3D Vision

BLG634E

Gözde ÜNAL

04.04.2022

Recap: Camera parameters;

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} f s_x & \overset{\downarrow}{f s_\theta} & \overset{\downarrow}{o_x} \\ 0 & f s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{\underline{\underline{K}} = K_s K_f} \overbrace{\underbrace{\begin{bmatrix} I_{3\times 3} & \underline{0} \end{bmatrix}}_{3\times 4} \underbrace{\begin{bmatrix} \underline{R} & \underline{T} \\ \underline{0}^T & 1 \end{bmatrix}_{4\times 4}}_{\bar{g} \in SE(3)}}^{[\underline{R} \quad t]} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \\ 1 \end{bmatrix}$$

$\underline{\underline{K}} = K_s K_f$ intrinsic calibration matrix

$\Pi_0 : 3\times 4$ perspective projection matrix

Intrinsic Camera Calibration: estimation of parameters of $\underline{\underline{K}}$.

Extrinsic Camera Calibration: estimation of $\bar{g}$.

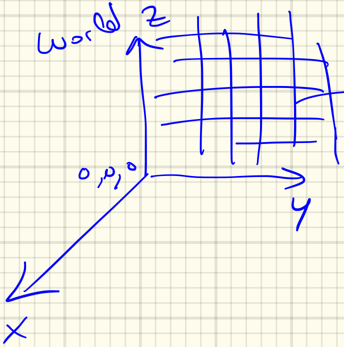
Overall, we have 11 parameters: 5 in $\underline{K}$
6 in $\bar{g}$.

→ We added $\overline{\text{Lens}}$ Distortion, parameters: k_1, k_2

↓ calibration

Zhang's Camera Calibration Method:

Typically



Calibration Object. → Planar : A checkerboard

→ Z=0 plane.

Σ

$$\underline{M} \longrightarrow \underline{\tilde{M}} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\underline{m} \longrightarrow \begin{pmatrix} u \\ v \end{pmatrix} \rightarrow \tilde{m} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$$

$$s \, \tilde{m} = \underline{\underline{A}} \, \underbrace{\begin{bmatrix} \underline{\underline{R}} & \underline{t} \end{bmatrix}_{3\times4}}_{\Pi_0 \, \bar{g}} \, \tilde{M}$$

$$\underline{\underline{H}} = \begin{pmatrix} \underline{h_1} & \underline{h_2} & h_3 \end{pmatrix}$$

$$K = \underline{\underline{A}} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

: Using homography btw the world plane and the image planes

1st estimate $\underline{\underline{A}}$ parameters: → closed for expressions $\left\{ \begin{array}{l} \alpha \\ \gamma \\ \beta \\ u_0 \\ v_0 \end{array} \right.$

↳ our initial estimates for a later nonlinear optimization.

→ Then we get initial estimates for: $\underline{r_1}$, $\underline{r_2}$ → $\underline{r_3} = \underline{r_1} \times \underline{r_2}$

project it onto SO(3) ⇐ $\underline{\underline{R}} = [\, r_1 \;\; r_2 \;\; \underset{\uparrow}{r_3} \,]$ ✓

$\underline{t}$ ✓

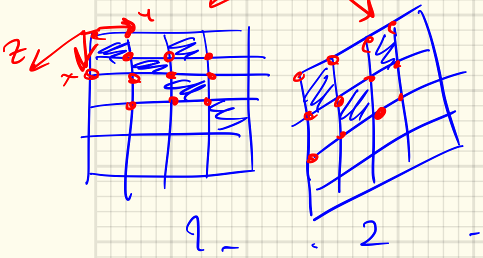
Now we have initial estimates for all intrinsics & extrinsics, use them in an nonlinear optimization:

$$\min J = \sum_{i=1}^{n} \sum_{j=1}^{m} \| \underset{\text{ith image} \swarrow \;\; \searrow \text{jth point}}{m_{ij}} - \hat{m}(\underset{\substack{\uparrow \\ \text{intrinsics} \\ \text{common} \\ \text{for all views}}}{\underline{\underline{A}}}, \underline{\underline{R}}_i, \underline{t}_i, \underline{M}_j) \|^2$$

extrinsics are different for each view i

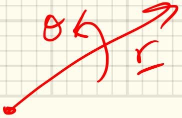
projection of point m_j onto the ith image.

we have n images of the model plane & m points on the model plane



... n views of the calibration object.

