Robot Perception

Multi-View Geometry

Dr. Chen Feng

cfeng@nyu.edu

ROB-GY 6203, Fall 2022

Overview

* Hands-on: AprilTag & camera calibration

+ Epipolar geometry

++ Fundamental matrix

+ Essential matrix

+ Planar Homography

+ PnP problem

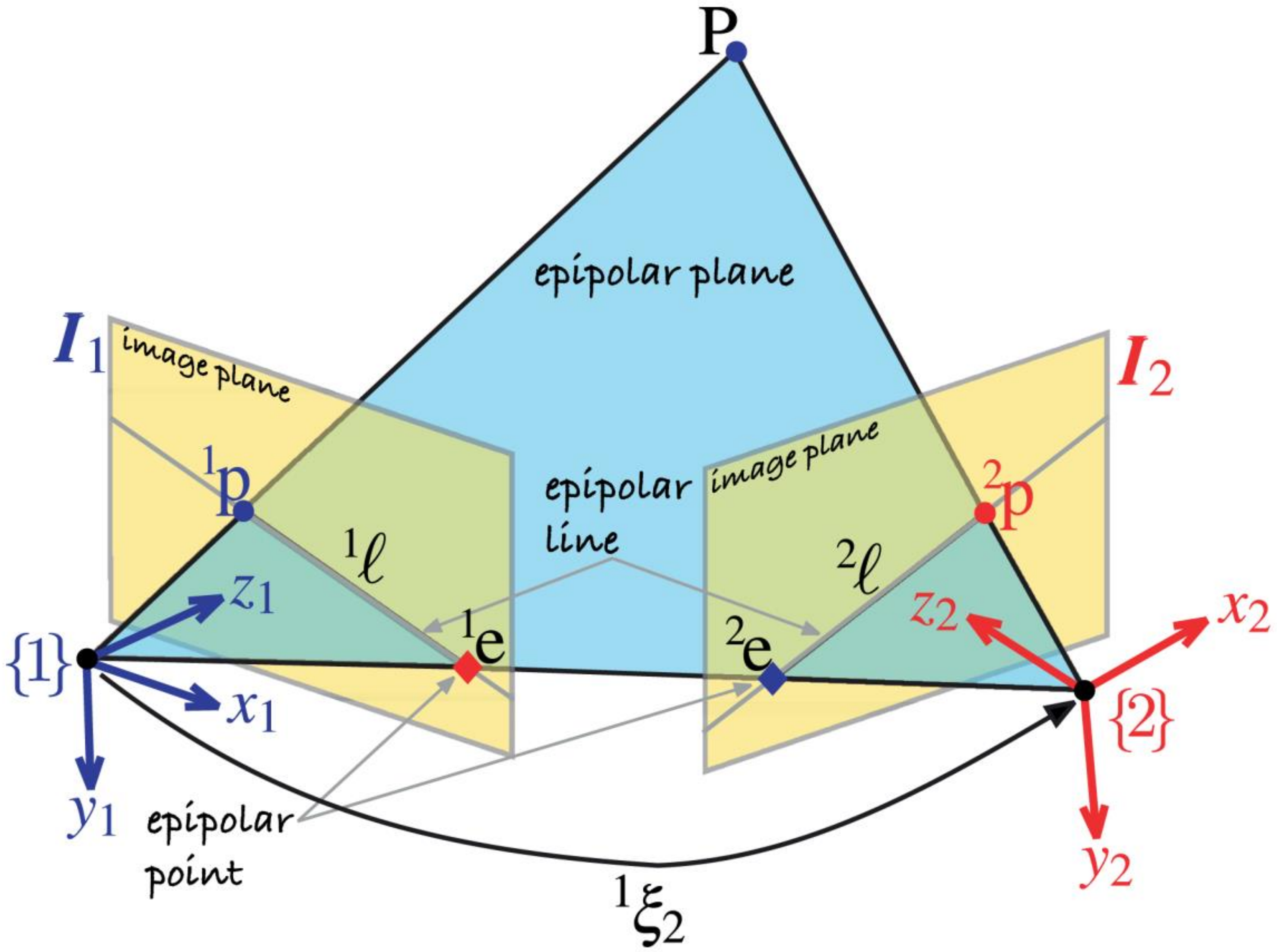
++ Hand-eye calibration

*: know how to code
++: know how to derive
+: know the concept

References

- HZ2003:
  - Section 9.1, 9.2, 9.3, 9.5, 9.6, 11.1, 11.2, 11.7
- Co2017:
  - Section 14.2, 11.2.3
- Sz2022:
  - Section 11.3, 11.2, 12.1
- FP2011:
  - Section 7.1, 8.1.2
- Radu Horaud, Fadi Dornaika. Hand-eye Calibration. International Journal of Robotics Research, SAGE Publications, 1995, 14 (3), pp.195–210.

