3D Vision

BLG634E

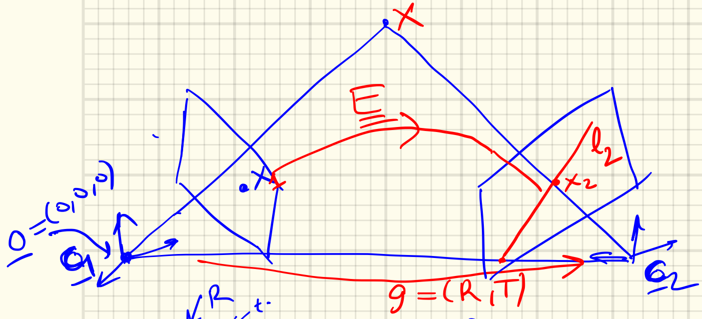
Gözde ÜNAL

11.04.2022

Recap $\boxed{\text{Calibrated}}$ $\boxed{\text{Epipolar Geometry}}$:

$= K$ is known

$$\underline{x}' = \underline{\underline{K}}\,\underline{x} \Rightarrow x = \underline{\underline{K}}^{-1} x'$$



$$\underline{X_2} = \underline{\underline{R}}\,\underline{X}_1 + \underline{T}$$ rigid body xform

$$\underline{X}_i = \lambda_i \underline{x}_i$$

$$\underline{\underline{P}}_1 = K\underbrace{[\overset{R}{I} \mid \overset{t}{0}]}_{g}, \quad \underline{c} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \underline{\underline{P_2}} = K[R \mid T]:$$

$$\lambda_2 \underline{x}_2 = \lambda_1 R\,\underline{x}_1 + \underline{T}$$

$\hat{T}$ (:

x_2^T

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

Essential Matrix

Thm (Epipolar Constraint)

Last time ✓

Properties of the Essential Matrix:

Given an essential matrix $E = \hat{T}R$ that defines an epipolar relation between two image points $\underline{x}_1, \underline{x}_2$;

(1) $\underline{\underline{E}}\, \underline{e}_1 = \underline{0}$, $\underline{\underline{E}}^T \underline{e}_2 = \underline{0}$ (Similar to fundamental matrix properties)

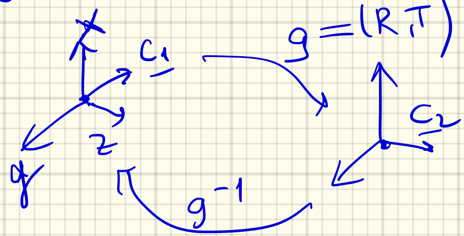
$(\underline{e}_2^T \underline{\underline{E}} = \underline{0})$

→ Also epipole $\underline{e}_2$, is the projection of camera center $\underline{C}_1$ onto image 2:

$\underline{C}_1 = \begin{pmatrix} \underline{0} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ in both world frame & camera frame

$\underline{e}_2 \overset{K=I}{=} [\underline{\underline{R}}, \underline{T}]\, \underline{C}_1 = \underline{T}$; $\boxed{\underline{e}_1 \sim \underline{T}}$ (equality up to a scale)

$\begin{bmatrix} \underline{0} \\ 1 \end{bmatrix}$



$\underline{C}_2^{world} = g\, C_1^{world} = \begin{bmatrix} \underline{T} \\ 1 \end{bmatrix}$

$g = \begin{bmatrix} R & T \\ \underline{0}^T & 1 \end{bmatrix}$

→ $\underline{e}_1$? $\underline{C}_2 = \begin{pmatrix} \underline{0} \\ 1 \end{pmatrix}$ in frame 2 coord.

$\underline{e}_1 = \underline{g}^{-1} \underline{C}_2 = [R^T \mid -R^T T] \begin{bmatrix} \underline{0} \\ 1 \end{bmatrix} = -R^T T$

$\boxed{e_2 \sim R^T T}$

② $\boxed{\underline{l_2} \sim \underline{\underline{E}}\, \underline{x_1}}$ & $\boxed{l_1 \sim \underline{\underline{E}}^T \underline{x_2}}$

l_2: epipolar lines $\in \mathbb{R}^3$; l_1: associated w/ two image points x_1 & x_2.