$\underline{\underline{R}}$: parameterized by exp. coord., i.e. by axis vector $\underline{r}$



$\| \underline{r} \| = \theta$: its magnitude is the rotation angle

$\Longrightarrow$ min J w.r.t. $\boxed{\underline{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_2 \end{bmatrix}}$ ← changing

$\rightarrow \boxed{\dfrac{\partial J}{\partial r}}$ $\nabla_{\underline{r}} J$ ← GD $\longrightarrow$ $r_{k+1} = r_k - \text{lR}\,\nabla_{\underline{r}} J$

You get an estimate $\boxed{\underline{r}}$.

to go the rotation matrix $\underline{\underline{\mathbb{R}}}$ ← use Rodrigues formula: $\boxed{e^{\hat{r}\theta}} = I + \cos\theta \ldots$ - - -

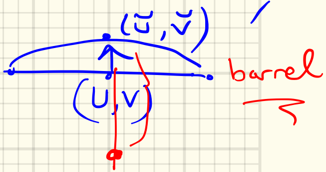
$\longrightarrow$ In Zhang: This nonlinear optim. problem is solved by LM algorithm

$$\underset{\underline{\underline{A}},\, \{\underline{r}_i,\, \underline{t}_i\}_{i=1}^{n}}{\arg\min}\; J$$

Dealing w/ Radial Distortion: Use a lens distortion model

Let (u,v) be the ideal (distortion free) pixel image coordinates,

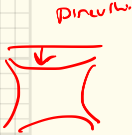
$(\breve{u}, \breve{v})$ the corresponding real (observed) pixel coord.



(x,y) :
$(\breve{x},\breve{y})$:
} (normalized) image coordinates.

$$\breve{x} = x + x\left(k_1 \overbrace{(x^2+y^2)}^{r^2 = (x^2+y^2)} + k_2 (x^2+y^2)^2\right)$$

$$\breve{y} = y + y\left(k_1 (x^2+y^2) + k_2 (x^2+y^2)^2\right)$$



- k_1, k_2 : are coefficients of radial lens distortion.
- The center of radial distortion is the same as the principal point.

Go to pixel coord:

$$\breve{u} = u_0 + \alpha \breve{x} + \gamma \breve{y}$$

$$\breve{v} = v_0 + \beta \breve{y}$$

next set $\gamma = 0$ (ignores the skew)

$$\Rightarrow \beta \breve{y} = \breve{v} - v_0 \longrightarrow$$

$$\breve{V} = V_0 + By + By\,k_1 r^2 + B\,y\,k_2 r^4$$

$$\breve{V} = V_0 + \underbrace{By}_{V-V_0}\left(1 + k_1 r^2 + k_2 r^4\right)$$

$$\left\{\begin{aligned} \breve{V} &= V + (V - V_0)\left(\underline{k_1}\, r^2 + \underline{k_2}\, r^4\right) \\ \breve{U} &= U + (U - U_0)\left(\underline{k_1}\, r^2 + \underline{k_2}\, r^4\right) \end{aligned}\right.$$

↳ Re-write to get :

$$\underbrace{\begin{bmatrix} (U-U_0)r^2 & (U-U_0)r^4 \\ (V-V_0)r^2 & (V-V_0)r^4 \end{bmatrix}}_{\underline{\underline{D}}} \underbrace{\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}}_{\underline{k}} = \underbrace{\begin{bmatrix} \breve{U} - U \\ \breve{V} - V \end{bmatrix}}_{\underline{d}}$$

$$\Longrightarrow \underline{k} = \left(\underline{\underline{D}}^T \underline{\underline{D}}\right)^{-1} \underline{\underline{D}}^T d \rightarrow$$ gives us k_1, k_2 initial estimates

⟹ Complete optimization / refinement :

$$J = \sum_{i=1}^{n} \sum_{j=1}^{m} \left\| \underline{m}_{ij} - \underbrace{\breve{\underline{m}}(\underline{\underline{A}}, k_1, k_2, \mathbb{R}_i, \underline{t}_i, \underline{M}_j)}_{\text{projection of model point } \underline{M}_j \text{ onto image } i} \right\|^2$$

$\swarrow \underline{r}_i$

$$\arg\min_{\underline{\underline{A}}, k_1, k_2, \underline{r}_i, \underline{t}_i} J$$: 13 parameters

→ Nonlinear minimization problem. (LM algorithm) e.g.