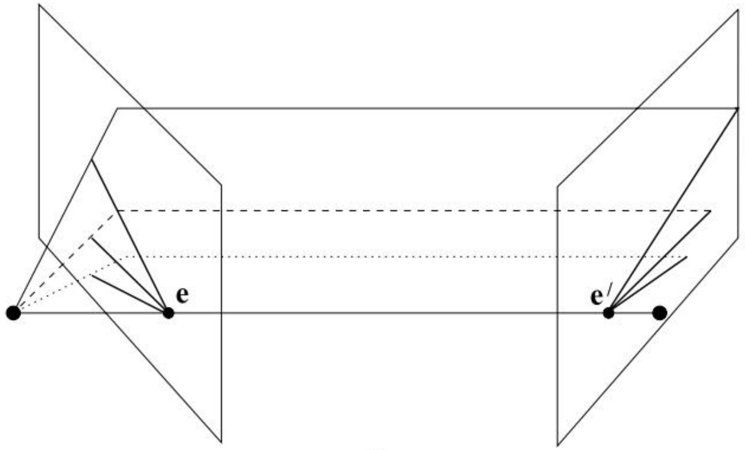
Two-view Geometry



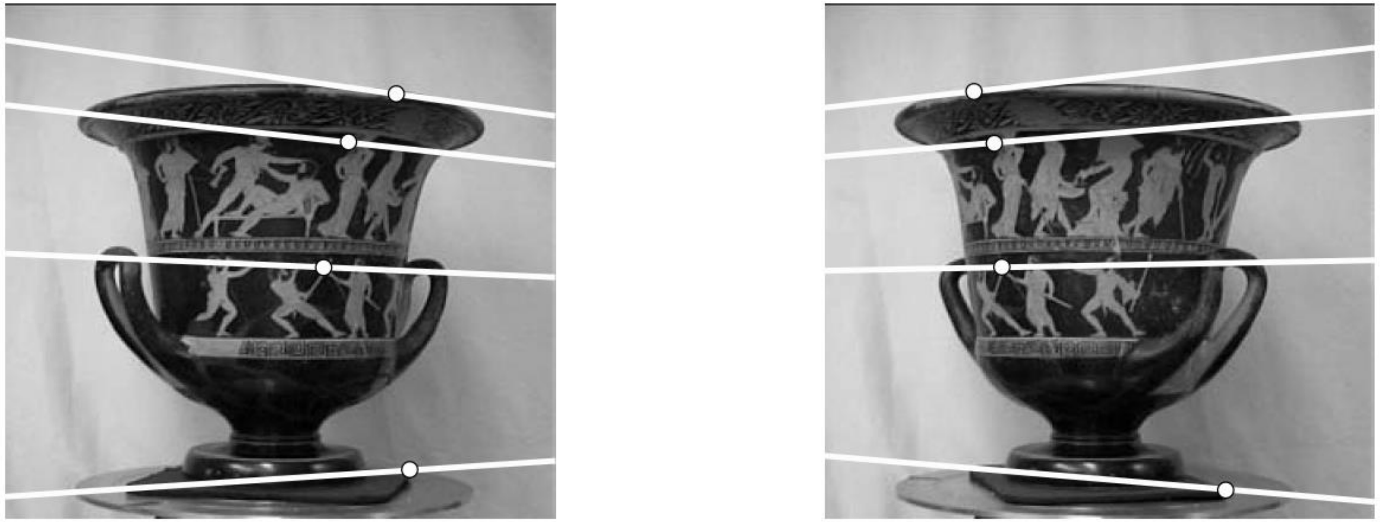
Corke 2011

Epipolar Geometry

- **Independent of scene structure**
- Only depends on the **cameras' internal parameters and relative pose**



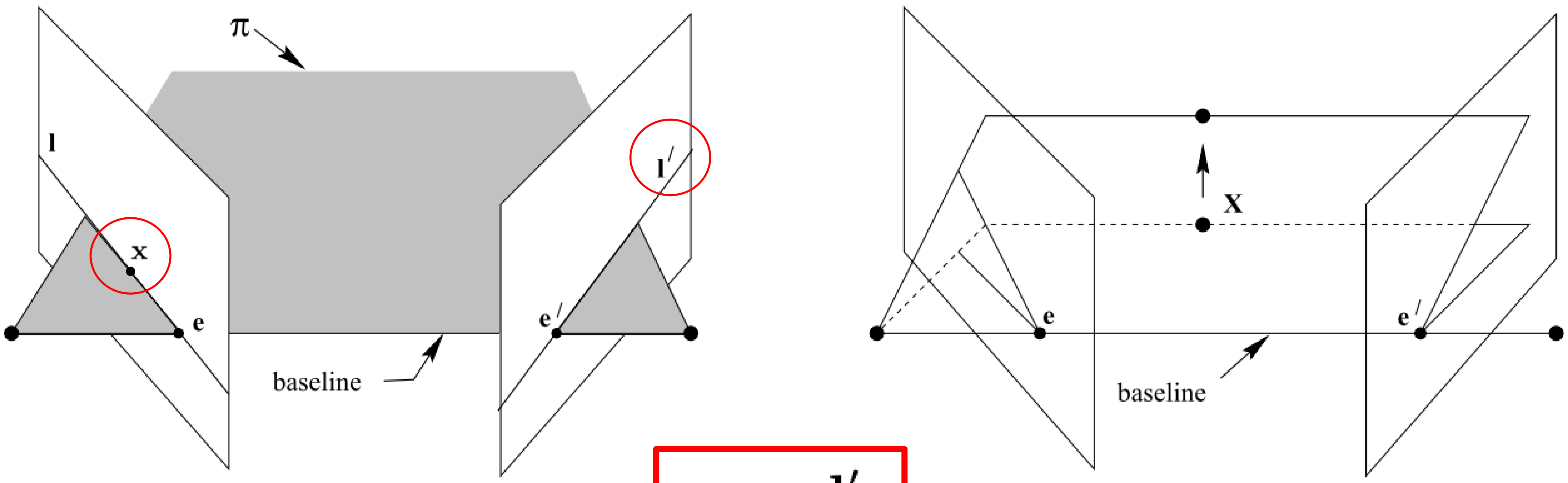
a



Hartley & Zisserman 2003

Epipolar Geometry

- From **a point x** on one image, epipolar geometry allows us to find on the other image **a corresponding line l'** that **MUST contain** the **corresponding point x', without knowing where x' is**



$$\boxed{\mathbf{x} \mapsto \mathbf{l}'}$$

Hartley & Zisserman 2003

Fundamental Matrix: the Algebra of Epipolar Geometry

Result 9.3. *The fundamental matrix satisfies the condition that for any pair of corresponding points* $\mathbf{x} \leftrightarrow \mathbf{x}'$ *in the two images*

$$\mathbf{x}'^{\mathsf{T}} \mathrm{F} \mathbf{x} = 0.$$

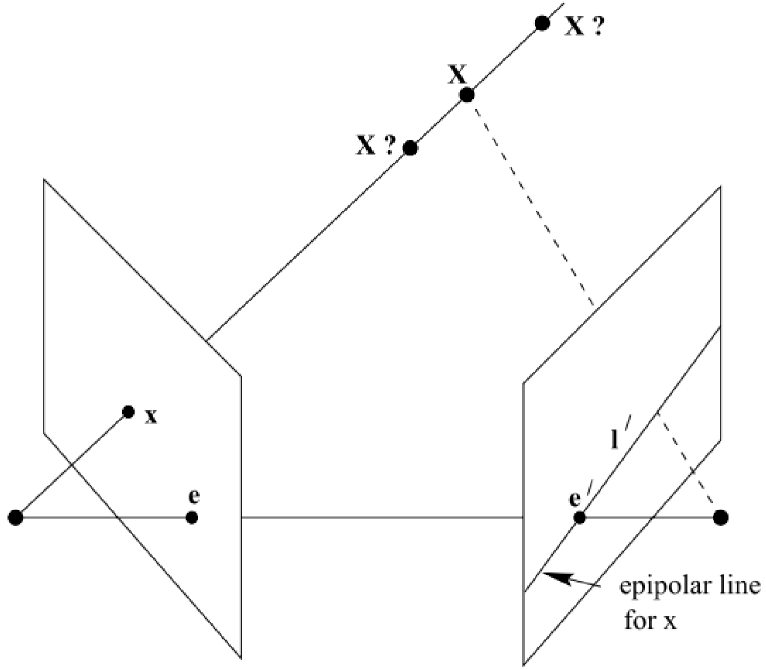
[LonguetHiggins-81] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, September 1981.