③ $l_i^T e_i = 0$, $i = 1, 2$ $\rightarrow$ $\underbrace{l_1^T}_{(E^T x_2)^T} e_1 = x_2^T \underbrace{E\, e_1}_{\underline{0}} = 0$

$l_i^T . x_i = 0$

★ Essential Matrix belongs to a special set of matrices in $\mathbb{R}^3$ called the Essential Space:

$$\mathcal{E} \triangleq \{ \hat{T} R \mid R \in SO(3), T \in \mathbb{R}^3 \}$$

$\Longrightarrow$

Thm $\overset{(MaS)}{(5.5)}$ Characterization of the Essential Matrix :

A non-zero matrix $\underline{\underline{E}} \in \mathbb{R}^{3\times3}$ is an essential matrix iff

$\underline{\underline{E}}$ has a singular value decomposition

b/c $\underline{\underline{E}}$ has rank 2

$$\underline{\underline{E}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T \quad w/ \quad \underline{\underline{\Sigma}} = \mathrm{diag}\{\sigma, \sigma, 0\}$$

3rd singular value is 0.

for some $\sigma \in \mathbb{R}^+$ and $\underline{\underline{U}}, \underline{\underline{V}}$ SO(3).

(proof: those interested Sec 5.1.2 [MaS] book [HZ] . . .

* Given $R \in SO(3)$, $T \in \mathbb{R}^3 \longrightarrow$ easy to construct $\boxed{\underline{\underline{E}} = \hat{T} R}$

Inverse problem: How to retrieve T & R from a given $\underline{\underline{E}}$?

But first, how to estimate essential matrix $\underline{\underline{E}}$?

The 8-point Linear Algorithm for estimating matrix:

$$\underline{\underline{E}} = \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix} \in \mathbb{R}^{3\times 3} \rightarrow \text{stack into a vector } \underline{E}^s:$$

$$\underline{E}^s = [e_1 \; e_4 \; e_7 \; e_2 \; e_5 \; e_8 \; e_3 \; e_6 \; e_9]^T \in \mathbb{R}^9$$

Let $\underline{x}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}, \underline{x}_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$

Use Kronecker product of 2 vectors $\otimes$; define

$$\underline{a} = [x_1 x_2, x_1 y_2, x_1 z_2, y_1 x_2, y_1 y_2, y_1 z_2, z_1 x_2, z_1 y_2, z_1 z_2]^T$$

Epipolar constraint $\underline{x}_2^T \underline{\underline{E}}\, \underline{x}_1 = 0 \rightarrow$ is linear in the entries of $\underline{\underline{E}}$; rewrite as

$$\rightarrow \boxed{\underline{a}^T \cdot \underline{E}^s = 0} \equiv x_2^T \underline{\underline{E}}\, \underline{x}_1 = 0$$

*Now given a set of image (corresp) points $(\underline{x}_1^j, \underline{x}_2^j), j = 1, \ldots, n,$

define a matrix $\underline{\underline{A}} \in \mathbb{R}^{n\times 9}$ associated w/ these measurements

$$\underline{\underline{A}} \triangleq \begin{bmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \vdots \\ \underline{a}_n \end{bmatrix}_{n\times 9} \longrightarrow \boxed{\underline{\underline{A}} \cdot \underline{E}^s = 0}$$

★ Linear Homogeneous equation.

→ Solve for $\underline{E}^s$

Rank of $\underline{\underline{A}} \in \mathbb{R}^{n\times 9} \rightarrow 8$

8 dof in $\underline{\underline{E}}$.

Given $n \geq 8$ corresponding points $\boxed{\text{min } 8}$

$$\underset{\underline{E}^s}{\arg\min} \; \|\underline{\underline{A}}\,\underline{E}^s\|^2 \quad \text{s.t.} \quad \|\underline{E}^s\| = 1$$

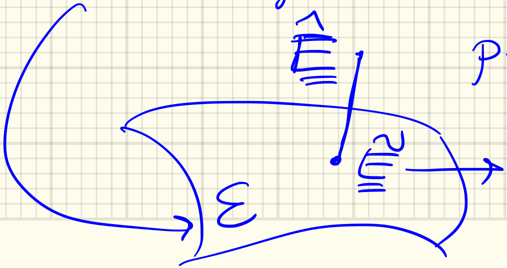
Soln: → Choose $\underline{E}^s$ to be the eigenvector of $\underline{\underline{A}}^T\underline{\underline{A}}$ that corresponds to its smallest eigenvalue.

$$\hat{\underline{E}}^s \Longrightarrow \hat{\underline{\underline{E}}} \quad \checkmark$$

We are not done yet.

why? b/c $\underline{\underline{E}}$ has to satisfy an additional constraint.