≡ → **Bundle Adjustment.**

HW 3 : Use this calibratio
for a basic Augmented Reality
exercise.

Important Note : $\hat{R} \leftarrow \underline{r}$ (Rodrigues) $\rightarrow \hat{R} \overset{?}{\in} SO(3) = \{ \underline{\underline{R}}_{3\times3} : R^T R = I, \; RR^T = I, \; \det(R) = +1 \}$

Project $\underline{\underline{\hat{R}}}$ into SO(3)) : $\hat{\mathbb{R}} = \underline{\underline{U}} \underbrace{\Sigma}_{} \underline{\underline{V}}^T$ (SVD)

$\begin{bmatrix} \ddots \end{bmatrix} \rightarrow \underline{\underline{I}}$

project

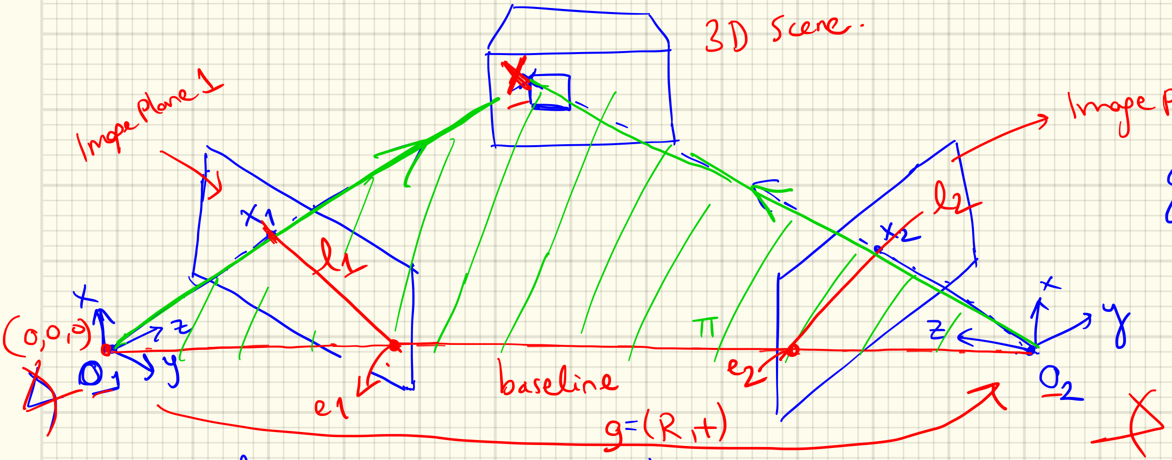
Let $\hat{R}_{\text{projected}} = \underline{\underline{U}}\,\underline{\underline{V}}^T$

≡ enforcing orthogonality & constraint of det(R) = +1

EPIPOLAR GEOMETRY & The Fundamental Matrix

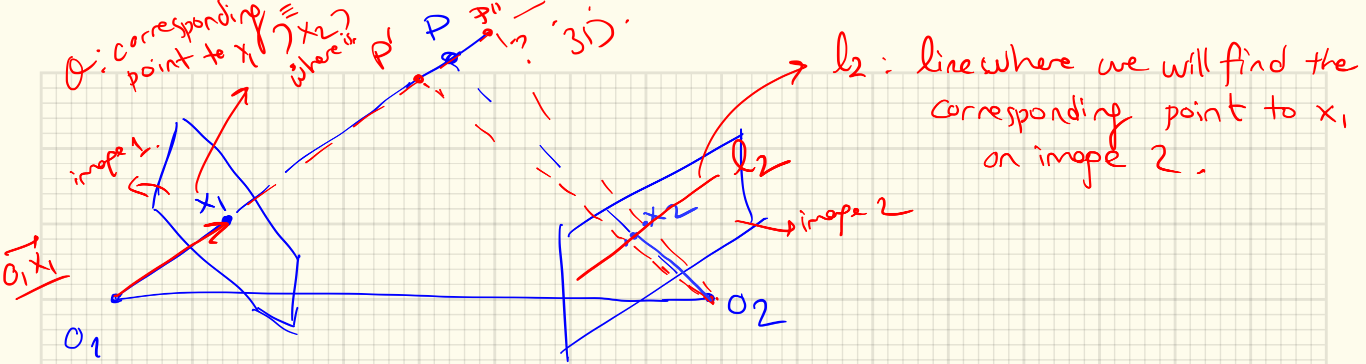
Reading [HZ Chap 9] [MaS] Chap 5.

