \times (K'RK^{-1}x)$$

using the identity: $(Ma) \times (Mb) = M^{-T}(a \times b)$

$$M^{-T} = (M^{-1})^T = (M^T)^{-1}$$

$$l' = (K')^{-T}(t \times (RK^{-1}x))$$

$$= \underbrace{K'^{-T}[t]_\times RK^{-1}}_{F}x$$

$\boxed{l' = Fx}$ The fundamental matrix which maps a point in the 1st camera image plane to a line in the 2nd camera image plane.

→ Points x and x' correspond ($x \leftrightarrow x'$), then $x'^T l' = 0$

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

⑱ The fundamental matrix "F".

F is a unique 3×3 rank 2 matrix that satisfies $x'^T F x = 0$ for all $x \leftrightarrow x'$

(a) Epipolar lines: $l' = Fx$ and $l = F^T x'$

(b) Epipoles: on all epipolar lines, thus $e'^T F x = 0 \;\; \forall x \Rightarrow e'^T F = 0$, similarly $Fe = 0$

(c) F has 7 DOF., i.e. 3×3 − 1 (homogeneous, upto scale) − 1 (rank(F)=2)

(d) F is a correlation, projective mapping from a point x to a line $l' = Fx$ (not a proper correlation i.e. not invertible)

(e) F relates (homogeneous) pixel coordinates of corresponding points.

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

(f) The transformation "Fx" provides all potential positions of corresponding point x' (F has rank 2)

$$l' = Fx$$

(g) Homogeneous coordinates x' of points on the line l' are computed in normal form by:

$$l'^T x = (l'_x \;\; l'_y \;\; l'_z)\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = l'_x x' + l'_y y' + l'_z \cancel{1} = 0$$

which provides the linear form

$$y'(x') = \frac{-l'_x \cdot x' - l'_z}{l'_y}$$

⑲ Essential matrix

Fundamental matrix: for points in pixel coordinates

$$(x'_{pix} \;\; y'_{pix} \;\; 1)\, F \begin{pmatrix} x_{pix} \\ y_{pix} \\ 1 \end{pmatrix} = 0$$

Essential matrix: for points in normalised image coordinates

$$(x'_n \;\; y'_n \;\; 1)\, E \begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} = 0$$

* Normalised image coordinates is a transformation b/w pixel coordinates, and camera coordinates using calibration matrix K

$$x_{pix} = K x_n \Rightarrow x_n = K^{-1} x_{pix}$$

$$x'^T_{pix} F x_{pix} = (K' x'_n)^T F (K x_n) = x'^T_n K'^T F K x_n = x'^T_n E x_n \qquad \boxed{E = K'^T F K}$$

moreover, we know that $F = K'^{-T}[t]_x R K^{-1}$

$$\boxed{\mathcal{E} = [t]_x R}$$

Another method to calculate the essential matrix is by "coplanarity constraint".

① Epipolar lines result from coplanarity constraint:

$$(x' \times t) \cdot x = (x' \times t)^T x = \begin{vmatrix} x' & t_x & x \\ y' & t_y & y \\ z' & t_z & z \end{vmatrix} = \det \begin{pmatrix} x' & t_x & x \\ y' & t_y & y \\ z' & t_z & z \end{pmatrix} = 0$$

→ x, x' and t form a plane, therefore the coplanarity equation holds true.
→ Note that the vectors x, x' and t should be in the same coordinate system (we choose the one of camera ($'$))

② Rotation b/w different normalized coordinate systems of cameras (c to c')

$$x_n^{c'} = R x_n = R \begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} \qquad x'_n = \begin{pmatrix} x'_n \\ y'_n \\ 1 \end{pmatrix}$$

↳ x_n expressed in coordinate system of c'

Due to coplanarity equation:

$$(x'_n \times t)^T x_n^{c'} = 0$$

$$\Leftrightarrow (x'_n \times t)^T R x_n = 0$$

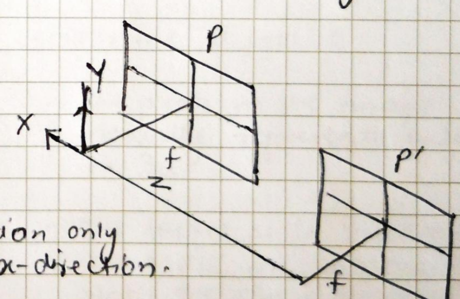
$$\Leftrightarrow x'^T_n [t]_x R x_n = 0 \qquad \Rightarrow \mathcal{E} = [t]_x R$$

(50) Compute the fundamental matrix for a parellel camera stereo rig:

→ $P = K[I|0]$ $\qquad P' = K'[R|t]$

→ $K = K' = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $\qquad R = I \qquad t = \begin{pmatrix} t_x \\ 0 \\ 0 \end{pmatrix}$

Translation only in the x-direction.



$$F = K'^{-T}[t]_x R K^{-1} = \begin{bmatrix} 1/f & 0 & 0 \\ 0 & 1/f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_x \\ 0 & t_x & 0 \end{bmatrix} \begin{bmatrix} 1/f & 0 & 0 \\ 0 & 1/f & 0 \\ 0 & 0 & 1 \end{bmatrix} = \lambda \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\lambda = \frac{t_x}{f^2}$$

For any point, we would have the following eq.

$$x'^T f x = (x' \; y' \; 1) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0$$

→ Reduces to $y = y'$ i.e. raster correspondences. (horizontal scan lines)

→ This also proves that given two parallel stereo cameras, the epipolar lines will be horizontal.

→ F is a rank 2 matrix :

(21) Fundamental matrix F: relative orientation b/w 2 views + camera ~~calibration~~ Parameters
→ 7 DOF (F has rank 2, det(F) = 0)
→ General case: intrinsic parameters not available.

Essential matrix E → relative orientation b/w 2 views.
→ 5 DOF
→ Calibrated case: intrinsic parameters available.