$\underline{\underline{E}}$ has to belong to $\mathcal{E}$: space of essential



project the estimated $\hat{E}$ from the 8pt algorithm onto $\mathcal{E}$.

How do you do this?

Digression:

Enforcing Constraints (in Computer Vision):

Often we generate estimates of a matrix $\underline{\underline{A}}$, e.g. a "orthogonal matrix", or the essential / fundamental matrix.

— Errors induced by noise and numerical computations alter the estimated matrix, say, $\hat{A}$, so that it no longer satisfies the given constraints

→ SVD allows us to find the "closest" matrix to $\hat{A}$ in the sense of Frobenius norm, which satisfies the properties exactly.

(Def (Frobenius norm) $\|A\|_F \triangleq \sqrt{\operatorname{trace}(A^T A)} = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$

$$= \sum_{i,j} a_{ij}^2 \Big)$$

Compute $\hat{A} = \underline{\underline{U}}\,\underline{\underline{D}}\,\underline{\underline{V}}^T$ → then choose an estimate

$\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{D'}}\,\underline{\underline{V}}^T$ w/ D' obtained by choosing the singular values of $\underline{D}$ to those expected when the constraints are satisfied.

e.g. $\underline{\underline{R}} \in \underbrace{SO(3)}$ say $\hat{\underline{\underline{R}}} = \underline{\underline{U}}\,\underline{\underline{D}}\,\underline{\underline{V}}^T$ set $\underline{\underline{D}} = \underline{\underline{I}}$

$\hat{R} = \underline{\underline{U}}\,\underline{\underline{V}}^T$

Note: If $\hat{A}$ is a good estimate, its singular values should not be too far from the expected ones.

Fixed Rank Approximation: Given a matrix $\underline{\underline{A}}$ of rank r, want to find a matrix $\underline{\underline{B}}$ of rank p, $p < r$, & the

$$\| A - B \|_F \text{ is minimal.}$$

The soln: $\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{D}}\,\underline{\underline{V}}^T \longrightarrow \underline{\underline{B}} = \underline{\underline{U}}\,\underline{\underline{D}}_{(p)}\,\underline{\underline{V}}^T$

$$\underline{\underline{D}} = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix} \rightarrow \underline{\underline{D}}_p = \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_p & & \\ & & & 0 & \\ & & & & 0 \ddots \end{bmatrix}$$

setting other singular values to 0

Thm 5.9 (MaS) [Projection onto the Essential Space]:
Given a real matrix $\underline{\underline{G}} \in \mathbb{R}^{3\times 3}$ w/ SVD $\underline{\underline{G}} = \underline{\underline{U}} \begin{bmatrix} d_1 & & 0 \\ & d_2 & \\ 0 & & d_3 \end{bmatrix} \underline{\underline{V}}^T$ w/ $\underline{\underline{U}}, \underline{\underline{V}} \in SO(3)$, $d_1 \geq d_2 \geq d_3$: singular values

Then the essential matrix $\underline{\underline{E}} \in \mathcal{E}$ that minimizes the error $\|\underline{\underline{G}} - \underline{\underline{E}}\|_F^2$ is given by $\underline{\underline{E}} = \underline{\underline{U}} \begin{bmatrix} \sigma & & \\ & \sigma & \\ & & 0 \end{bmatrix} \underline{\underline{V}}^T$ w/ $\sigma = \left(\frac{d_1 + d_2}{2}\right)$.

pf. (MaS) book.

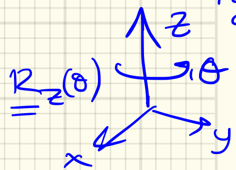
→ Now we have $\tilde{E}$: an estimated essential matrix in $\mathcal{E}$.

SVD on $\tilde{E} = U \Sigma V^T$, $U, V \in SO(3)$,

Thm 5.7 (Recovery of the camera pose from the Essential Matrix)

$E = \hat{T} R$. $\exists$ exactly 2 relative poses (R, T) w/ $R \in SO(3)$, $T \in \mathbb{R}^3$ corresponding to a non-zero essential matrix $E \in \mathcal{E}$.