- Instrinsic projective geometry between two views (of the same scene)
- Fundamental matrix F encapsulates that geometry.



$g=(R,t)$
rigid transform btw the two cameras
$g \in SE(3)$.

- Coordinates of projections of a point X, and the two camera optical centers, and the point X itself form a triangle / plane.
- **Baseline**: the line joining two camera centers
- **Epipole**: Point of intersection of the baseline w/ the image plane.



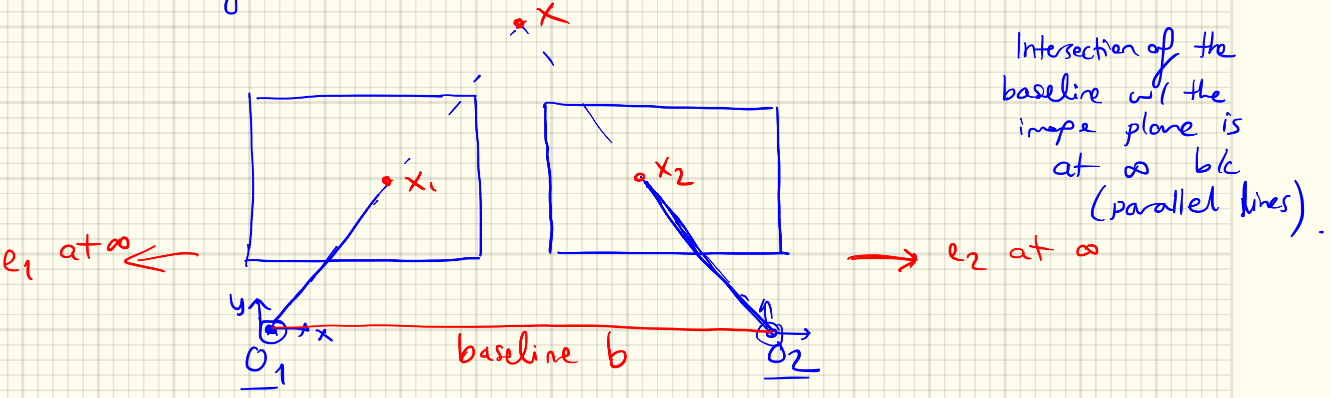
Q: corresponding point to x_1 ≡ x_2? where is x_2?

l_2: line where we will find the corresponding point to x_1 on image 2.

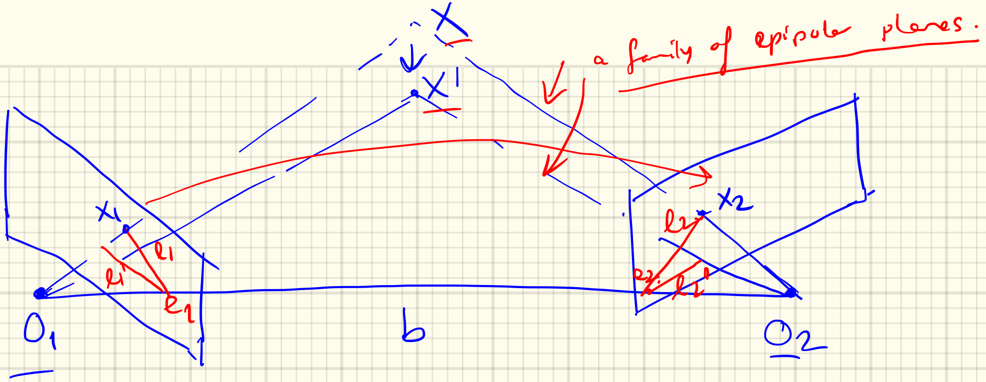
≡ → Epipole: is the image in one view of the camera center of the other view.

