# CS4277 / CS5477
# 3D Computer Vision

## Lecture 6:
## The fundamental and essential matrices

Assoc. Prof. Lee Gim Hee

AY 2022/23

Semester 2

# Course Schedule

| Week | Date | Topic | Assignments |
|---|---|---|---|
| 1 | 11 Jan | 2D and 1D projective geometry | **Assignment 0:** Getting started with Python (Ungraded) |
| 2 | 18 Jan | 3D projective geometry, Circular points and Absolute conic | |
| 3 | 25 Jan | Rigid body motion and Robust homography estimation | |
| 4 | 01 Feb | Camera models and calibration | **Assignment 1:** Metric rectification and robust homography (10%)<br>**Due:** 2359hrs, 07 Feb |
| 5 | 08 Feb | Single view metrology | **Assignment 2**: Affine 3D measurement from vanishing line and point (10%)<br>**Due:** 2359hrs, 14 Feb |
| 6 | 15 Feb | The Fundamental and Essential matrices | |
| - | 22 Feb | Semester Break | No lecture |
| **7** | **01 Mar** | **Mid-term Quiz (20%)**<br>Lecture: Generalized cameras | **In-person Quiz (LT 15, 1900hrs – 2000hrs)**<br>Lecture: 2000hrs – 2130hrs |
| 8 | 08 Mar | Absolute pose estimation from points or lines | |
| 9 | 15 Mar | Three-view geometry from points and/or lines | |
| 10 | 22 Mar | Structure-from-Motion (SfM) and bundle adjustment | **Assignment 3:** SfM and Bundle adjustment (10%)<br>**Due:** 2359hrs, 28 Mar |
| 11 | 29 Mar | Two-view and multi-view stereo | **Assignment 4:** Dense 3D model from multi-view stereo (10%)<br>**Due:** 2359hrs, 04 Apr |
| 12 | 05 Apr | 3D Point Cloud Processing | |
| 13 | 12 Apr | Neural Field Representations | |

Final Exam: 03 MAY 2023

# Learning Outcomes

- Students should be able to:

1. Describe the epipolar geometry between two views.
2. Estimate fundamental / essential matrix with 8 point correspondences.
3. Decompose fundamental matrix into the camera matrices of two views.
4. Find rotation and translation between two views from the essential matrix.
5. Recover 3D structures with linear triangulation, and do stratified reconstruction from uncalibrated reconstruction.

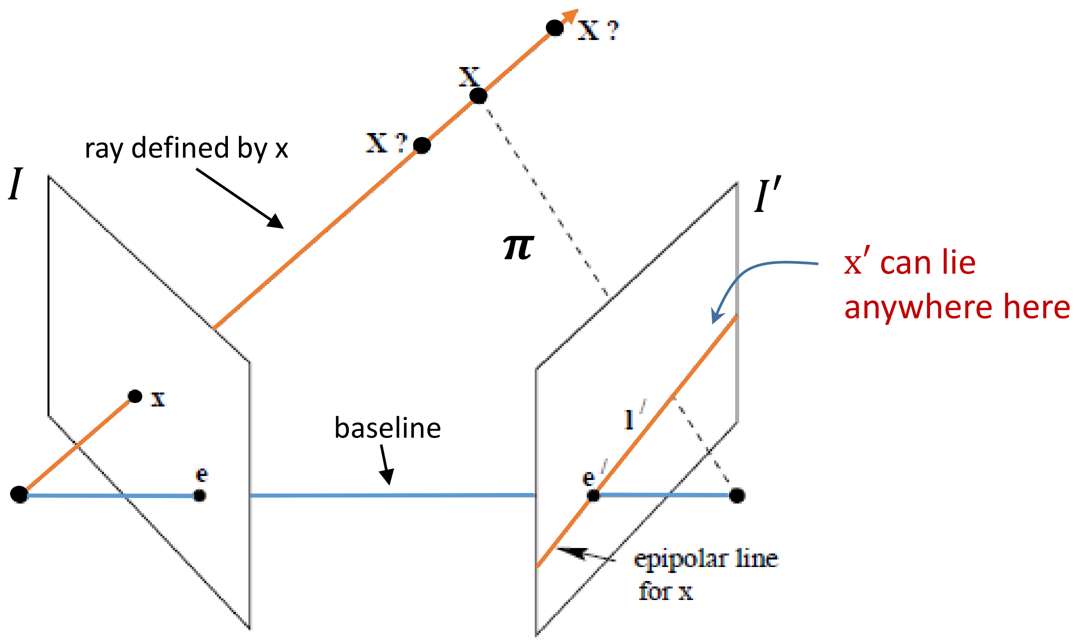
# Acknowledgements

- A lot of slides and content of this lecture are adopted from:

1. R. Hartley, and A. Zisserman: “Multiple view geometry in computer vision”, Chapter 9, 10, 11.

2. Y. Ma, S. Soatto, J. Kosecka, S. S. Sastry, “ An invitation to 3-D vision”, Chapter 6.

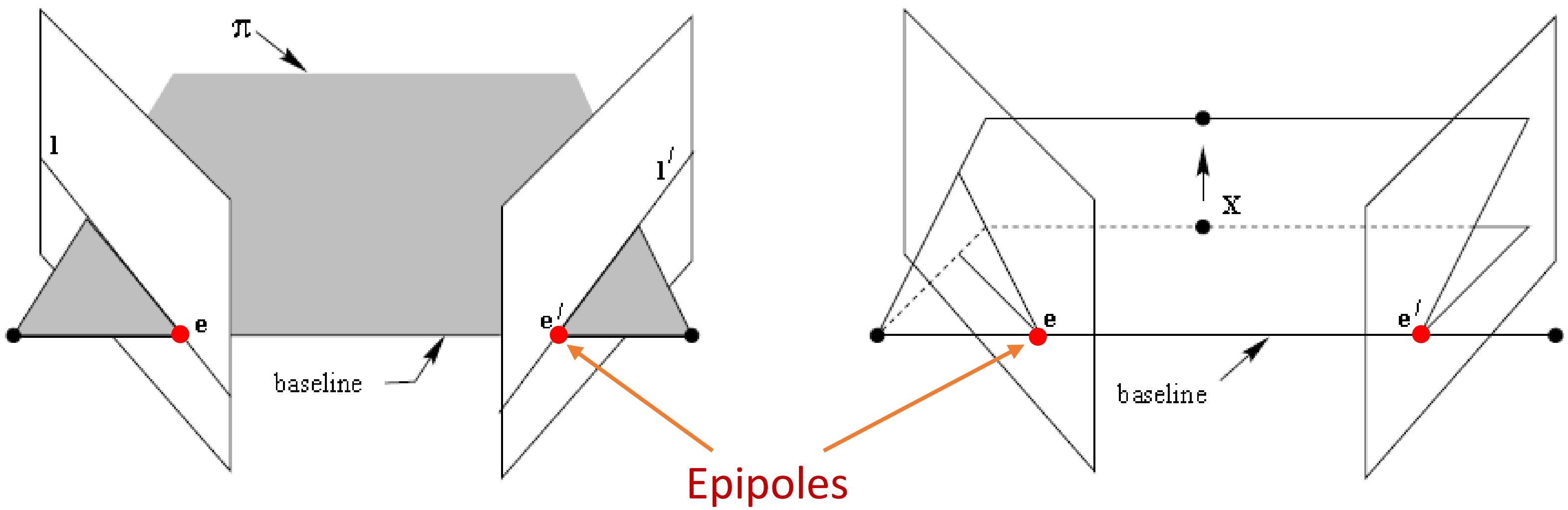
# The Epipolar Geometry




- The image point $\mathrm{x}$ in $I$ back-projects to a ray, and this ray projects to $I'$ as the epipolar line $\mathbf{l}'$.
- The corresponding point $\mathbf{x}'$ can lie anywhere on $\mathbf{l}'$.
- Epipolar plane $\boldsymbol{\pi}$ is determined by the baseline and ray defined by $\mathbf{x}$.

Image Source: R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision"

# The Epipolar Geometry: Terminology

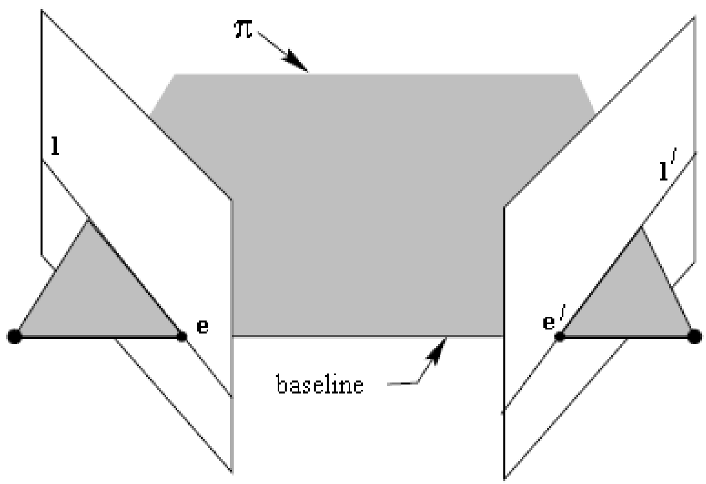


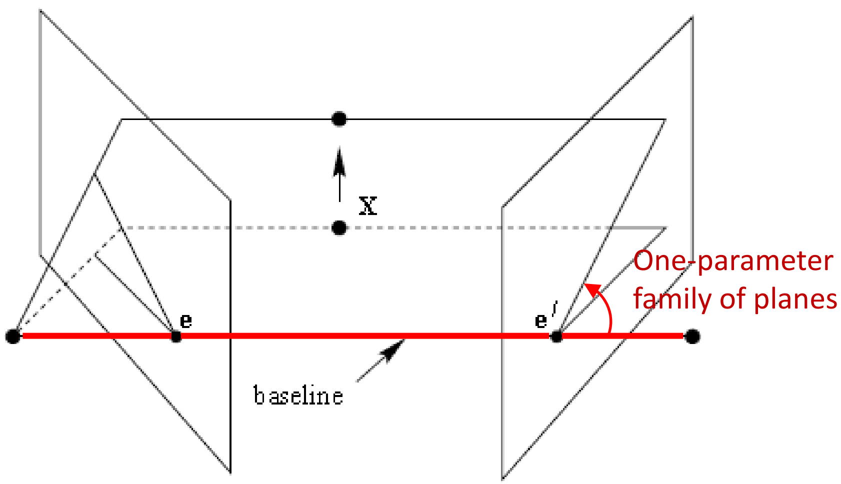

**Epipoles (e, e′):**

- Point of intersection of the line joining the camera centers (baseline) with the image plane.
- Equivalently, it is the image in one view of the camera center of the other view.
- Also, the vanishing point of the baseline (translation) direction.