Define: $R_z(\theta) \triangleq e^{\hat{e}_3 \theta}$ w/ $e_3 = [0, 0, 1]^T \in \mathbb{R}^3$



rotation around axis z.

$$R_z\left(+\frac{\pi}{2}\right) = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$(\hat{T}_1, R_1) = \Big(\underbrace{U R_z\left(+\tfrac{\pi}{2}\right) \Sigma U^T}_{\hat{T}_1},\ \underbrace{U R_z^T\left(+\tfrac{\pi}{2}\right) V^T}_{R_1}\Big)$$

$$(\hat{T}_2, R_2) = \Big(\underbrace{U R_z\left(-\tfrac{\pi}{2}\right) \Sigma U^T}_{\hat{T}_2},\ \underbrace{U R_z^T\left(-\tfrac{\pi}{2}\right) V^T}_{R_2}\Big)$$

One can verify that $\hat{T}_1 R_1 = \hat{T}_2 R_2 = E$ ✓

$\hat{T}_1 R_1$

eg. $\underline{U}\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\underline{V}^T = \ldots$

LHS = RHS

$$E = \underline{U}\begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{bmatrix}\underline{V}^T \quad ✓$$

$\underline{E}$ is an essential matrix.

★ $-E$ is also a solution. 4 possibilities in $\hat{T}, R$ decomposition.

4 possible solutions (T, R) pick a pair $\underline{x}_1, \underline{x}_2$ corresp. image points

$$d_2 x_2 = d_1 R x_1 + T \rightarrow d_1, d_2$$

3 solns: will yield either $-d_1$ or $-$ve d_2 or both. Hence, only 1 of the (R, T) solution satisfies the positive depth constraint. You pick that one.

Summary

$$\{x_1^j, x_2^j\}_{j=1}^n \xrightarrow{\text{estimate}} \underline{\hat{E}} \xrightarrow{\text{project}} \underline{E} \in \mathcal{E} \longrightarrow$$

you resolve T, R from $\underline{E}$

Started w/ image measurement

$\equiv$ Camera pose

extrinsic calibration

Note: W/o loss of generality, $\underline{T}$ can be rescaled to be unit length.

$$x_2^T \hat{T} R x_1 = 0 \rightarrow \cancel{x}_2^T (\alpha \hat{T}) R x_1 = 0$$

$$\hat{T} \rightarrow \alpha \hat{T} \equiv T \longrightarrow \alpha T \quad \checkmark \qquad \underline{T} = \begin{bmatrix} T^x \\ T^y \\ T^z \end{bmatrix}$$

$\therefore$ Space of essential matrices is 5-dimensional.

$\underline{\underline{E}} = \underline{\underline{\hat{T}}}\,\underline{\underline{R}}$, $\quad R$ has 3 dof
$\qquad\qquad\qquad\quad T$ has 2 dof

$\therefore \underline{\underline{E}}$ is defined up to a scale
$\quad \underline{T} \qquad$ " $\qquad\qquad$ "

a typical choice to fix this ambiguity is to assume $\|\underline{T}\| = 1 \rightarrow \|E\| = 1$.
(unit translation)

Uncalibrated Epipolar Geometry (Chap 6 MaS. → only 6.1, 6.2, 6.A)

$(K \neq I)$

Similar derivation to calibrated case: direct elimination of the unknown depths λ_1 & λ_2 from the rigid body eqn:

$$\lambda_2 \underline{x_2} = R \lambda_1 \underline{x_1} + \underline{T} \quad , \quad \lambda \underline{x} = \underline{X}$$

multiply both sides by $\underline{\underline{K}}$

$$\lambda_2 \underbrace{\underline{\underline{K}} x_2}_{x_2'} = \underline{\underline{K}}\,\underline{\underline{R}}\,\lambda_1 \underbrace{x_1}_{K^{-1}Kx_1} + \underbrace{\underline{\underline{K}}\,\underline{T}}_{} \triangleq T' = \underline{\underline{K}}\,\underline{T}$$

pixel coordinates → x_2'

$$\lambda_2 x_2' = \underline{\underline{K}}\,\underline{\underline{R}}\,\underline{\underline{K}}^{-1} \lambda_1 x_1' + \underline{T}'$$

$\hat{T}'$

$$\lambda_2 \underbrace{\hat{T}' x_2'}_{\perp\, x_2'} = \underbrace{\hat{T}' KRK^{-1}}_{\underline{\underline{F}}} \lambda_1 x_1' + 0$$