→ Epipoles may lie outside of the visible image (image sensor)

eg. a translating camera, translation is parallel to the image plane



Intersection of the baseline w/ the image plane is at ∞ b/c (parallel lines).



epipolar pencil of planes

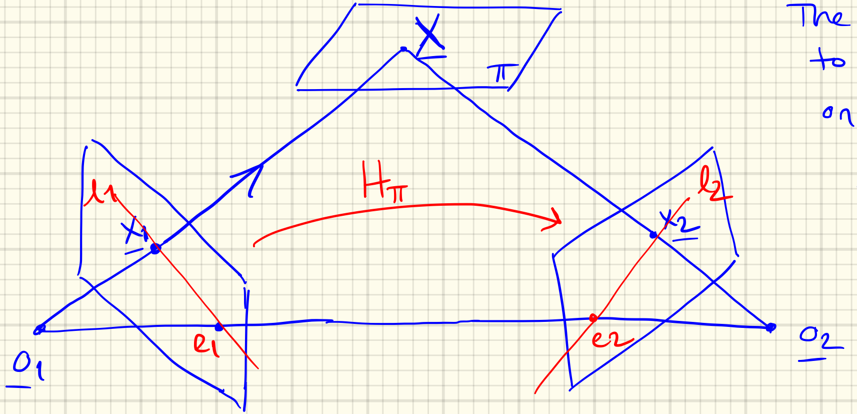
Epipolar plane. The plane $(\underline{O}_1, \underline{O}_2, \underline{X})$ containing the baseline.

(Def) Epipolar Line: intersection of an epipolar plane w/ the image plane

Geometric Derivation: Consider a plane π in 3D space:

we know that

The projected point $\underline{x_2}$, corresp. to 3D point $\underline{X}$ must lie on on the epipolar line l_2, l_2 corresponds to the image of the ray $(\underline{o_1}, \underline{x_1})$



− ∃ a 2D homography $\underline{\underline{H_\pi}}$ that maps $\underline{x_1}$ to $\underline{x_2}$. Given $\underline{x_2}$, the epipolar line $\underline{l_2}$ passing thru $\underline{x_2}$ (& the epipole $\underline{e_2}$) can be written as:

$$\underline{l_2} = \underline{e_2} \underset{\substack{\uparrow \\ \text{cross product}}}{\times} \underline{x_2} = \underline{\underline{\hat{e}_2}}\, \underline{x_2} \quad , \quad \text{since} \quad \underline{x_2} = \underline{\underline{H_\pi}}\, \underline{x_1}$$

$$\Rightarrow \underset{\uparrow}{\underline{l_2}} = \underbrace{\underline{\underline{\hat{e}_2}}\, \underline{\underline{H_\pi}}}_{\triangleq\, \underline{\underline{F}}}\, x_1 = \underline{\underline{F}}\, \underline{x_1}$$

$\underline{\underline{F}}$: Fundamental Matrix

Since $\hat{e}_2$ has rank 2 & H_π has rank 3 → $\underline{\underline{F}}$ has rank $\underline{2}$.

$\underline{\underline{F}}$: a map $\underline{x_1} \longrightarrow \underline{l_2}$:

Fundamental matrix satisfies the condition for any pair of corresponding points $\underline{x_1} \longleftrightarrow \underline{x_2}$ in the images :

$\forall \; \underline{x_1} \leftrightarrow x_2$

$$\boxed{\underline{x_2}^T \, \underline{\underline{F}} \, \underline{x_1} = 0}$$

b/c $\underline{l_2} = \underline{\underline{F}} \, \underline{x_1}$ ✓

b/c $\underline{x_2}$ lies on l_2 : $x_2^T l_2 = 0$ }

$\underline{\underline{F}}$: 3x3 matrix of $\underline{\text{rank } 2}$, homogeneous matrix → 8 d.o.f.

det (F) = 0 → 7 d.o.f.

* The importance of $\underline{x_2}^T \underline{\underline{F}} x_1 = 0$ constraint: it enables $\underline{\underline{F}}$ to be computed from image correspondences alone w/o reference to camera matrices.