Hartley & Zisserman 2003

Fundamental Matrix: the Algebra of Epipolar Geometry

Result 9.3. *The fundamental matrix satisfies the condition that for any pair of corresponding points* $\mathbf{x} \leftrightarrow \mathbf{x}'$ *in the two images*

$$\mathbf{x}'^{\mathsf{T}} \mathrm{F} \mathbf{x} = 0.$$



$$\boxed{\mathbf{x} \mapsto \mathbf{l}'}$$

Hartley & Zisserman 2003

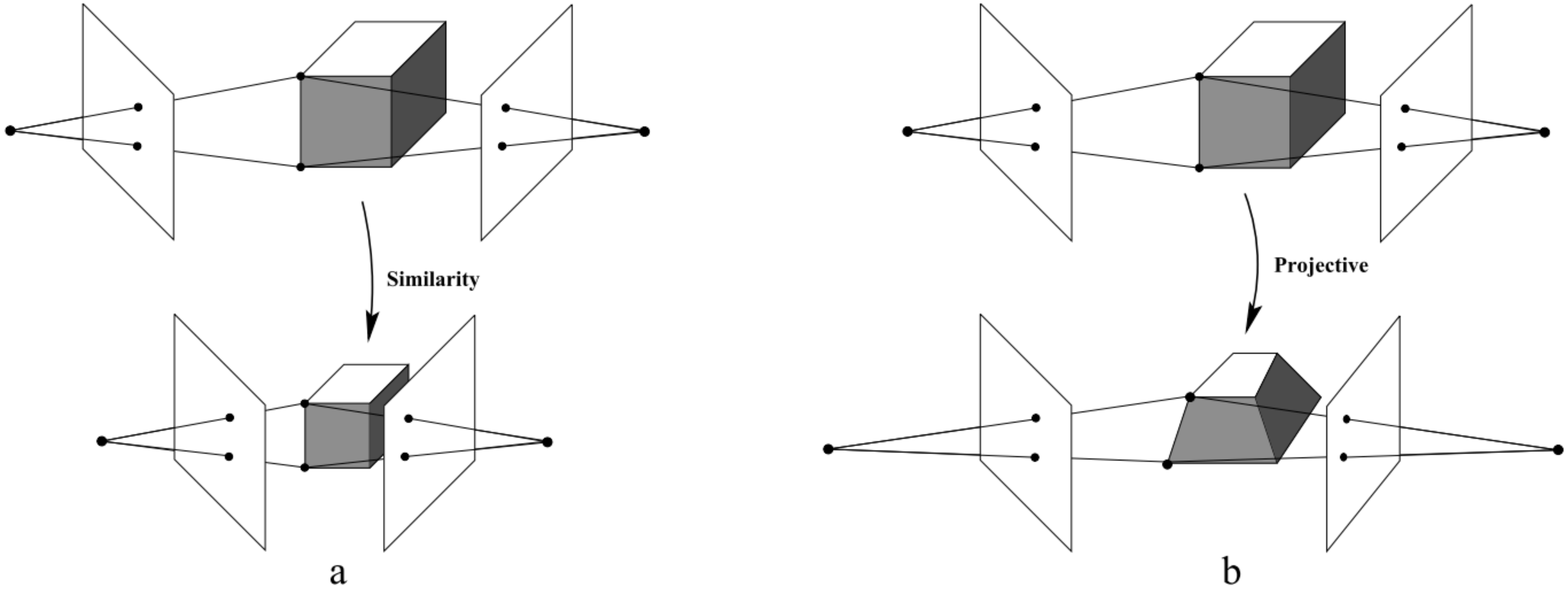
Properties of Fundamental Matrix

- $\boxed{\text{F is a rank 2 homogeneous matrix with 7 degrees of freedom.}}$
- **Point correspondence**: If $\mathbf{x}$ and $\mathbf{x}'$ are corresponding image points, then $\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$.
- **Epipolar lines**:
  - $\boxed{\mathbf{l}' = \mathtt{F}\mathbf{x} \text{ is the epipolar line corresponding to } \mathbf{x}.}$
  - $\mathbf{l} = \mathtt{F}^{\mathsf{T}}\mathbf{x}'$ is the epipolar line corresponding to $\mathbf{x}'$.
- **Epipoles**:
  - $\boxed{\mathtt{F}\mathbf{e} = \mathbf{0}.}$
  - $\mathtt{F}^{\mathsf{T}}\mathbf{e}' = \mathbf{0}$.
- **Computation from camera matrices** $\mathtt{P}, \mathtt{P}'$:
  - General cameras,
    $\mathtt{F} = [\mathbf{e}']_\times \mathtt{P}'\mathtt{P}^+$, where $\mathtt{P}^+$ is the pseudo-inverse of $\mathtt{P}$, and $\mathbf{e}' = \mathtt{P}'\mathbf{C}$, with $\mathtt{P}\mathbf{C} = \mathbf{0}$.
  - Canonical cameras, $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, $\mathtt{P}' = [\mathtt{M} \mid \mathbf{m}]$,
    $\mathtt{F} = [\mathbf{e}']_\times \mathtt{M} = \mathtt{M}^{-\mathsf{T}}[\mathbf{e}]_\times$, where $\mathbf{e}' = \mathbf{m}$ and $\mathbf{e} = \mathtt{M}^{-1}\mathbf{m}$.
  - Cameras not at infinity $\mathtt{P} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]$, $\mathtt{P}' = \mathtt{K}'[\mathtt{R} \mid \mathbf{t}]$,
    $\mathtt{F} = \mathtt{K}'^{-\mathsf{T}}[\mathbf{t}]_\times \mathtt{R}\mathtt{K}^{-1} = [\mathtt{K}'\mathbf{t}]_\times \mathtt{K}'\mathtt{R}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}\mathtt{R}\mathtt{K}^{\mathsf{T}}[\mathtt{K}\mathtt{R}^{\mathsf{T}}\mathbf{t}]_\times$.

Geometrically, F represents a mapping from the 2-dimensional projective plane $\mathbb{P}^2$ of the first image to the pencil of epipolar lines through the epipole $\mathbf{e}'$. Thus, it represents a mapping from a 2-dimensional onto a 1-dimensional projective space, and hence must have rank 2.

Hartley & Zisserman 2003

Projective Ambiguity of Fundamental Matrix



Result 9.8. *If* $\mathtt{H}$ *is a* 4×4 *matrix representing a projective transformation of 3-space, then the fundamental matrices corresponding to the pairs of camera matrices* $(\mathtt{P}, \mathtt{P}')$ *and* $(\mathtt{PH}, \mathtt{P}'\mathtt{H})$ *are the same.*

Hartley & Zisserman 2003

Normalized Coordinate and Essential Matrix

$$\mathbf{x} = \mathtt{P}\mathbf{X}$$

$$\mathtt{P} = \mathtt{K}[\mathtt{R} \mid \mathbf{t}]$$

Normalized coordinate $\hat{\mathbf{x}} = \mathtt{K}^{-1}\mathbf{x} = [\mathtt{R} \mid \mathbf{t}]\mathbf{X}$

$$\hat{\mathbf{x}}'^{\mathsf{T}}\mathtt{E}\hat{\mathbf{x}} = 0$$

Result 9.17. *A 3×3 matrix is an essential matrix if and only if two of its singular values are equal, and the third is zero.*