$x_2'^T$

$$0 \Rightarrow x_2'^T \boxed{\hat{T}' KRK^{-1}} x_1' = 0$$

$\triangleq \underline{\underline{F}}$: Fundamental matrix

We are in pixel coord.

$$\boxed{x_2'^T \underline{\underline{F}} x_1' = 0}$$

→ or direct substitution of $\boxed{\underline{x} = \underline{\underline{K}}^{-1}\underline{x}'}$ into the calibrated epipolar constraint:

calibrated EG $\quad \underline{x}_2^T \underline{\underline{\hat{T}}}\,\underline{\underline{R}}\,\underline{x}_1 = 0 \quad \Rightarrow \quad x_2'^T \underbrace{K^{-T}\,\hat{T}\,R\,K^{-1}}\,x_1' = 0$

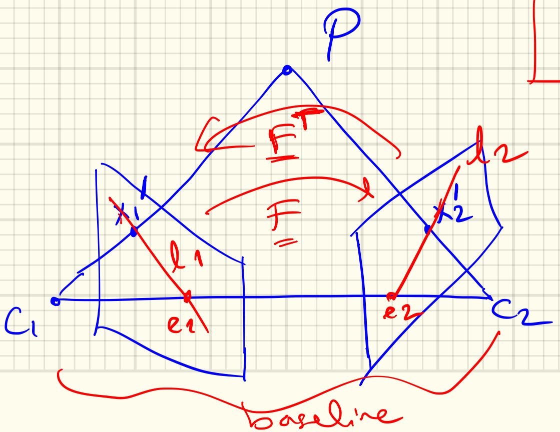
$$\underline{\underline{F}} \triangleq K^{-T}\,\hat{T}\,R\,K^{-1}$$

before: $F \triangleq \hat{T}'KRK^{-1}$, $\widehat{(KT)}$

Check when $\underline{\underline{K}} = I \Rightarrow \underline{\underline{F}} = \underline{\underline{E}}$

multiply from both sides.

$$\Rightarrow \quad \underline{\underline{E}} = K^T F K$$



$$\boxed{l_2 = \underline{\underline{F}}\,x_1'}$$

$$\boxed{l_1 = F^T x_2'}$$

$\underline{\underline{F}}\,\underline{e}_1 = \underline{0}$, $\underline{e}_2^T\,\underline{\underline{F}} = 0$

→ $\underline{\underline{F}}$: product of a skew sym matrix $\hat{T}'$ of rank 2, and a matrix $\underline{\underline{K}}\,\underline{\underline{R}}\,\underline{\underline{K}}^{-1} \in \mathbb{R}^3$ of rank 3 → $\underline{\underline{F}}$ has rank 2.

(same as $\underline{\underline{E}}$)

$\underline{\underline{F}}$ can be characterized by SVD : $\underline{\underline{F}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T$

w/ $\underline{\underline{\Sigma}} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\sigma_1, \sigma_2 \in \mathbb{R}^+$

$\sigma_1 \geq \sigma_2$

(In contrast to $\underline{\underline{E}}$ where $\sigma_1 = \sigma_2 = \sigma$)

∴ Any rank 2 3×3 matrix can be a fundamental matrix.

Again we can use $\boxed{\text{8-pt Algorithm}}$ to estimate $\underline{\underline{F}}$

★ w/ $\underline{\underline{E}}$, we were able to decompose it into $\hat{T}$, R ⟶ recovered camera pose.

In the Uncalibrated case, we cannot simply do that :

Why?

⟶

$\underline{\underline{F}}$ has at most 8 free parameters, but it is composed of
$\underline{\underline{K}}$ (5 dof), $\underline{\underline{R}}$ (3 dof), $\underline{T}$ (2 dof)

$\therefore$ From 8 dof in $\underline{\underline{F}}$, we cannot recover 10 dof in $\underline{\underline{K}}, \underline{\underline{R}}, \underline{T}$.

$\rightarrow$ you can only recover upto a projective ambiguity

eg. Stratified Reconstruction: first recover a projective xform, then an affine, until Euclidean w/ some scene constraints/assumptions.