Properties of $\underline{\underline{F}}$:

1) Point correspondence : If $\underline{x}_1 \longleftrightarrow \underline{x}_2$ then $\boxed{\underline{x}_2^T \underline{\underline{F}}\,\underline{x}_1 = 0}$

$$\downarrow$$

$$\boxed{\underline{x}_1^T \underbrace{\underline{\underline{F}}^T x_2}_{\ell_1} = 0}$$

2) $\underline{\ell}_2 = \underline{\underline{F}}\,\underline{x}_1$: is the epipolar line corresp. to $\underline{x}_1$

$\underline{\ell}_1 = \underline{\underline{F}}^T \underline{x}_2$ " " " " " $\underline{x}_2$.

(Take transpose of $(x_2^T F x_1 = 0)^T \rightarrow \underline{x}_1^T \underbrace{F^T x_2} = 0$

$x_1^T \underline{\ell}_1 = 0$ ✓

3) $\underline{\underline{F}}$ is a rank 2 homogeneous matrix (7 dof)

4) The epipolar line $\underline{\ell}_2 = \underline{\underline{F}}\,\underline{x}$ contains the epipole $\underline{e}_2$.

$\Rightarrow e_2^T \ell_2 = 0 = \underline{e}_2^T \underline{\underline{F}}\,\underline{x} = (\underline{e}_2^T \underline{\underline{F}})\,\underline{x} = 0 \quad \forall \underline{x}$ (on the 1st image) other than e_1.

$\Longrightarrow \boxed{\underline{e}_2^T \underline{\underline{F}} = \underline{0}}$: $\underline{e}_2$ left-null vector of $\underline{\underline{F}}$

or ($\underline{\underline{F}}^T \underline{e}_2 = \underline{0}$

Similarly : $\boxed{\underline{\underline{F}}\,e_1 = \underline{0}}$: $\underline{e}_1$ is the right-null vector of $\underline{\underline{F}}$

Recall null-spac
of a matrix A.
$y : A\underline{y} = 0$

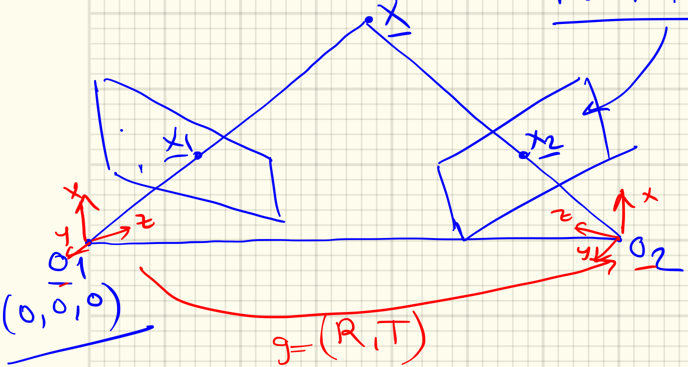
5) $\underline{\underline{F}}$ maps from $\underline{x}_1$ to $\underline{\ell}_2$ & $\underline{\underline{F}}^T$ maps from $\underline{x}_2$ to $\underline{\ell}_1$.

Epipolar Geometry w/ Calibrated Cameras : (Chap 5 [MaS])

Calibrated = $\underline{\underline{K}}$ is known, i.e. we can set $\underline{\underline{K}} = \underline{\underline{I}}$

Recall $x' = \underline{\underline{K}}\,\underline{x} \rightarrow \underline{x} = \boxed{\underline{\underline{K}}^{-1}}\, x'$ ↖ pixel coord. ✓ correspondences.

move to image coord.



$$\underline{\underline{P}} = \underbrace{\underline{\underline{K}}}_{3\times3}\,\underbrace{\Pi_0}_{3\times4}\begin{bmatrix}\underline{\underline{R}} & \underline{T}\\ \underline{0} & 1\end{bmatrix}_{4\times4}$$

Camera projection matrix

$$\lambda x' = \underline{\underline{P}}\,\underline{X}_0 = \underline{\underline{K}}\,\underbrace{[\underline{\underline{R}}, \underline{T}]}_{3\times4}\,\underline{X}_0$$

$$\lambda\,\underline{x} = \overset{K=I}{\Pi_0} \cdot \underline{X} \rightarrow \begin{pmatrix}X\\Y\\Z\\1\end{pmatrix}$$

homogeneous coord. $\begin{pmatrix}x\\y\\1\end{pmatrix}$ in camera frame

Assumption: 1) Static scene.