Image Source: R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision"

# The Epipolar Geometry: Terminology



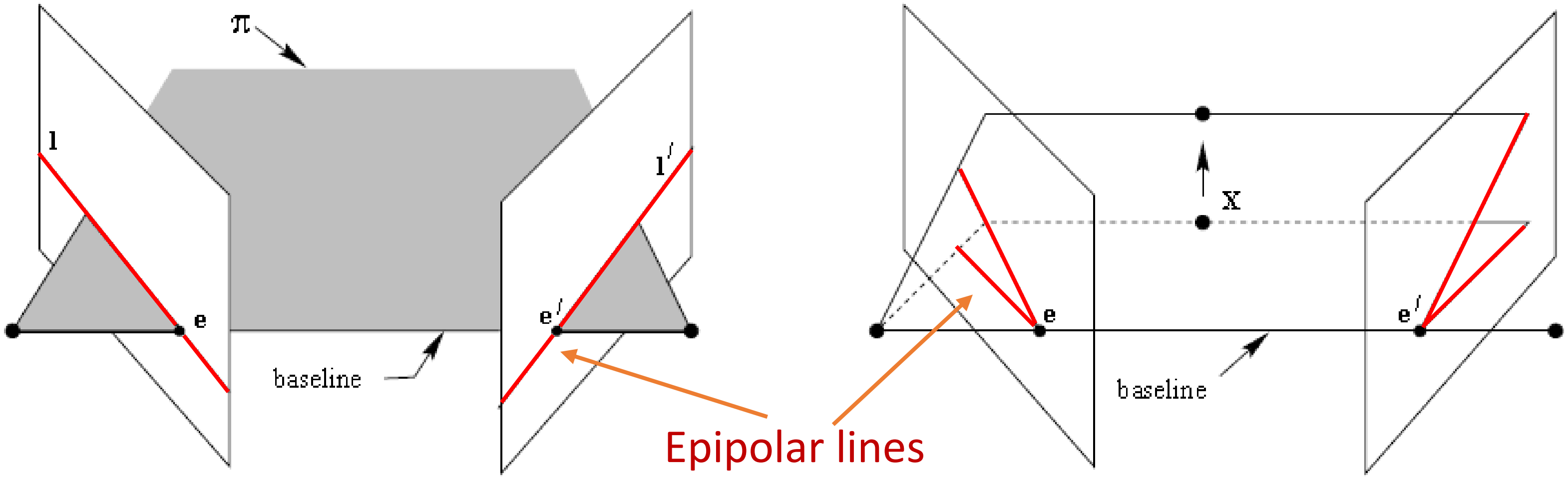




**Epipolar plane $\pi$ :**

- A plane containing the baseline.
- There is a one-parameter family (a pencil) of epipolar planes.

Image Source: R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision"

# The Epipolar Geometry: Terminology




**Epipolar lines $(\mathbf{l}, \mathbf{l}')$ :**

- The intersection of an epipolar plane with the image plane.
- All epipolar lines intersect at the epipole.
- An epipolar plane intersects the left and right image plane in epipolar lines, and defines the correspondences between the lines.

Image Source: R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# The Fundamental Matrix

- The fundamental matrix is the algebraic representation of epipolar geometry.

- Gives the projective mapping relationship between a point $\mathbf{x}$ on one image to a line $\mathbf{l}'$ on the other.

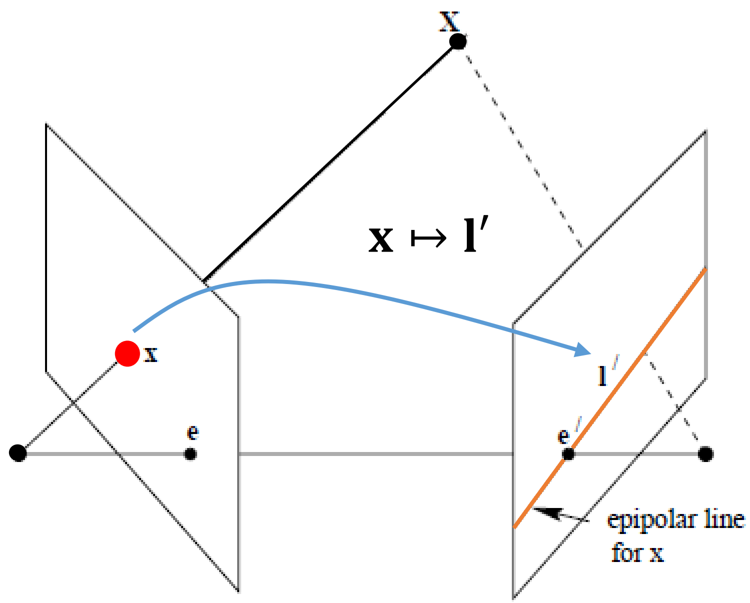



Image Source: R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision"

# F Matrix: Geometric Derivation

- The mapping $\mathbf{x} \mapsto \mathbf{l}'$ may be decomposed into two steps:

1. The point $\mathbf{x}$ is mapped to some point $\mathbf{x}'$ in the other image lying on the epipolar line $\mathbf{l}'$; this point $\mathbf{x}'$ is a potential match for the point $\mathbf{x}$.

2. The epipolar line $\mathbf{l}'$ is obtained as the line joining $\mathbf{x}'$ to the epipole $\mathbf{e}'$.

# F Matrix: Geometric Derivation

**Step 1: Point transfer via a plane.**

- Consider a plane $\boldsymbol{\pi}$ in space not passing through either of the two camera centres and contains the point $\mathbf{X}$.
- Thus, there is a 2D homography $\mathrm{H}_{\boldsymbol{\pi}}$ mapping each $\mathbf{x}_i$ to $\mathbf{x}'_i$.

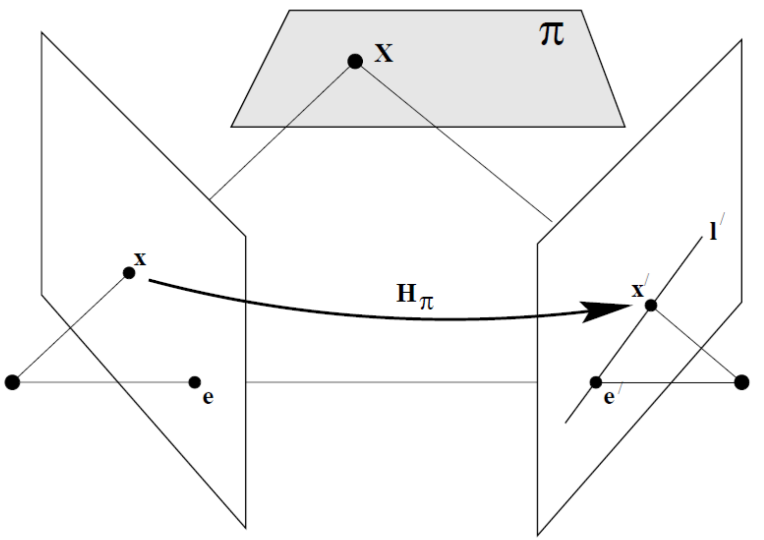



Image Source: R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision"

# F Matrix: Geometric Derivation

**Step 2: Constructing the epipolar line.**

- Given the point $\mathbf{x}'$, the epipolar line $\mathbf{l}'$ passing through $\mathbf{x}'$ and the epipole $\mathbf{e}'$ can be written as $\mathbf{l}' = \mathbf{e}' \times \mathbf{x}' = [\mathbf{e}']_\times \mathbf{x}'$.
- Since $\mathbf{x}'$ may be written as $\mathbf{x}' = \mathrm{H}_{\boldsymbol{\pi}}\mathbf{x}$, we have:

$$\mathbf{l}' = [\mathbf{e}']_\times \mathrm{H}_{\boldsymbol{\pi}} \mathbf{x} = \mathrm{F}\mathbf{x} \ ,$$

where we define $\mathrm{F} = [\mathbf{e}']_\times \mathrm{H}_{\boldsymbol{\pi}}$ as the fundamental matrix.

# Cross-Product as Matrix Multiplication

- Vector cross product can be expressed as the product of a skew-symmetric matrix and a vector:

$$\mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times \mathbf{b} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

# F Matrix: Geometric Derivation

- The fundamental matrix $\mathrm{F}$ may be written as:

$$\mathrm{F} = [\mathbf{e}']_{\times} \mathrm{H}_{\boldsymbol{\pi}},$$

- where $\mathrm{H}_{\boldsymbol{\pi}}$ is the transfer mapping from one image to another via any plane.

- Furthermore, since $[\mathbf{e}']_{\times}$ has rank 2 and $\mathrm{H}_{\boldsymbol{\pi}}$ rank 3, $\mathrm{F}$ is a matrix of rank 2.

# F Matrix: Geometric Derivation

- Geometrically, $\mathrm{F}$ 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 of $\mathbb{P}^2 \mapsto \mathbb{P}^1$, and hence must have rank 2.

- **Note:** The plane is simply used here as a means of defining a point map from one image to another, but not required for F to exist.

# F Matrix: Algebraic Derivation

- The form of the fundamental matrix in terms of the two camera projection matrices, $\mathrm{P}$ and $\mathrm{P}'$, may be derived algebraically.