$$\mathtt{E} = \mathtt{K}'^{\mathsf{T}}\mathtt{F}\mathtt{K}$$

$$\boxed{\mathtt{E} = [\mathbf{t}]_\times \mathtt{R}}$$

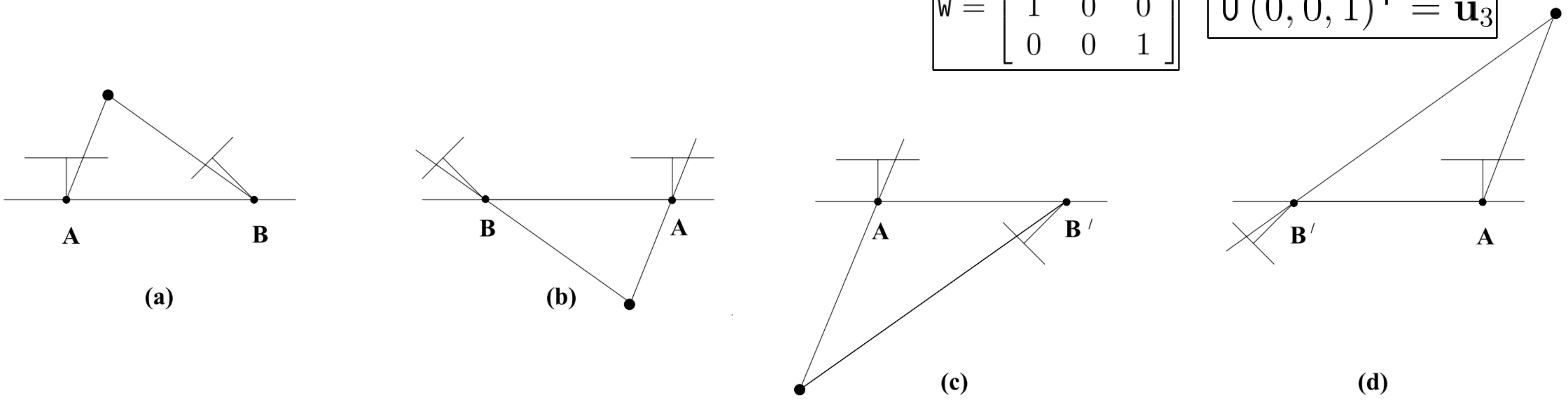
Hartley & Zisserman 2003

Extracting Relative Camera Pose from E-matrix

Result 9.19. *For a given essential matrix* $\mathtt{E} = \mathtt{U}\,\mathrm{diag}(1,1,0)\mathtt{V}^\mathsf{T}$, *and first camera matrix* $\mathtt{P} = [\mathtt{I} \mid \mathbf{0}]$, *there are four possible choices for the second camera matrix* $\mathtt{P}'$, *namely*

$$\mathtt{P}' = [\mathtt{UWV}^\mathsf{T} \mid +\mathbf{u}_3] \text{ or } [\mathtt{UWV}^\mathsf{T} \mid -\mathbf{u}_3] \text{ or } [\mathtt{UW}^\mathsf{T}\mathtt{V}^\mathsf{T} \mid +\mathbf{u}_3] \text{ or } [\mathtt{UW}^\mathsf{T}\mathtt{V}^\mathsf{T} \mid -\mathbf{u}_3].$$

$$\boxed{\mathtt{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \qquad \boxed{\mathtt{U}\,(0,0,1)^\mathsf{T} = \mathbf{u}_3}$$



(a) (b) (c) (d)

Hartley & Zisserman 2003

Estimating F-matrix

- Find multiple $\mathbf{X} \leftrightarrow \mathbf{X}'$ correspondences **(≥ 7)** between two images

$$\mathbf{x}'^{\mathsf{T}}\mathtt{F}\mathbf{x} = 0$$

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

$$\mathtt{A}\mathbf{f} = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix}\mathbf{f} = \mathbf{0}$$

- **Enforce rank-2 constraint by SVD**

Hartley & Zisserman 2003

Normalized 8-point Algorithm

Objective

Given $n \geq 8$ image point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$, determine the fundamental matrix F such that $\mathbf{x}'^{\top}_i \mathtt{F} \mathbf{x}_i = 0$.

Algorithm

(i) **Normalization:** Transform the image coordinates according to $\hat{\mathbf{x}}_i = \mathtt{T}\mathbf{x}_i$ and $\hat{\mathbf{x}}'_i = \mathtt{T}'\mathbf{x}'_i$, where T and T′ are normalizing transformations consisting of a translation and scaling.

(ii) Find the fundamental matrix $\hat{\mathtt{F}}'$ corresponding to the matches $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ by

(a) **Linear solution:** Determine $\hat{\mathtt{F}}$ from the singular vector corresponding to the smallest singular value of $\hat{\mathtt{A}}$, where $\hat{\mathtt{A}}$ is composed from the matches $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ as defined in (11.3).

(b) **Constraint enforcement:** Replace $\hat{\mathtt{F}}$ by $\hat{\mathtt{F}}'$ such that $\det \hat{\mathtt{F}}' = 0$ using the SVD (see section 11.1.1).

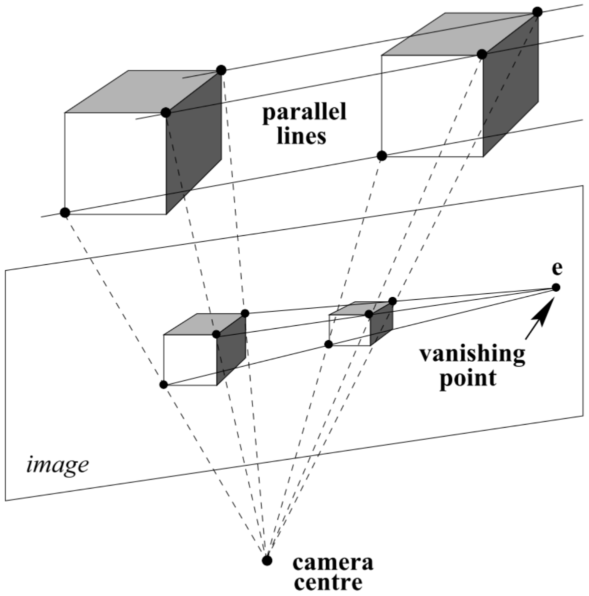
(iii) **Denormalization:** Set $\mathtt{F} = \mathtt{T}'^{\top}\hat{\mathtt{F}}'\mathtt{T}$. Matrix F is the fundamental matrix corresponding to the original data $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$.