The image pt $\underline{x}$ differs from 3D coord. of the point by an unknown depth or scale $\lambda \in \mathbb{R}^+$.

2) Positions of corresp. feature points across the two images available

Let $(\underline{x}_1, \underline{x}_2)$ be corresp. points in two views ⟶

they are related by a precise geometric relationship :

– Let world frame is in one of the camera centers, & the other is positioned & oriented accord. to $g = (R, T)$ = a Euclidean transform

1st camera $\underline{\underline{P}}_1 = \underline{\underline{K}}\,[\,\underline{\underline{I}} \mid \underline{0}\,]$, $\underline{c}_1 = \begin{pmatrix} \underline{0} \\ 1 \end{pmatrix}$ world origin is at camera 1

2nd camera $\underline{\underline{P}}_2 = \underline{\underline{K}}\,[\,\underline{\underline{R}} \mid \underline{T}\,]$,

– Let 3D coordinates of a point p relative to the 2 cameras

$\underline{X}_1 \in \mathbb{R}^3$, $\underline{X}_2 \in \mathbb{R}^3$ $\rightarrow$ $\boxed{\underline{X}_2 = \underline{\underline{R}}\,\underline{X}_1 + \underline{T}}$ rigid body xform

Let $\underline{x}_1, \underline{x}_2$ be the homogeneous coord of the projection of the same point p into the two image planes : b/c $\boxed{\underline{X}_i = \lambda_i x_i}$, $i = 1, 2$, $\lambda_1, \lambda_2 > 0$

$\rightarrow$ $\boxed{\lambda_2 \underline{x}_2 = \lambda_1 \underline{\underline{R}}\,\underline{x}_1 + \underline{T}}$; we want to eliminate depths λ_i, pre-multiply both sides $\underline{\underline{\hat{T}}}$

to obtain $\boxed{\lambda_2 \underline{\underline{\hat{T}}}\,\underline{x}_2 = \lambda_1 \underline{\underline{\hat{T}}}\,\underline{\underline{R}}\,\underline{x}_1} + \overset{0}{\cancel{\underline{\underline{\hat{T}}}\,\underline{T}}} \equiv \underline{T} \times \underline{T}$

Vector $\underline{\underline{\hat{T}}}\,\underline{x}_2 = \underline{T} \times \underline{x}_2$ ⊥ to $\underline{x}_2$ → premultiply the last eqn premultiply by $\underline{x}_2^T$:

$$\lambda_2 \underbrace{\hat{T}\, \underline{x}_2}_{0} = \lambda_1 \underbrace{\hat{\underline{\underline{T}}}\, \underline{\underline{R}}\, \underline{x}_1}$$

$$\Rightarrow \quad x_2^T \underbrace{\hat{\underline{\underline{T}}}\, \underline{\underline{R}}}\, x_1 = 0$$

$\triangleq \underline{\underline{E}}$: called Essential Matrix

(~ Fundamental matrix for the calibrated case)

$$\boxed{\underline{x}_2^T \, \underline{\underline{E}} \, \underline{x}_1 = 0}$$

Essential Constraint → for a Calibrated Camera

(~ Fundamental Constraint)

$$\underline{\underline{E}} \triangleq \hat{\underline{\underline{T}}}\, \underline{\underline{R}}$$

i.e. intrinsics K are known

Theorem 5.1 (Epipolar Constraint) Consider 2 images $\underline{x}_1, \underline{x}_2$ of the same 3D point p from camera positions w/ relative pose $(\underline{\underline{R}}, \underline{T})$ where $\underline{\underline{R}} \in SO(3)$, relative orientation, $\underline{T} \in \mathbb{R}^3$, relative translation (position)

Then $\underline{x}_1, \underline{x}_2$ satisfy the epipolar constraint: (equation)

$$\underline{x}_2^T \, \underline{\underline{E}} \, \underline{x}_1 = 0,$$

$$\underline{\underline{E}} = \underline{\underline{\hat{T}}} \, \underline{\underline{R}} :$$

☆

Essential matrix encodes the relative $\underline{\text{pose}}$ (R, T) between the two cameras.