- The back-projected ray from $\mathbf{x}$ is given by:

$$\mathbf{X}(\lambda) = \mathrm{P}^{+}\mathbf{x} + \lambda\mathbf{C} \, ,$$

where

- $\mathrm{P}^{+}$ is the pseudo-inverse of $\mathrm{P}$, i.e. $\mathrm{P}\mathrm{P}^{+} = \mathrm{I}$
- $\mathbf{C}$ the null-vector of $\mathrm{P}$, i.e. the camera center, $\mathrm{P}\mathbf{C} = \mathbf{0}$
- The ray is parametrized by the scalar $\lambda$

# F Matrix: Algebraic Derivation

- Let's consider two points on the ray: $\mathrm{P}^{+}\mathbf{x}$ at $\lambda = 0$ and the first camera center $\mathbf{C}$ at $\lambda = \infty$.

- These two points are imaged by the second camera $\mathrm{P}'$ at $\mathrm{P}'\mathrm{P}^{+}\mathbf{x}$ and $\mathrm{P}'\mathbf{C}$, respectively in the second view.

- The epipolar line is the line joining these two projected points, i.e. $\mathbf{l}' = (\mathrm{P}'\mathbf{C}) \times (\mathrm{P}'\mathrm{P}^{+}\mathbf{x})$.

- The point $\mathrm{P}'\mathbf{C}$ is the epipole in the second image, i.e. $\mathbf{e}'$.

# F Matrix: Algebraic Derivation

- Thus, $\mathbf{l}' = [\mathbf{e}']_\times (\mathtt{P}'\mathtt{P}^+)\mathbf{x} = \mathtt{F}\mathbf{x}$, where $\mathtt{F}$ is the matrix

$$\mathtt{F} = [\mathbf{e}']_\times \mathtt{P}'\mathtt{P}^+ .$$

- This is similar to $\mathtt{F} = [\mathbf{e}']_\times \mathtt{H}_{\boldsymbol{\pi}}$ that we have derived geometrically.

- We can see that the homography takes the form

$$\mathtt{H}_{\boldsymbol{\pi}} = \mathtt{P}'\mathtt{P}^+$$

in terms of the two camera matrices.

# F Matrix: Algebraic Derivation

- **Remarks:** Note that this derivation breaks down in the case where the two camera centres are the same.

**Proof:**

$\mathbf{e}' = \mathrm{P}'\mathbf{C} = \mathbf{0}$, when $\mathbf{C} = \mathbf{C}'$. It follows that:

$$\mathrm{F} = [\mathrm{e}']_{\times}\mathrm{P}'\mathrm{P}^{+} = \mathbf{0}. \qquad \square$$

# F Matrix: Algebraic Derivation

**Example:** Suppose the camera matrices are those of a calibrated stereo rig with the world origin at the first camera

$$\mathtt{P} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}] \qquad \mathtt{P}' = \mathtt{K}'[\mathtt{R} \mid \mathbf{t}].$$

Then

$$\mathtt{P}^{+} = \begin{bmatrix} \mathtt{K}^{-1} \\ \mathbf{0}^{\mathsf{T}} \end{bmatrix} \qquad \mathbf{C} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}$$

and

$$\begin{aligned} \mathtt{F} &= [\mathtt{P}'\mathbf{C}]_{\times}\mathtt{P}'\mathtt{P}^{+} \\ &= [\mathtt{K}'\mathbf{t}]_{\times}\mathtt{K}'\mathtt{R}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}[\mathbf{t}]_{\times}\mathtt{R}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}\mathtt{R}[\mathtt{R}^{\mathsf{T}}\mathbf{t}]_{\times}\mathtt{K}^{-1} = \mathtt{K}'^{-\mathsf{T}}\mathtt{R}\mathtt{K}^{\mathsf{T}}[\mathtt{K}\mathtt{R}^{\mathsf{T}}\mathbf{t}]_{\times} \end{aligned}$$

# F Matrix: Algebraic Derivation

- Note that the epipoles (defined as the image of the other camera centre) are:

$$\mathbf{e} = \mathrm{P}\begin{pmatrix} -\mathrm{R}^\mathsf{T}\mathbf{t} \\ 1 \end{pmatrix} = \text{-}\mathrm{K}\mathrm{R}^\mathsf{T}\mathbf{t} \qquad \mathbf{e}' = \mathrm{P}'\begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} = \mathrm{K}'\mathbf{t}.$$

Thus

$$\mathrm{F} = [\mathbf{e}']_\times \mathrm{K}'\mathrm{R}\mathrm{K}^{-1} = \mathrm{K}'^{-\mathsf{T}}[\mathbf{t}]_\times \mathrm{R}\mathrm{K}^{-1} = \mathrm{K}'^{-\mathsf{T}}\mathrm{R}[\mathrm{R}^\mathsf{T}\mathbf{t}]_\times \mathrm{K}^{-1} = \mathrm{K}'^{-\mathsf{T}}\mathrm{R}\mathrm{K}^\mathsf{T}[\mathbf{e}]_\times.$$

# Correspondence Condition

- For any pair of corresponding points $\mathbf{x} \leftrightarrow \mathbf{x}'$ in two images:

$$\mathbf{x}'^{\top} \mathrm{F} \mathbf{x} = 0$$

**Proof:**

$\mathbf{x}'$ lies on the epipolar line $\mathbf{l}' = \mathrm{F}\mathbf{x}$ corresponding to the point $\mathbf{x}$

$$\Rightarrow 0 = \mathbf{x}'^{\top} \mathbf{l}' = \mathbf{x}'^{\top} \mathrm{F} \mathbf{x} = 0$$

# Correspondence Condition

- The importance of the relation $\mathbf{x}'^{\top}\mathrm{F}\mathbf{x} = 0$ is that it gives a way of characterizing the fundamental matrix without reference to the camera matrices.

- That is the relation is only in terms of corresponding image points, and this enables F to be computed from image correspondences alone.

- We will discuss the details later: how many correspondences are required to compute $\mathrm{F}$ from $\mathbf{x}'^{\top}\mathrm{F}\mathbf{x} = 0$ ?

# Properties of the F Matrix

- **Transpose**:
  - $\mathrm{F}$ is the fundamental matrix of the pair of cameras $(\mathrm{P}, \mathrm{P}')$
  - $\mathrm{F}^\top$ is the fundamental matrix of the pair in the opposite order: $(\mathrm{P}', \mathrm{P})$
- **Epipolar lines**:
  - For any point $\mathbf{x}$ in first image, corresponding epipolar line is $\mathbf{l}' = \mathrm{F}\mathbf{x}$
  - $\mathbf{l} = \mathrm{F}^\top \mathbf{x}'$ represents epipolar line corresponding to $\mathbf{x}'$ in second image
- **Epipole**:
  - For any point $\mathbf{x}$ (other than $\mathbf{e}$) the epipolar line $\mathbf{l}' = \mathrm{F}\mathbf{x}$ contains the epipole $\mathbf{e}'$
  - $\mathbf{e}'$ satisfies $\mathbf{e}'^\top(\mathrm{F}\mathbf{x}) = (\mathbf{e}'^\top \mathrm{F})\mathbf{x} = 0$ for all $\mathbf{x}$
  - $\mathbf{e}'^\top \mathrm{F} = \mathbf{0}$, i.e. $\mathbf{e}'$ is the left null-vector of $\mathrm{F}$
  - $\mathrm{F}\mathbf{e} = \mathbf{0}$, i.e. $\mathbf{e}$ is the right null-vector of $\mathrm{F}$

# Properties of the F Matrix

- **7 degrees of freedom (9 elements – 2 dof)**:
  - 3 x 3 homogenoueous matrix with 8 independent ratios ⇒ -1 dof
  - $\det(\mathrm{F}) = 0$ ⇒ -1 dof
- **Not a proper correlation (not invertible)**:
  - Projective map taking a point to a line
  - A point in first image $\mathbf{x}$ defines a line in the second $\mathbf{l} = \mathrm{F}\mathbf{x}$, i.e. epipolar line of $\mathbf{x}$
  - If $\mathbf{l}$ and $\mathbf{l}'$ are corresponding epipolar lines, then any point $\mathbf{x}$ on $\mathbf{l}$ is mapped to the same line $\mathbf{l}'$
  - This means no inverse mapping, and $\mathrm{F}$ is not of full rank

# Summary of F Matrix Properties

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

Source: Page 246, R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

The fundamental matrix
Used in stereo geometry
A matrix with nine entries
It's square with size 3 by 3
Has seven degrees of freedom
It has a rank deficiency
It's only of rank two
Call the matrix F and you'll see...

Two points that correspond
Column vectors called x and x-prime
x-prime transpose times F times x
Equals zero every time

The epipolar constraint
Involves epipolar lines
Postmultiplying F by x
Results in vector l-prime
It's the epipolar line
In the other view passing through x-prime
A three component vector
Of homogeneous design

The left and right nullspaces of F
Are the epipoles e-prime and e
All of the epipolar lines
Should pass through these