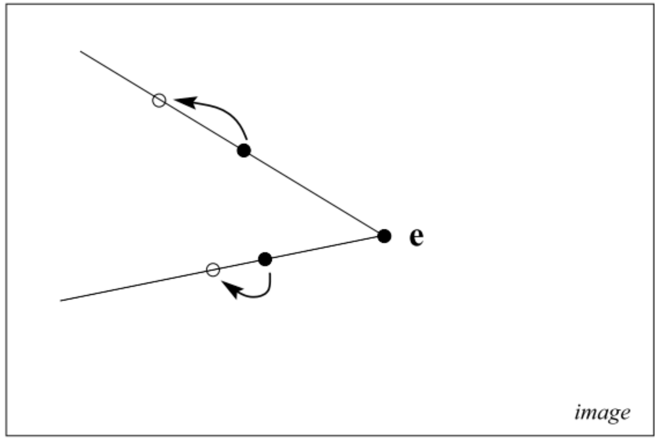
Hartley & Zisserman 2003

Special Case of F-matrix Computation

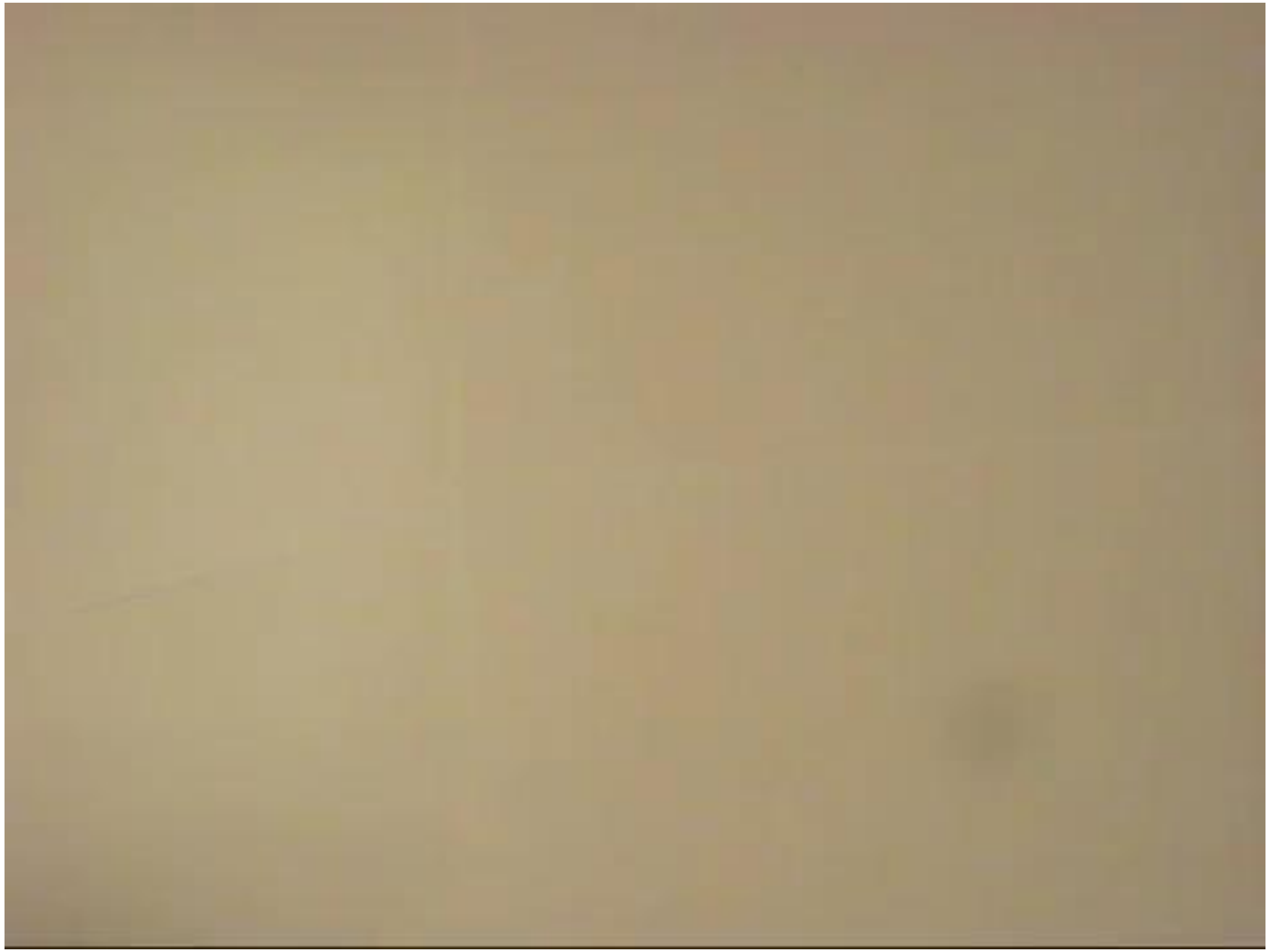
- Special cases of motion
  - Simplify the computation of F-matrix

- Pure Translation: $\boxed{\mathrm{F} = [\mathrm{e}']_\times}$
  - the epipole can be estimated from the **image motion of two points**