(See Chap 6.4 those interested) (out of our scope)

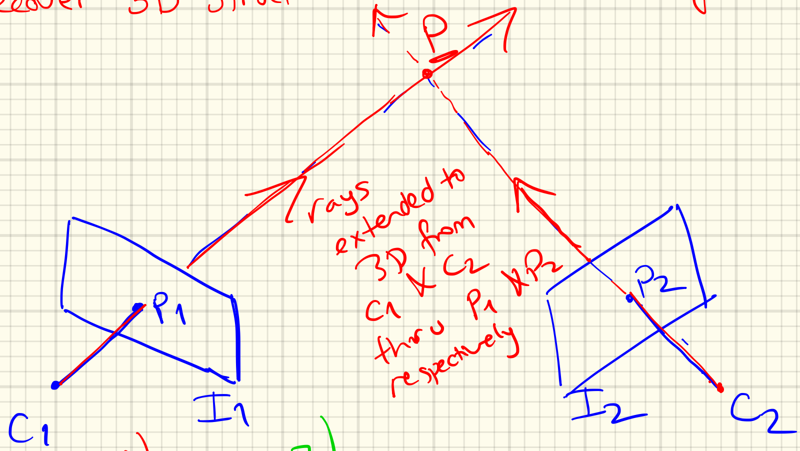
Wise $\bigstar$ $\rightarrow$ Recommendation: Work w/ Calibrated cameras.

$\rightarrow$ hence $\underline{\underline{K}}$ is known.

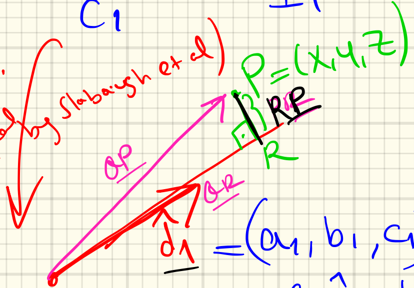
Use 8_pt algorithm to estimate E (or F) then recover T & R.

Idea of $\boxed{\text{Triangulation}}$ to Recover 3D Structure → 3D Coordinates of the Structure points

We assume we have fully-calibrated cameras (K, R, T)



Notation: (in the technical report by Slabaugh et al)



$\underline{d_1} = (a_1, b_1, c_1)$

$a_1 \hat{x} + b_1 \hat{y} + c_1 \hat{z}$

$O = (x_1, y_1, z_1)$

$$\underline{OP} = \begin{pmatrix} x - x_1 \\ y - y_1 \\ z - z_1 \end{pmatrix} \rightarrow \|OP\|$$

$\underline{OR}$: projection of $\underline{OP}$ onto $\hat{\underline{d_1}}$ ray

$$\underline{OR} = (\underline{OP} \cdot \underline{d_1})\, d_1$$

← unit vector.

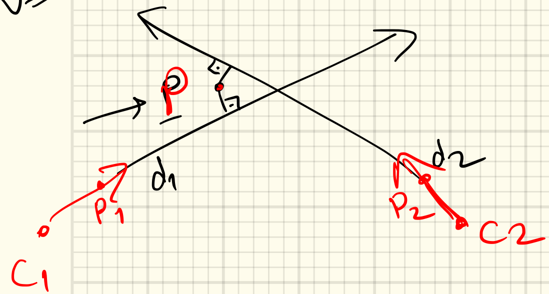
$$\|RP\|^2 = \|OP\|^2 - \|OR\|^2$$

$$\Rightarrow \|RP\|^2 = (x-x_1)^2 + (y-y_1)^2 + (z-z_1)^2 - \underbrace{\left[a_1(x-x_1) + b_1(y-y_1) + c_1(z-z_1)\right]^2}$$

Cost fn:

$$E = \sum_{i=1}^{N \rightarrow \text{images}} (x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2 - \left[a_i(x-x_i) + b_i(y-y_i) + c_i(z-z_i)\right]^2$$

N=2



$$\underset{\underbrace{x,y,z}_{P}}{\arg\min}\ E \ : \longrightarrow$$ closed form solution exists.

$\Longrightarrow$ P. ✓ 3D Point.

Final Notes: We finished Epipolar Geometry (Calibrated / Uncalibrated)

→ Use normalized 8-point algorithm. ✓

→ Use RANSAC w/ 8pt algo ✓

— Fundamental Matrix Song 🙂

→ $\boxed{\text{Triangulation}}$ → A direct method

→ "Optimal" Ray Intersection for Computing 3D Points

Camera parameters

→ $\boxed{\text{Bundle Adjustment}}$: estimating Camera matrices $\underline{M}$ & 3D Points $\underline{P}$.

nonlinear least squares optimization
↳ iteratively solved.