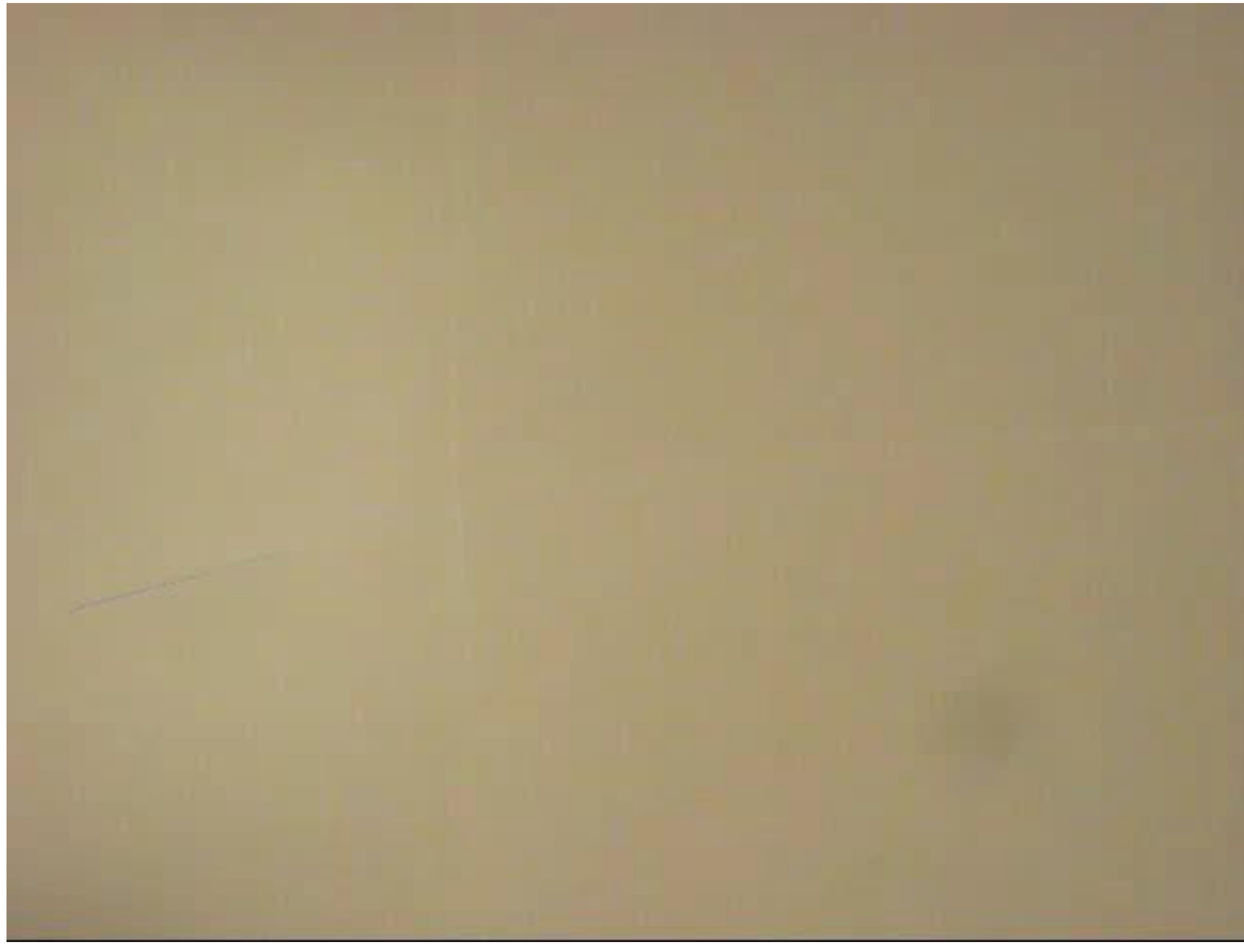
Here's a linear estimation example:
Take a set of 8 point samples
Construct a matrix, take the SVD
And the elements of F are in the last column of V

If you try to estimate
F with a coplanar set of points
Your sample set will be degenerate
And will not bring you joy

When doing the estimation
If you don't perform rank deprivation
Your epipolar lines
And the epipoles will not coincide

But if your scene has three views
The trifocal tensor is what you'd use
Constraints from the third view act like glue
That can't be determined from just two views

# The Fundamental Matrix Song



Source: http://danielwedge.com/fmatrix/

# The Epipolar Line Homography

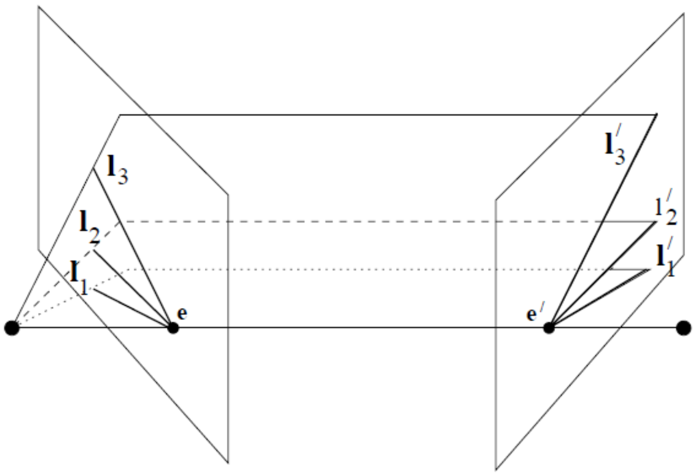
- Suppose $\mathbf{l}$ and $\mathbf{l}'$ are corresponding epipolar lines, and $\mathbf{k}$ is any line not passing through the epipole $\mathbf{e}$, then $\mathbf{l}$ and $\mathbf{l}'$ are related by $\mathbf{l}' = \mathrm{F}[\mathbf{k}]_\times \mathbf{l}$, where $\mathrm{F}[\mathbf{k}]_\times$ is a homography.
- Symmetrically, $\mathbf{l} = \mathrm{F}^\top [\mathbf{k}'] \times \mathbf{l}'$.

**Proof:**

The expression $[\mathbf{k}]_\times \mathbf{l} = \mathbf{k} \times \mathbf{l}$ is the point of intersection of the two lines **k** and **l**, and hence a point on the epipolar line $\mathbf{l}$ – call it $\mathbf{x}$.

Hence, $\mathrm{F}[\mathbf{k}]_\times \mathbf{l} \;=\; \mathrm{F}\mathbf{x}$ is the epipolar line corresponding to the point $\mathbf{x}$, namely the line $\mathbf{l}$.

# The Epipolar Line Homography




Homography in $\mathbb{P}^2$:

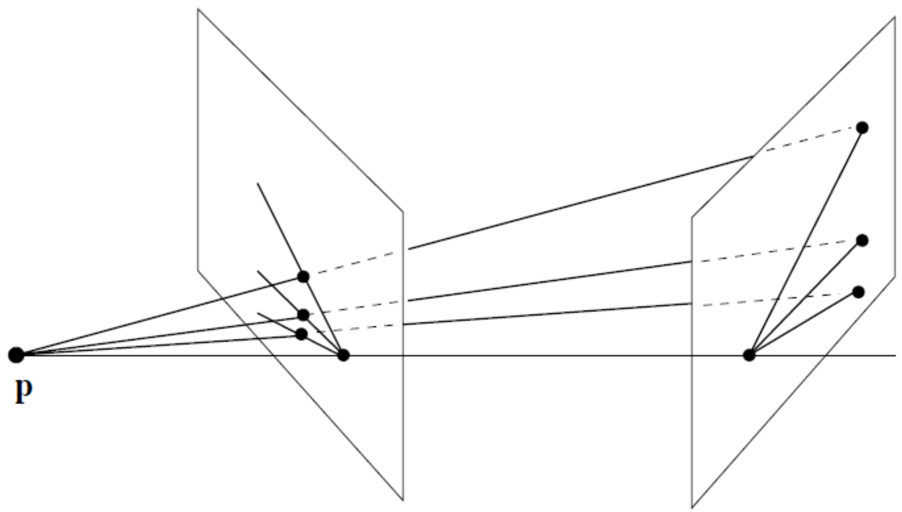
$$\mathbf{l}' = \mathrm{F}[\mathbf{k}]_\times \mathbf{l}$$

$$\mathbf{l} = \mathrm{F}^\mathsf{T}[\mathbf{k}'] \times \mathbf{l}'$$

- There is a pencil of epipolar lines in each image centred on the epipole.
- The correspondence between epipolar lines, $\mathbf{l}_i \leftrightarrow \mathbf{l}_i$, is defined by the pencil of planes with axis the baseline.

Source: Page 246, R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# The Epipolar Line Homography




- The corresponding lines are related by a perspectivity with centre any point $\mathbf{p}$ on the baseline.
- It follows that the correspondence between epipolar lines in the pencils is a 1D homography (c.f. Lecture 1 on cross-ratio).

Source: Page 246, R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# Special Motion: Pure Translation

- Suppose the motion of the cameras is a pure translation with no rotation ($\mathrm{R} = \mathrm{I}$) and no change in the internal parameters ($\mathrm{K} = \mathrm{K}'$).

- The two cameras are $\mathrm{P} = \mathrm{K}[\mathrm{I} \mid \mathbf{0}]$ and $\mathrm{P}' = \mathrm{K}[\,\mathrm{I} \mid \mathbf{t}\,]$, and

$$\mathrm{F} = [\mathbf{e}']_\times \mathrm{K}\mathrm{K}^{-1} = [\mathbf{e}']_\times .$$

- F is skew-symmetric and has only 2 degrees of freedom, which correspond to the position of the epipole.

# Special Motion: Pure Translation

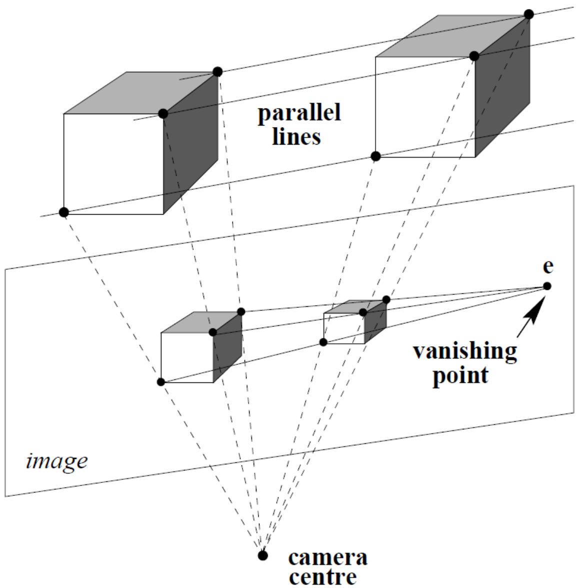
- The epipolar line of $\mathbf{x}$ is $\mathbf{l}' = \mathrm{F}\mathbf{x} = [\mathbf{e}]_\times \mathbf{x}$, and $\mathbf{x}'$ lies on this line since $\mathbf{x}'^\top [\mathbf{e}]_\times \mathbf{x} = 0$.

- That is $\mathbf{x}$, $\mathbf{x}'$ and $\mathbf{e} = \mathbf{e}'$ are collinear (assuming both images are overlaid on top of each other).

- This collinearity property is termed auto-epipolar, and **does not hold** for general motion.