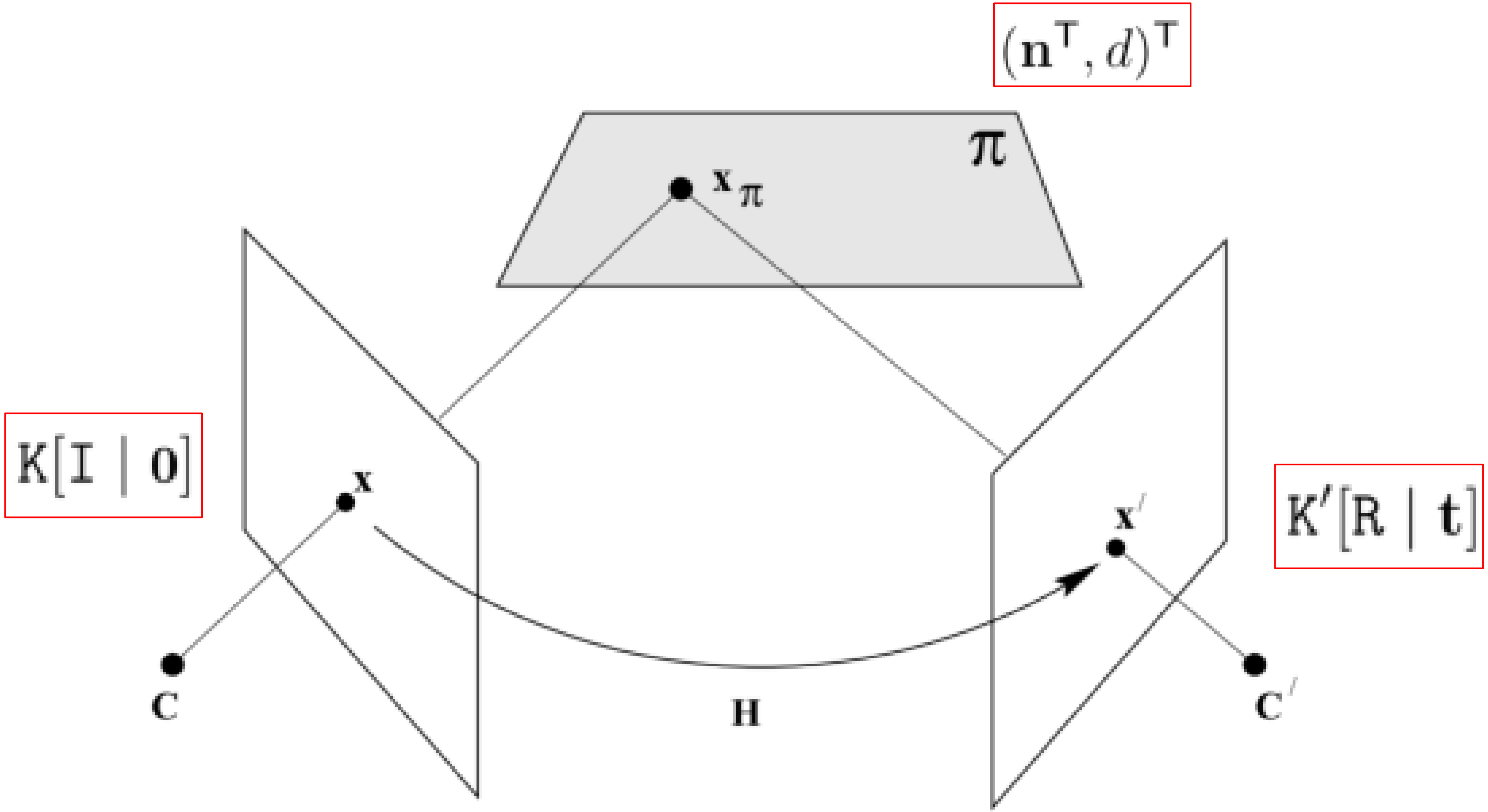


The Fundamental Matrix Song – Daniel Wedge



Video from: https://youtu.be/DgGV3l82NTk

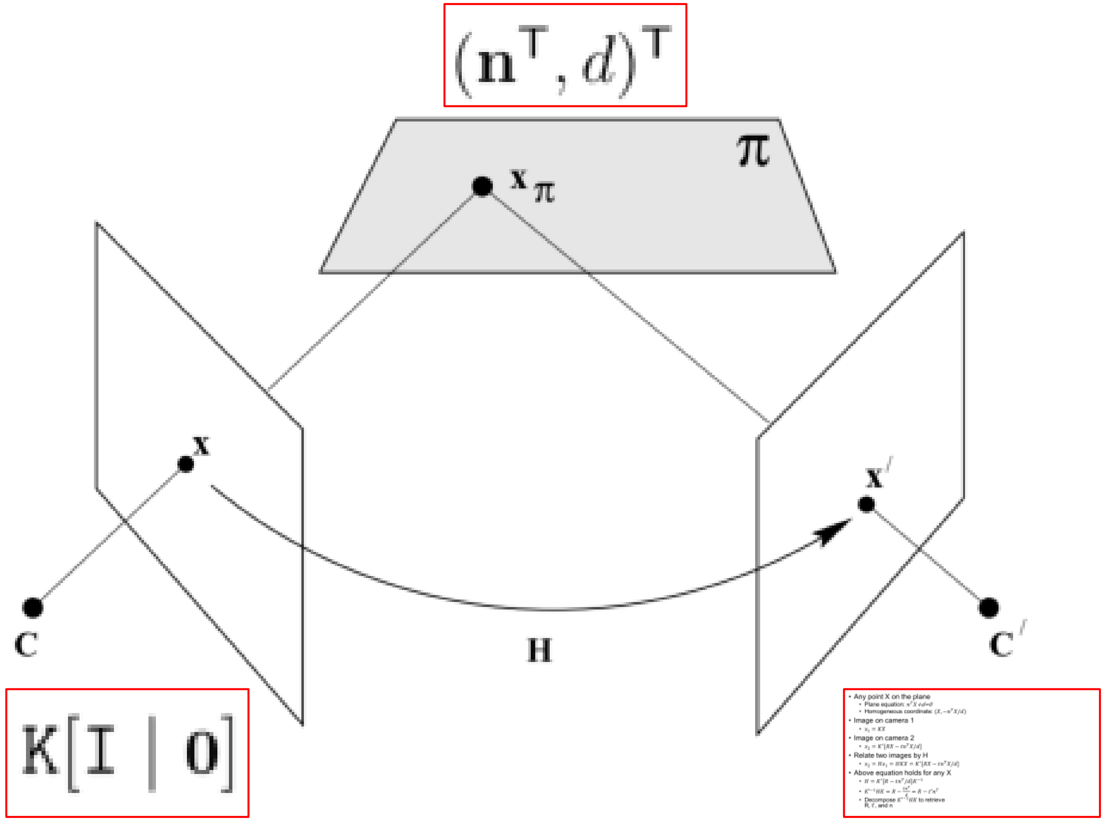
Planar Homography



Hartley & Zisserman 2003

Planar Homography

- Any point X on the plane
  - Plane equation: $n^T X + d = 0$
  - Homogeneous coordinate: $(X, -n^T X/d)$
- Image on camera 1
  - $x_1 = KX$
- Image on camera 2
  - $x_2 = K'[RX - tn^T X/d]$
- Relate two images by H
  - $x_2 = Hx_1 = HKX = K'[RX - tn^T X/d]$
- Above equation holds for any X
  - $H = K'[R - tn^T/d]K^{-1}$
  - $K'^{-1}HK = R - \frac{tn^T}{d} = R - t'n^T$
  - Decompose $K'^{-1}HK$ to retrieve R, t', and n



Hartley & Zisserman 2003

Malis, Ezio, and Manuel Vargas. "Deeper understanding of the homography decomposition for vision-based control." (2007).

Perspective-n-point (PnP) Problem

- **A calibrated camera is an angular sensor**
  - $\hat{\boldsymbol{x}}_{\boldsymbol{i}} = \boldsymbol{K}^{-1}\boldsymbol{x}_{\boldsymbol{i}} / \|\boldsymbol{K}^{-1}\boldsymbol{x}_{\boldsymbol{i}}\|$
- Visual angles between any pair of image points must be the same as the angle between corresponding 3D points
  - $\angle \boldsymbol{x}_i \boldsymbol{c} \boldsymbol{x}_j \equiv \angle \boldsymbol{p}_i \boldsymbol{c} \boldsymbol{p}_j$
- 3 pair of 2D-3D correspondences leads to 4 possible solutions
  - P3P problem
- Useful for estimating object pose
  - [**R**, **t**]

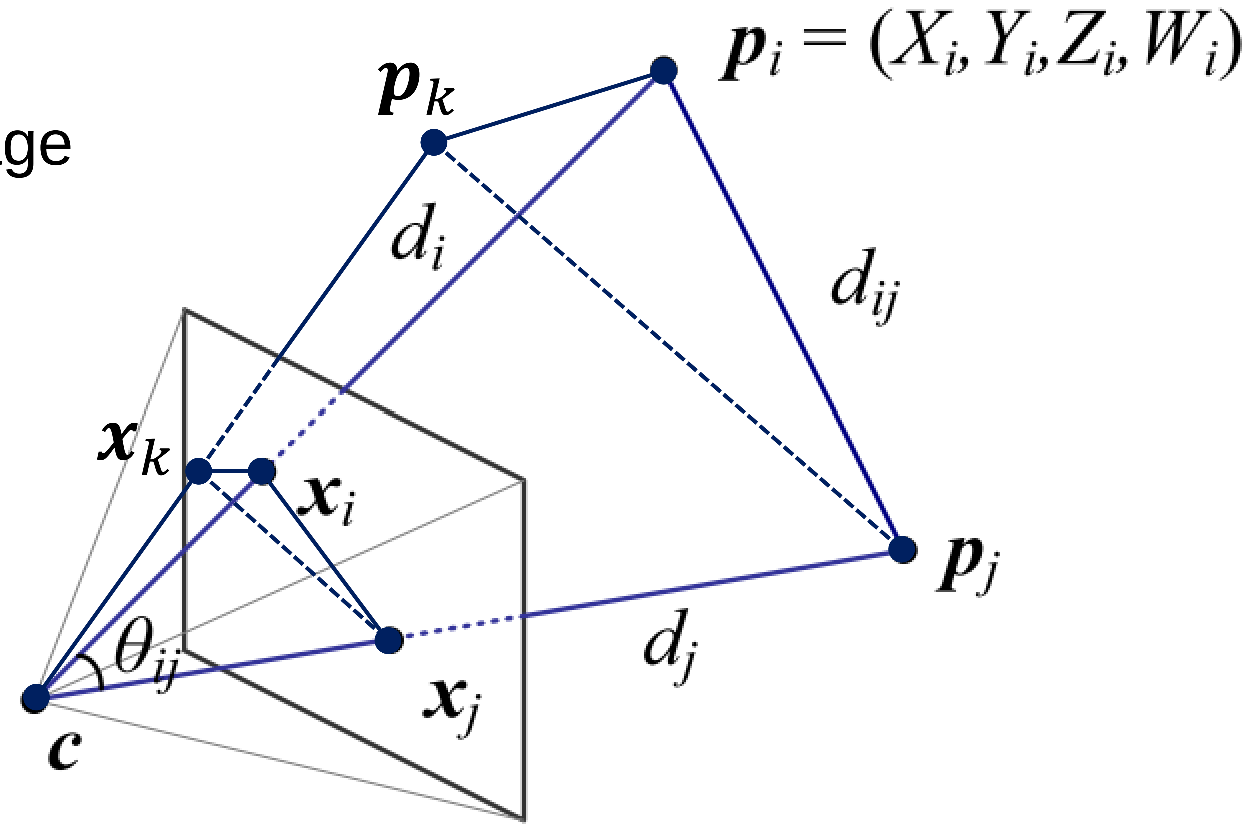


Image modified from Szeliski 2011

Hand-eye Calibration

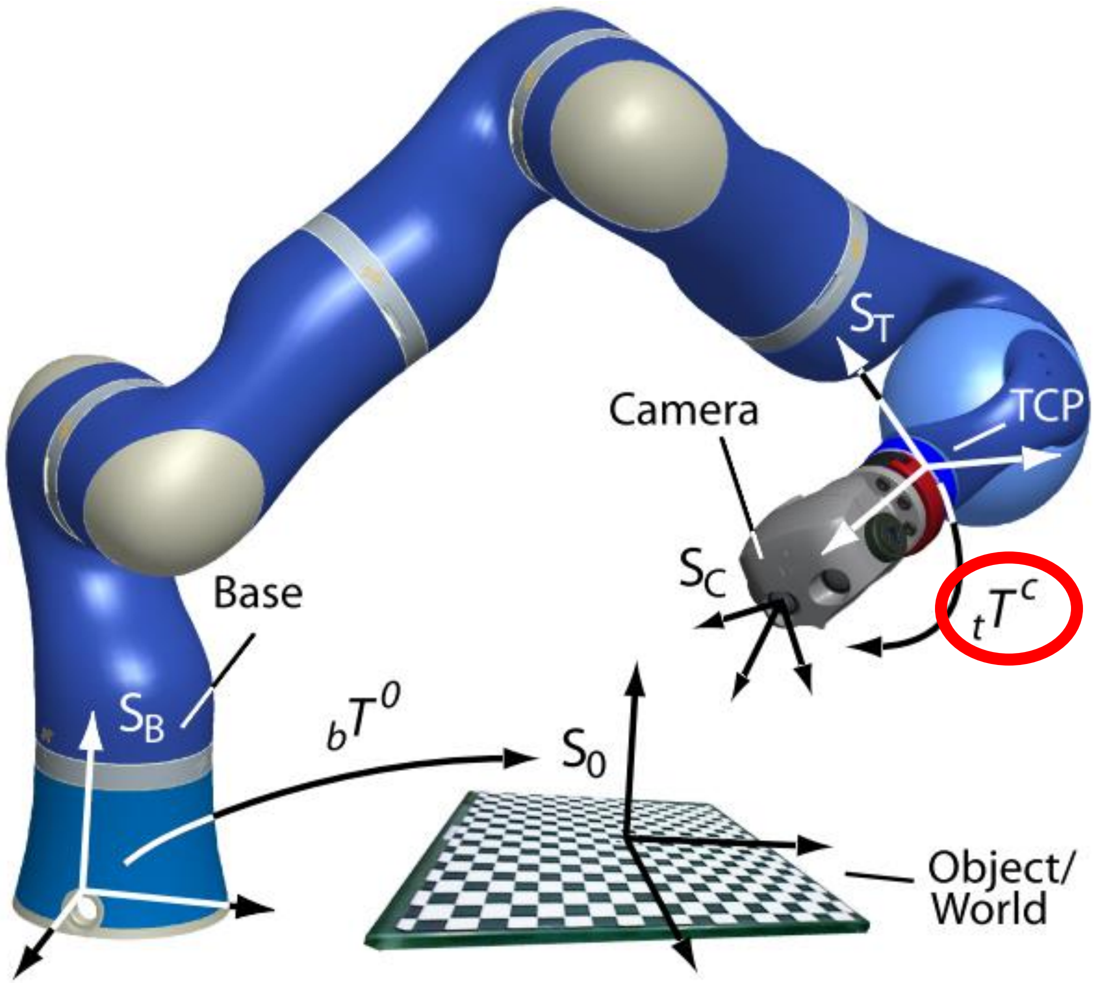


Image from Strobl & Hirzinger, ICRA'06

Hand-eye Calibration

- For any static point P in the calibration frame
  - $B_1X^{-1}A_1P = B_2X^{-1}A_2P, \forall P$
    - $B_2^{-1}B_1X^{-1} = X^{-1}A_2A_1^{-1}$
    - $XB_1^{-1}B_2 = A_1A_2^{-1}X$
- $\boldsymbol{AX = XB}$
  - $A = A_1A_2^{-1}$
  - $B = B_1^{-1}B_2$
- Solve X by observing multiple (A,B)

$$\begin{cases} A_{12}X & = & XB_{12} \\ & \vdots & \\ A_{i-1\ i}X & = & XB_{i-1\ i} \\ & \vdots & \\ A_{n-1\ n}X & = & XB_{n-1\ n} \end{cases}$$