# Special Motion: Pure Translation

**Example 1:**



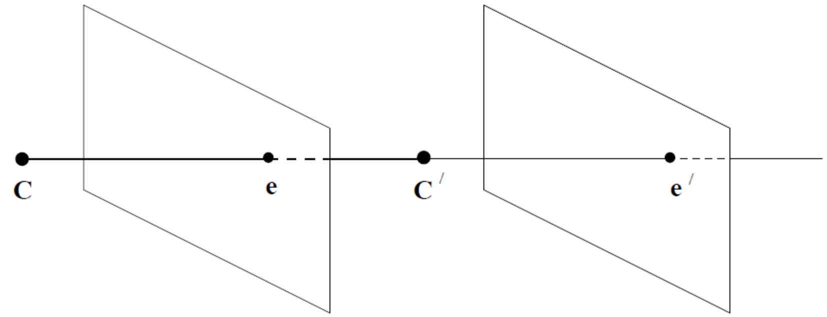

- We may consider the equivalent situation of pure translation.
- Camera is stationary, and the world undergoes a translation $-\mathbf{t}$.
- 3D points appear to slide along parallel rails.
- The images of these parallel lines intersect in a vanishing point corresponding to the translation direction.
- The epipole $\mathbf{e}$ is the vanishing point.

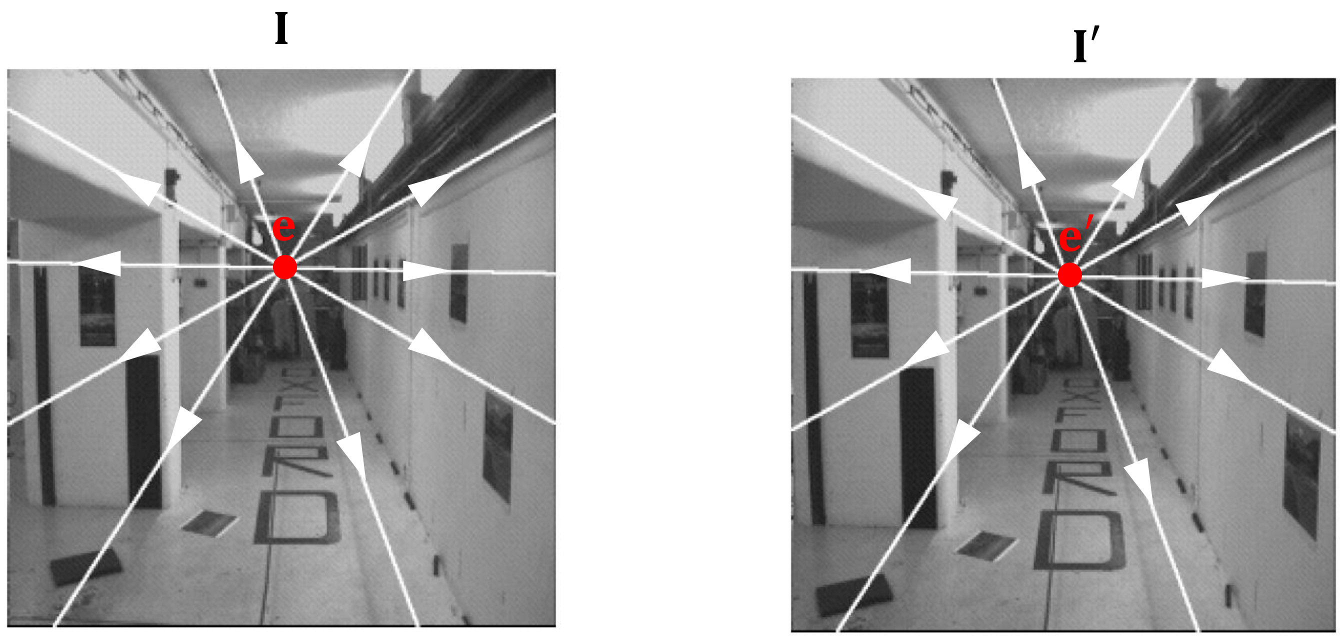
Source: Page 246, R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision"

# Special Motion: Pure Translation

**Example 2:**

Epipole remains fixed and point correspondences appear to move radially along the epipolar lines.

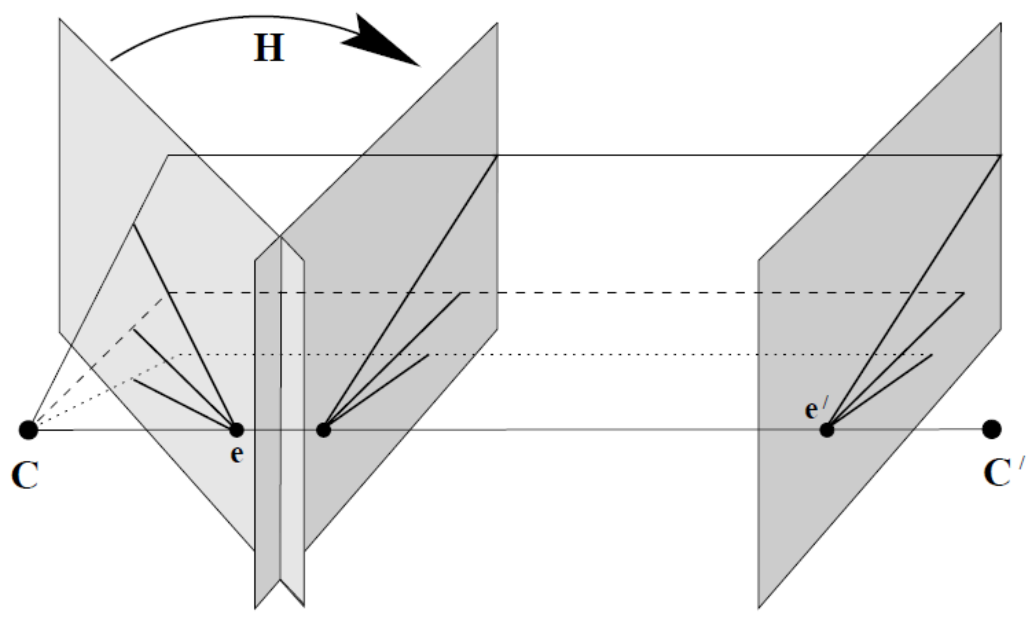



Source: Page 246, R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# General Motion

- A general motion and its effect on the fundamental matrix can be separate into a pure rotation followed by a pure translation.




Source: Page 246, R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision"

# General Motion

- Now the the two cameras are given by $\mathrm{P} = \mathrm{K}[\mathrm{I} \mid \mathbf{0}]$ and $\mathrm{P} = \mathrm{K}[\mathrm{R} \mid \mathbf{t}]$.

- The pure rotation may be simulated by: $\mathtt{H} = \mathtt{K}'\mathtt{R}\mathtt{K}^{-1} = \mathtt{H}_\infty$, where $\mathrm{H}_\infty$ is the infinite homography.

- As seen earlier, the fundamental matrix $\tilde{\mathrm{F}}$ under pure translation is given by $\tilde{\mathrm{F}} = [\mathbf{e}']_\times$.

- Since $\mathrm{F} = [\mathbf{e}']_\times \mathrm{K}'\mathrm{R}\mathrm{K}^{-1}$ (c.f. algebraic derivation), we have $\mathrm{F} = \tilde{\mathrm{F}}\mathrm{H}_\infty = [\mathbf{e}']_\times \mathrm{H}_\infty$.

Source: Page 246, R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# Retrieving the Camera Matrices

- To this point we have examined the properties of $\mathrm{F}$ and of image relations for a point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$.

- We now turn to one of the most significant properties of $\mathrm{F}$, that the matrix may be used to determine the camera matrices of the two views.

# Projective Invariance

- The fundamental matrices corresponding to the pairs of camera matrices $(\mathrm{P}, \mathrm{P}')$ and $(\mathrm{PH}, \mathrm{P}'\mathrm{H})$ are the same.
- $\mathrm{H}$ is a $4 \times 4$ matrix representing a projective transformation of 3-space.

**Proof:**

- Observe that $\mathrm{P}\mathbf{X} = (\mathrm{PH})(\mathrm{H}^{-1}\mathbf{X})$, and similarly for $\mathrm{P}'$.
- Thus, if $\mathbf{x} \leftrightarrow \mathbf{x}'$ are matched points with respect to the pair of cameras $(\mathrm{P}, \mathrm{P}')$, corresponding to a 3D point $\mathbf{X}$.
- Then they are also matched points with respect to the pair of cameras $(\mathrm{PH}, \mathrm{P}'\mathrm{H})$, corresponding to the point $\mathrm{H}^{-1}\mathbf{X}$.

$\square$

# Projective Invariance

- Thus, although a pair of camera matrices $(\mathrm{P}, \mathrm{P}')$ uniquely determine a fundamental matrix $\mathrm{F}$, the converse is not true.

- The fundamental matrix determines the pair of camera matrices at best up to right-multiplication by a 3D projective transformation.

Given: $(\mathrm{P}, \mathrm{P}') \xrightarrow{\text{Unique}} \mathrm{F}$,

Given: $\mathrm{F} \xrightarrow{\text{Not Unique}} (\mathrm{P}, \mathrm{P}')$ or $(\mathrm{PH}, \mathrm{P}'\mathrm{H})$

# Canonical Form of Camera Matrices

- The fundamental matrix corresponding to a pair of camera matrices $\mathrm{P} = [\mathrm{I} \mid \mathbf{0}]$ and $\mathrm{P}' = [\mathrm{M} \mid \mathbf{m}]$ is equal to $[\mathbf{m}]_{\times}\mathrm{M}$.

**Proof:**

$$\mathbf{e}' = \mathrm{P}'\mathrm{C} = [\, \mathrm{M} \mid \mathbf{m}\,][0,0,0,1]^{\mathsf{T}} = \mathbf{m}$$

$$\mathrm{F} = [\mathbf{e}']_{\times}\mathrm{P}'\mathrm{P}^{+} = [\mathbf{m}]_{\times}\,[\mathrm{M} \mid \mathbf{m}] \begin{bmatrix} \mathrm{I}_{3\times3} \\ 0_{3\times1} \end{bmatrix} = [\mathbf{m}]_{\times}\mathrm{M},$$

where $\mathrm{P}^{+} = \begin{bmatrix} \mathrm{I}_{3\times3} \\ 0_{3\times1} \end{bmatrix}$ since $\mathrm{PP}^{+} = \mathrm{I}$.

$\square$

# Projective Ambiguity of Cameras Given F

**Theorem:**

- Let $(\mathrm{P}, \mathrm{P}')$ and $(\tilde{\mathrm{P}}, \tilde{\mathrm{P}}')$ be two pairs of camera matrices such that $\mathrm{F}$ is the fundamental matrix corresponding to each of these pairs.

- Then, there exists a non-singular $4 \times 4$ matrix $\mathrm{H}$ such that $\tilde{\mathrm{P}} = \mathrm{PH}$ and $\tilde{\mathrm{P}}' = \mathrm{P}'\mathrm{H}$ .

# Projective Ambiguity of Cameras Given F

**Proof:**

- Suppose that a given fundamental matrix $\mathrm{F}$ corresponds to two different pairs of camera matrices $(\mathrm{P}, \mathrm{P}')$ and $(\tilde{\mathrm{P}}, \tilde{\mathrm{P}}')$.

- And the two pair of camera matrices is in canonical form with $\mathrm{P} = \tilde{\mathrm{P}} = [\mathrm{I} \mid \mathbf{0}]$, $\mathrm{P}' = [\mathrm{A} \mid \mathbf{a}]$ and $\tilde{\mathrm{P}}' = [\widetilde{\mathrm{A}} \mid \tilde{\mathbf{a}}]$.

- According to result of canonical cameras earlier, the fundamental matrix may then be written $\mathrm{F} = [\mathbf{a}]_\times \mathrm{A} = [\tilde{\mathbf{a}}]_\times \widetilde{\mathrm{A}}$.

# Projective Ambiguity of Cameras Given F

**Proof (cont.):**

- We will need the following lemma:

**Lemma:**

Suppose the rank 2 matrix $\mathrm{F}$ can be decomposed in two different ways as $\mathrm{F} = [\mathbf{a}]_\times \mathrm{A}$ and $\mathrm{F} = [\tilde{\mathbf{a}}]_\times \widetilde{\mathrm{A}}$;

then $\tilde{\mathbf{a}} = k\mathbf{a}$ and $\widetilde{\mathrm{A}} = k^{-1}(\mathrm{A} + \mathbf{a}\mathbf{v}^\top)$ for some non-zero constant $k$ and 3-vector $\mathbf{v}$.

# Projective Ambiguity of Cameras Given F

**(Lemma) Proof:**

First, note that $\mathbf{a}^\top \mathrm{F} = \mathbf{a}^\top [\mathbf{a}]_\times \mathrm{A} = \mathbf{0}$, and similarly, $\tilde{\mathbf{a}}^\top \mathrm{F} = \mathbf{0}$. Since $\mathrm{F}$ has rank 2, it follows that $\tilde{\mathbf{a}} = k\mathbf{a}$ as required.

Next, from $[\mathbf{a}]_\times \mathrm{A} = [\tilde{\mathbf{a}}]_\times \widetilde{\mathrm{A}}$, it follows that $[\mathbf{a}]_\times (k\widetilde{\mathrm{A}} - \mathrm{A}) = 0$, and so $k\widetilde{\mathrm{A}} - \mathrm{A} = \mathbf{a}\mathbf{v}^\top$ for some $\mathbf{v}$. Hence, $\widetilde{\mathrm{A}} = k^{-1}(\mathrm{A} + \mathbf{a}\mathbf{v}^\top)$ as required.

$\square$

# Projective Ambiguity of Cameras Given F

**Proof (cont.):**

- Applying this result to the two camera matrices $\mathrm{P}$ and $\tilde{\mathrm{P}}$ shows that $\mathrm{P}' = [\mathrm{A} \mid \mathbf{a}]$ and $\tilde{\mathrm{P}}' = [k^{-1}(\mathrm{A} + \mathbf{a}\mathbf{v}^\top) \mid k\mathbf{a}]$ if they are to generate the same F.

- Now let $\mathrm{H} = \begin{bmatrix} k^{-1}\mathrm{I} & \mathbf{0} \\ k^{-1}\mathbf{v}^\top & k \end{bmatrix}$, we then we can verify that $\mathrm{PH} = k^{-1}[\mathrm{I} \mid \mathbf{0}] = k^{-1}\tilde{\mathrm{P}}$.

# Projective Ambiguity of Cameras Given F

**Proof (cont.):**

- And furthermore,

$$\mathtt{P}'\mathtt{H} = [\mathtt{A} \mid \mathbf{a}]\mathtt{H} = [k^{-1}(\mathtt{A} + \mathbf{a}\mathbf{v}^\mathsf{T}) \mid k\mathbf{a}] = [\tilde{\mathtt{A}} \mid \tilde{\mathbf{a}}] = \tilde{\mathtt{P}}'$$

so that the pairs $\mathrm{P}$, $\mathrm{P}'$ and $\tilde{\mathrm{P}}$, $\tilde{\mathrm{P}}'$ are indeed projectively related.

$\square$

# Decomposition of F Matrix

- A non-zero matrix $\mathrm{F}$ is the fundamental matrix corresponding to a pair of camera matrices $\mathrm{P}$ and $\mathrm{P}'$ if and only if $\mathrm{P}'^{\top}\mathrm{F}\mathrm{P}$ is skew-symmetric.

**Proof:**

The condition that $\mathrm{P}'^{\top}\mathrm{F}\mathrm{P}$ is skew-symmetric is equivalent to $\mathbf{X}^{\top}\mathrm{P}'^{\top}\mathrm{F}\mathrm{P}\mathbf{X} = 0$ for all $\mathbf{X}$.

Setting $\mathbf{x}' = \mathrm{P}'\mathbf{X}$ and $\mathbf{x} = \mathrm{P}\mathbf{X}$, this is equivalent to $\mathbf{x}'^{\top}\mathrm{F}\mathbf{x} = 0$, which is the defining equation for the fundamental matrix.

$\square$

# Decomposition of F Matrix

- The camera matrices corresponding to a fundamental matrix F may be chosen as $\mathrm{P} = [\mathrm{I} \mid \mathbf{0}]$ and $\mathrm{P}' = [[\mathbf{e}']_\times \mathrm{F} \mid \mathbf{e}']$.

**Proof:**

We may verify that

$$[\mathrm{SF} \mid \mathrm{e}']^\mathsf{T} \mathrm{F} [\mathrm{I} \mid 0] = \begin{bmatrix} \mathrm{F}^\mathsf{T}\mathrm{S}^\mathsf{T}\mathrm{F} & \mathbf{0} \\ \mathrm{e}'^\mathsf{T}\mathrm{F} & 0 \end{bmatrix} = \begin{bmatrix} \mathrm{F}^\mathsf{T}\mathrm{S}^\mathsf{T}\mathrm{F} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 0 \end{bmatrix}, \text{ where } \mathrm{S} = [\mathbf{e}']_\times.$$

which is skew-symmetric and hence F is a valid fundamental matrix (as we have proven previously).

$\square$

# Decomposition of F Matrix

- According to the Lemma seen earlier:

  $\mathrm{F}$ can be decomposed in two different ways as $\mathrm{F} = [\mathbf{a}]_\times \mathrm{A}$ and $\mathrm{F} = [\tilde{\mathbf{a}}]_\times \widetilde{\mathrm{A}}$, where $\tilde{\mathbf{a}} = k\mathbf{a}$ and $\widetilde{\mathrm{A}} = k^{-1}(\mathrm{A} + \mathbf{a}\mathbf{v}^\top)$ for some non-zero constant $k$ and 3-vector **v**.

- The general formula for a pair of canonic camera matrices corresponding to a fundamental matrix F is given by:

$$\mathrm{P} = [\mathrm{I} \mid \mathbf{0}] \quad \mathrm{P}' = [[\mathbf{e}']_\times \mathrm{F} + \mathbf{e}'\mathbf{v}^\top \mid \lambda \mathbf{e}']$$

  where **v** is any 3-vector, and λ a non-zero scalar.

# Essential Matrix

- **Normalized coordinates:** Known calibration matrices $\mathrm{K}$ and $\mathrm{K}'$ $\Rightarrow$ we can write $\mathbf{x} \leftrightarrow \mathbf{x}'$ as $\mathrm{K}^{-1}\mathbf{x} \leftrightarrow \mathrm{K}'^{-1}\mathbf{x}'$, i.e. $\hat{\mathbf{x}} \longleftrightarrow \hat{\mathbf{x}}'$:

$$\mathbf{x}'^{\top}\boxed{\mathrm{F}}\mathbf{x} = \mathbf{x}'^{\top}\boxed{\mathrm{K}'^{-\top}\mathrm{E}\mathrm{K}^{-1}}\mathbf{x} = 0$$

$$\hat{\mathbf{x}}'^{\top}\boxed{\mathrm{E}}\hat{\mathbf{x}} = 0$$

$$\hat{\mathbf{x}}'^{\top}\boxed{[\mathbf{t}]_{\times}\mathrm{R}}\hat{\mathbf{x}} = 0$$

- E is the Essential Matrix which can be expressed in terms of the relative transformation between two image frames.