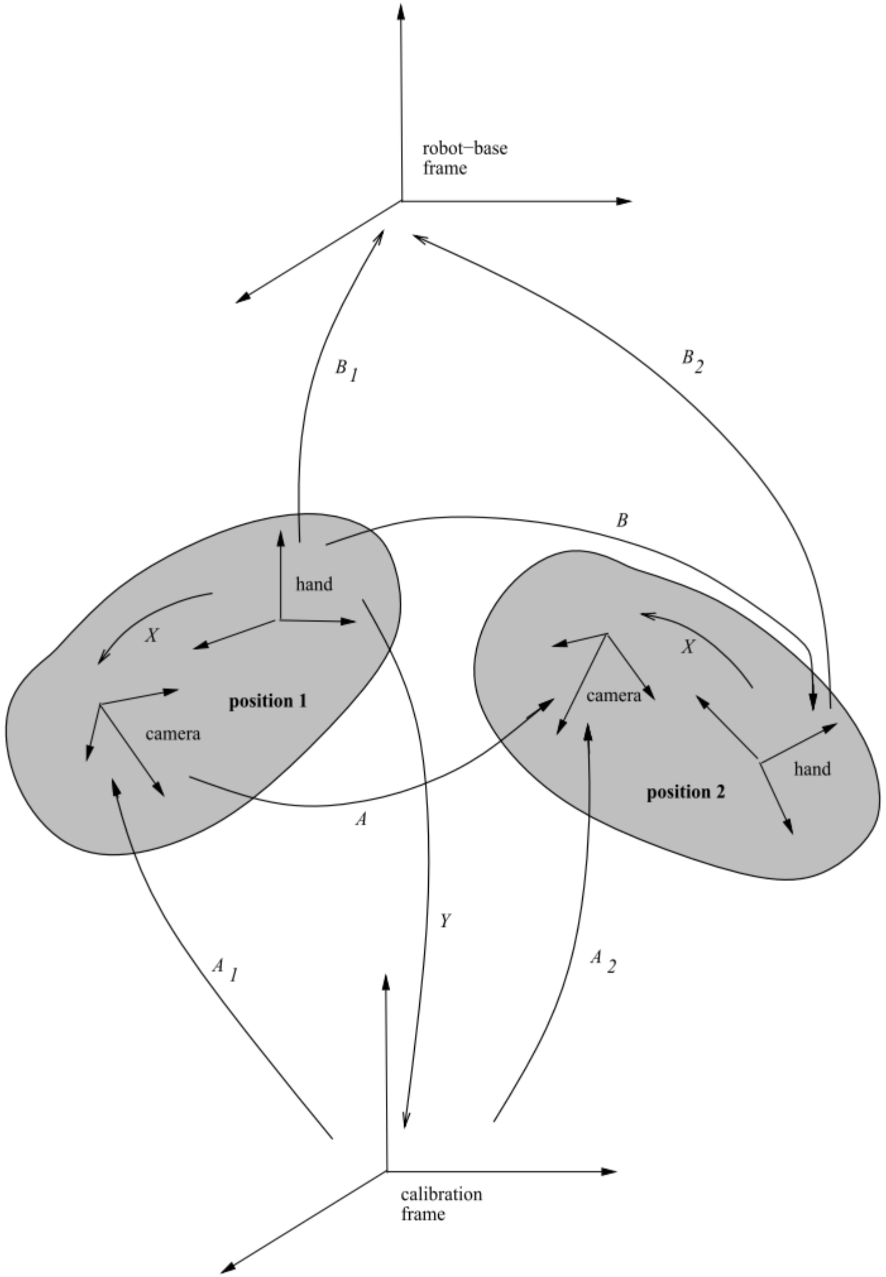


Image from Horaud 1995

Solve AX=XB

- $A = \begin{bmatrix} R_A & t_A \\ 0 & 1 \end{bmatrix}$, and so on for B and X
- $\begin{bmatrix} R_A & t_A \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_X & t_X \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R_X & t_X \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_B & t_B \\ 0 & 1 \end{bmatrix}$

- $R_A R_X = R_X R_B$
- $R_A t_X + t_A = R_X t_B + t_X \Rightarrow (R_A - I) t_X = R_X t_B - t_A$

- Solve rotation first!

Solve $R_A R_X = R_X R_B$

- Let n_B be R_B's rotation axis
  - $n_B = R_B n_B$
- Multiply both sides by R_X
  - $R_X n_B = R_X R_B n_B$
  - $= R_A R_X n_B$
- Consider R_A's rotation axis: $n_A = R_A n_A$
- So $R_X n_B$ turns out to be R_A's rotation axis
  - $\mathrm{n_A} = R_X n_B$
- R_X can be solved with multiple **(≥ 2)** pairs of (A, B)!
  - Polar decomposition on a covariance matrix $\boldsymbol{N}_A \boldsymbol{N}_B^T$
  - Orthogonal Procrustes problem

Orthogonal Procrustes Problem

$\boldsymbol{A}: 3 \times n, \boldsymbol{B}: 3 \times n, \boldsymbol{\Omega}: 3 \times 3$

$\langle \mathbf{A}, \mathbf{B} \rangle_\mathrm{F} = \sum_{i,j} \overline{A_{ij}} B_{ij}\,, = \mathrm{tr}\left(\overline{\mathbf{A}^T} \mathbf{B}\right)$

$$
\begin{aligned}
R &= \arg\min_\Omega \|\Omega A - B\|_F \quad \text{subject to} \quad \Omega^T\Omega = I, \\
&= \arg\min_\Omega \langle \Omega A - B, \Omega A - B \rangle \\
&= \arg\min_\Omega \|\Omega A\|_F^2 + \|B\|_F^2 - 2\langle \Omega A, B \rangle \\
&\boxed{\begin{aligned} &= \arg\min_\Omega \|A\|_F^2 + \|B\|_F^2 - 2\langle \Omega A, B \rangle \\ &= \arg\max_\Omega \langle \Omega, BA^T \rangle \end{aligned}} \\
&= \arg\max_\Omega \langle \Omega, U\Sigma V^T \rangle \\
&= \arg\max_\Omega \langle U^T\Omega V, \Sigma \rangle \\
&= \arg\max_\Omega \langle S, \Sigma \rangle \quad \text{where } S = U^T\Omega V
\end{aligned}
$$

More generally, the trace is *invariant under cyclic permutations*, i.e.,

$$\mathrm{tr}(ABCD) = \mathrm{tr}(BCDA) = \mathrm{tr}(CDAB) = \mathrm{tr}(DABC).$$

$$\xRightarrow{S^*=I} \quad \begin{aligned} I &= U^T R V \\ R &= UV^T \end{aligned}$$

Next Week

+ Feature detection
  - Harris/FAST/DoG
+ Feature description & matching
  - SIFT/SURF

++ Linear & total least square

* RANSAC
  - Intuitions behind RANSAC
  - How RANSAC works
  - Why minimal solution is important
  - More example problems
  - Variations

*: know how to code
++: know how to derive
+: know the concept

References for Next Week

- HZ2003:
  - Section 4.7, 4.8, 11.6
- Corke 2011:
  - Section 14.2.3
- Sz2022:
  - Section 7.1, 7.2, 8.1.4

- DeTone, D., Malisiewicz, T. and Rabinovich, A., 2018. Superpoint: Self-supervised interest point detection and description. In *CVPR workshops* (pp. 224-236).
- Sarlin, P.E., DeTone, D., Malisiewicz, T. and Rabinovich, A., 2020. Superglue: Learning feature matching with graph neural networks. In *CVPR* (pp. 4938-4947).