# Essential Matrix

**Proof:**

Previously we seen $\mathrm{F} = [\mathbf{e}']_\times \mathrm{P}'\mathrm{P}^+$, since $\mathrm{P} = \mathrm{K}[\mathrm{I} \mid 0]$ and $\mathrm{P}' = \mathrm{K}'[\mathrm{R} \mid \mathbf{t}]$, we have:

$$\mathrm{P}^+ = \begin{bmatrix} \mathrm{K}^{-1} \\ 0_{1\times 3} \end{bmatrix}, \qquad \mathbf{C} = \begin{bmatrix} 0_{3\times 1} \\ 1 \end{bmatrix}$$

and

$$\mathrm{F} = [\mathbf{e}']_\times \mathrm{P}'\mathrm{P}^+ = [\mathrm{P}'\mathbf{C}]_\times \mathrm{P}'\mathrm{P}^+$$

$$= [\mathrm{K}'\mathbf{t}]_\times \mathrm{K}'\mathrm{R}\mathrm{K}^{-1} = \mathrm{K}'^{-\mathsf{T}} [\mathrm{t}]_\times \mathrm{R}\mathrm{K}^{-1}$$

# Properties of the Essential Matrix

- Five degree of freedom (3+3-1):
  - R and **t** have 3 degree of freedom each
  - But there is an overall scale ambiguity ⇒ -1 dof
- Singular values:
  - A 3 x 3 matrix is an essential matrix iff two of its singular values are equal, and the third is zero

# Decomposition of E Matrix

- Extract R and t from the essential matrix E.

$$\mathrm{E} = \lfloor \mathbf{t} \rfloor_\times \mathrm{R}$$

Let us factorize $\lfloor \mathbf{t} \rfloor_\times$ and R into:

$$\mathrm{E} = \lfloor \mathrm{t} \rfloor_\times \mathrm{R} = \underbrace{(\mathrm{UZU}^\top)}_{\text{Skew-symmetric matrix}}\underbrace{(\mathrm{UXV}^\top)}_{\text{Some rotation matrix}} = \underbrace{\mathrm{U(ZX)V}^\top}_{\text{SVD of E}}$$

Since E is known up to a scale and ignoring the sign, we can set:

$$\mathrm{ZX} = \begin{cases} \mathrm{ZW} = \mathrm{diag}(1,1,0) \\ \mathrm{ZW}^\top = \mathrm{diag}(-1,-1,0) \end{cases}, \quad \text{where} \quad \mathrm{Z} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \mathrm{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

# Decomposition of E Matrix

- U and V are known from SVD of E.

**Recovery of t:** $[\mathbf{t}]_\times = \mathrm{UZU}^\top$

Since U is orthogonal and $[\mathbf{t}]_\times$ is skew-symmetric, we get:

$$\mathbf{t} = \pm\mathbf{U}_3, \quad \text{i.e. third column of U}$$

**Recovery of R:** $\mathrm{R} = \mathrm{UWV}^\top$, or $\mathrm{R} = \mathrm{UW}^\top\mathrm{V}^\top$

Make sure that R is in the Right-Hand Coordinate:

If det(R) < 0, then R = -R.

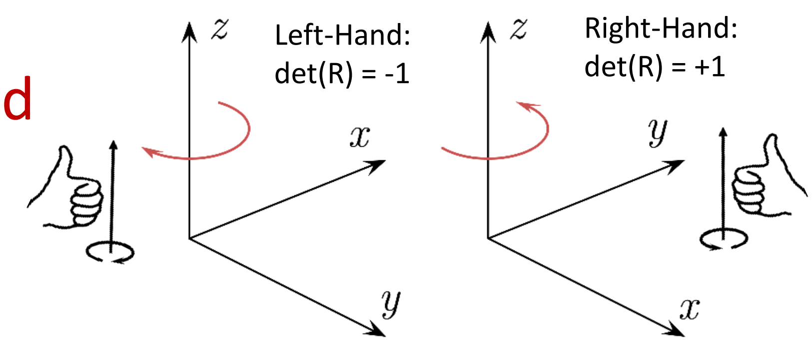
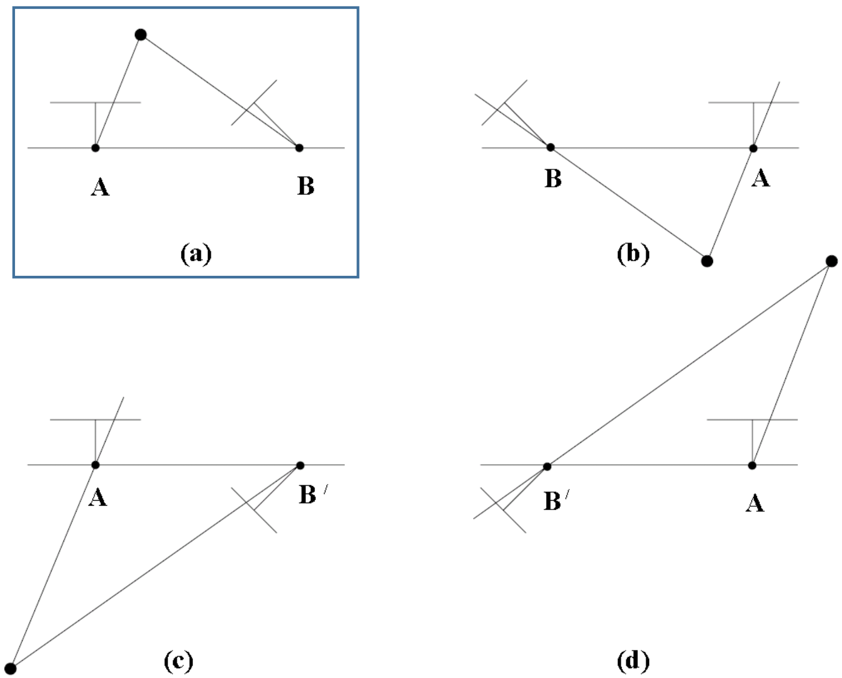



Image Source: https://en.wikipedia.org/wiki/Right-hand_rule

# Decomposition of E Matrix

## Four Possible Solutions for P':

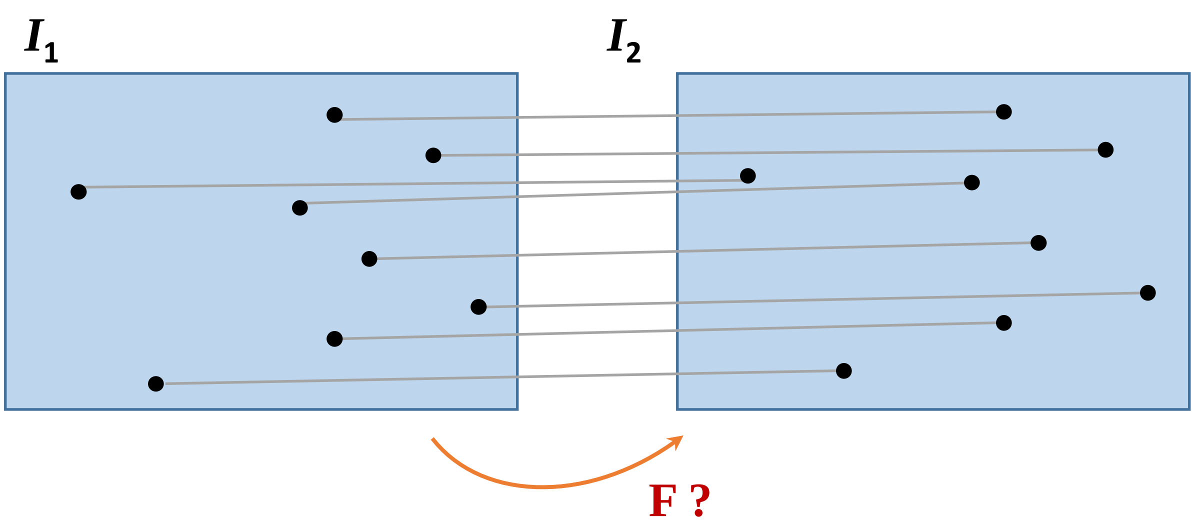
$$P' = [UWV^T \mid +\mathbf{u}_3] \text{ or } [UWV^T \mid -\mathbf{u}_3] \text{ or } [UW^TV^T \mid +\mathbf{u}_3] \text{ or } [UW^TV^T \mid -\mathbf{u}_3]$$




Only 1 of the 4 solutions is physically correct, i.e. the 3D point appears in front of both cameras.

# Linear 8-Point Algorithm for F Matrix

- **Given**: A set of points correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}_i'$ between two images.
- **Compute**: The Fundamental matrix F.



# Linear 8-Point Algorithm for F Matrix

- For any pair of matching points $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ in two images, the 3x3 fundamental matrix is defined by the equation:

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

- Let $\mathbf{x} = (x, y, 1)^{\mathsf{T}}$ and $\mathbf{x}' = (x', y', 1)^{\mathsf{T}}$, we rewrite the above equation as:

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

- Let **f** be the 9-vector made up of the entries of F in row-major order, we get:

$$(x'x, x'y, x', y'x, y'y, y', x, y, 1)\,\mathbf{f} = 0.$$

# Linear 8-Point Algorithm for F Matrix

- From a set of *n* point matches, we obtain a set of linear equations of the form:

$$\mathbf{Af} = \begin{bmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x'_n x_n & x'_n y_n & x'_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{bmatrix} \mathbf{f} = \mathbf{0}$$

- A is a n x 9 matrix.
- For a non-trivial solution to exist, rank(A)=8 since f is a 9-vector.
- A minimum of 8-point correspondences is needed to solve for f.

# Linear 8-Point Algorithm for F Matrix

- For noisy data, we obtain the solution of f by finding the least-squares solution.
- Least-squares solution for **f** is the singular vector corresponding to the smallest singular value of A.
- That is the last column of V in the SVD $A = UDV^T$.
- Similar to homography estimation, data normalization is needed.

# Singularity Constraint of F Matrix

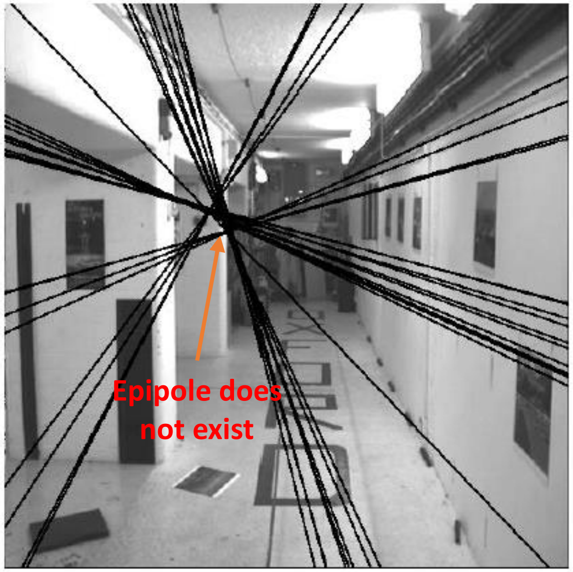
- An important property of the fundamental matrix is that it is singular, i.e. rank(F) = 2.

- **Problem:** Least-squares solution in general will NOT give rank(F)=2.

**Recall:**

- Right and left nullspaces of F gives the epipoles, i.e. $\mathrm{F}\mathbf{e} = 0$, and $\mathrm{F}^\top \mathbf{e}' = 0$.
- Since e is a 3-vector, epipole exists if rank(F)=2.

$\mathrm{rank}(\mathrm{F}) \neq 2$




$\mathrm{rank}(\mathrm{F}) = 2$

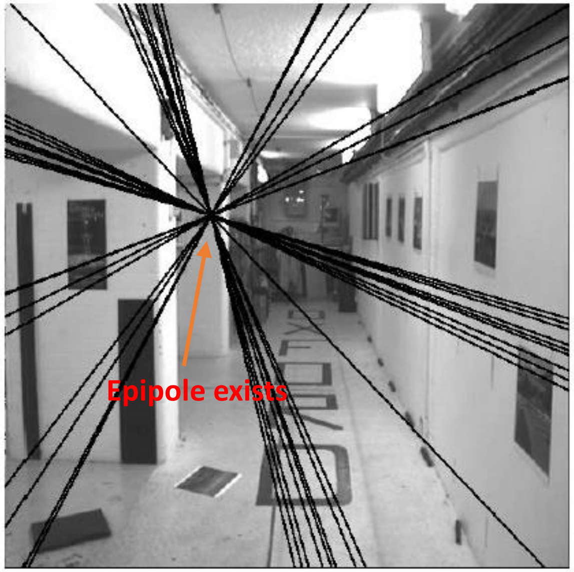



Image Source: R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# Singularity Constraint of F Matrix

- Most convenient way is to correct the matrix F found by the SVD solution from A.

- Matrix F is replaced by the matrix F' that minimizes the Frobenius norm:

$$\min_{\mathrm{F}'} \|\mathrm{F} - \mathrm{F}'\|, \quad \text{s.t.} \qquad \det(\mathrm{F}') = 0$$

**Steps:**

1. Take SVD of F, i.e. $\mathrm{F} = \mathrm{UDV}^\top$, where $\mathrm{D} = \mathrm{diag}(r, s, t)$ satisfying $r \geq s \geq t$.
2. Then, $\mathrm{F}' = \mathrm{U}\mathrm{diag}(r, s, 0)\mathrm{V}^\top$ minimizes the Frobenius norm of $\mathrm{F} - \mathrm{F}'$.

# Normalized 8-Point Algorithm for F Matrix

<u>Objective</u>

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

<u>Algorithm</u>

**(i)** **Normalization:** Transform the image coordinates according to $\hat{\mathbf{x}}_i = \mathrm{T}\mathbf{x}_i$ and $\hat{\mathbf{x}}_i' = \mathrm{T}'\mathbf{x}_i'$, where $\mathrm{T}$ and $\mathrm{T}'$ are normalizing transformations.

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

- **a)** **Linear 8-point algorithm**.
- **b)** **Enforcing singularity constraint**.

Note: RANSAC should be used for robust estimation!

(iii) Denormalization: Set $\mathrm{F} = \mathrm{T}'^{\top} \hat{\mathrm{F}}' \mathrm{T}$.

$$\mathrm{T} = \begin{bmatrix} s & 0 & -sc_x \\ 0 & s & -sc_y \\ 0 & 0 & 1 \end{bmatrix}$$

c: centroid of all data points

$$s = \frac{\sqrt{2}}{\bar{d}}$$

where $\bar{d}$ : mean distance of all points from centroid.

# Normalized 8-Point Algorithm for E Matrix

<u>Objective</u>

Given $n \geq 8$ image point correspondences $\{x_i \leftrightarrow x_i'\}$ and the camera calibration matrices K and K', compute the essential matrix E such that $\mathbf{x}_i'^{\top}\, \mathrm{K}'^{-\top} \mathrm{E}\, \mathrm{K}^{-1} \mathbf{x}_i = 0$.

<u>Algorithm</u>

**(i)** **Normalized Coordinates:** For each correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}_i'$, compute $\mathrm{K}^{-1}\mathbf{x}_i \leftrightarrow \mathrm{K}'^{-1}\mathbf{x}_i'$, i.e. $\hat{\mathbf{x}} \leftrightarrow \hat{\mathbf{x}}'$.

(ii) Find the essential matrix $\mathrm{E}$ corresponding to the matches $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}_i'$ by:

- **a)** **Linear 8-point algorithm**.
- **b)** ***Enforcing singularity constraint**.

(iii) Decompose E to get R and **t**, thus forming P and P'.

*Singular constraint for E matrix is different from F matrix. See next slide for more detail.

# Singularity Constraint of E Matrix

**Problem:**

- In general, the essential matrix E obtained from the linear 8-point algorithm will NOT have two similar singular values, and third is zero.

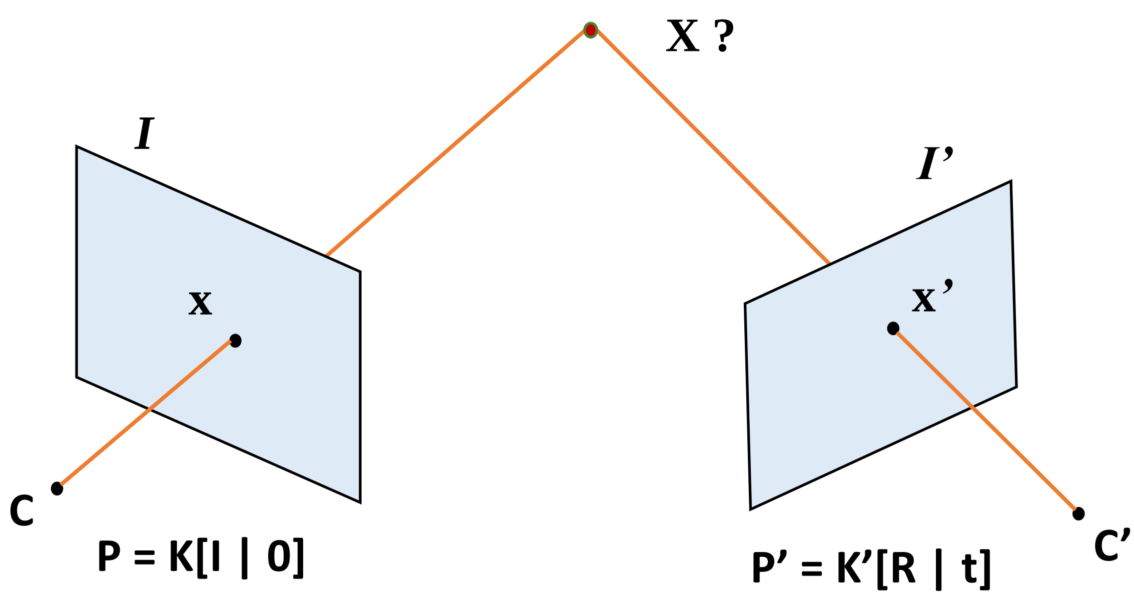
**Solution:**

1. Take SVD of E, i.e. $\mathrm{E} = \mathrm{UDV}^\top$, where $\mathrm{D} = \mathrm{diag}(a, b, c)$ with $a \geq b \geq c$.
2. The closest essential matrix to E in Frobenius norm is given $\hat{\mathrm{E}} = \mathrm{U}\hat{\mathrm{D}}\mathrm{V}^\top$, where

$$\hat{\mathrm{D}} = \mathrm{diag}(\frac{a+b}{2}, \frac{a+b}{2}, 0)$$

# 3D Structure Computation

- **Given:** The point correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ and camera projection matrices P and P' of two images.
- **Find:** The 3D structure points $\mathbf{X}_i$ that corresponds to each 2D point correspondence.



# 3D Structure Computation

## Linear Triangulation Method

- In each image, we have a measurement:

$$\mathbf{x} = \mathrm{P}\mathbf{X},\ \mathbf{x}' = \mathrm{P}'\mathbf{X}$$

- Unknown scale factor is eliminated by a cross-product, i.e. $\mathbf{x} \times (\mathrm{P}\mathbf{X}) = 0$ to give:

$$\begin{aligned} x(\mathbf{p}^{3\top}\mathbf{X}) - (\mathbf{p}^{1\top}\mathbf{X}) &= 0 \\ y(\mathbf{p}^{3\top}\mathbf{X}) - (\mathbf{p}^{2\top}\mathbf{X}) &= 0 \\ x(\mathbf{p}^{2\top}\mathbf{X}) - y(\mathbf{p}^{1\top}\mathbf{X}) &= 0 \end{aligned}$$

- $\mathrm{P}^{i\top}$ are rows of P.
- Two of the three equations are linearly independent.

# 3D Structure Computation

## Linear Triangulation Method

- An equation of the form $A\mathbf{X} = 0$ can be formed:

$$\mathtt{A} = \begin{bmatrix} x\mathbf{p}^{3\top} - \mathbf{p}^{1\top} \\ y\mathbf{p}^{3\top} - \mathbf{p}^{2\top} \\ x'\mathbf{p}'^{3\top} - \mathbf{p}'^{1\top} \\ y'\mathbf{p}'^{3\top} - \mathbf{p}'^{2\top} \end{bmatrix}$$

- Two equations from each image, giving a total of four equations in four homogeneous unknowns, i.e. $\mathbf{X} = [X\ Y\ Z\ 1]^{\top}$.

- Solution given by the right singular vector that corresponds to the smallest singular value of $A$, i.e. $\boldsymbol{v}_4$.

- $\mathbf{X} = \boldsymbol{v}_4 / v_{4w} \Rightarrow$ to make last element of $\mathbf{X}$ equal to 1.

# Reconstruction (Similarity) Ambiguity

- **Known Calibration:** Scene determined by the image is only up to a similarity transformation (rotation, translation and scaling).

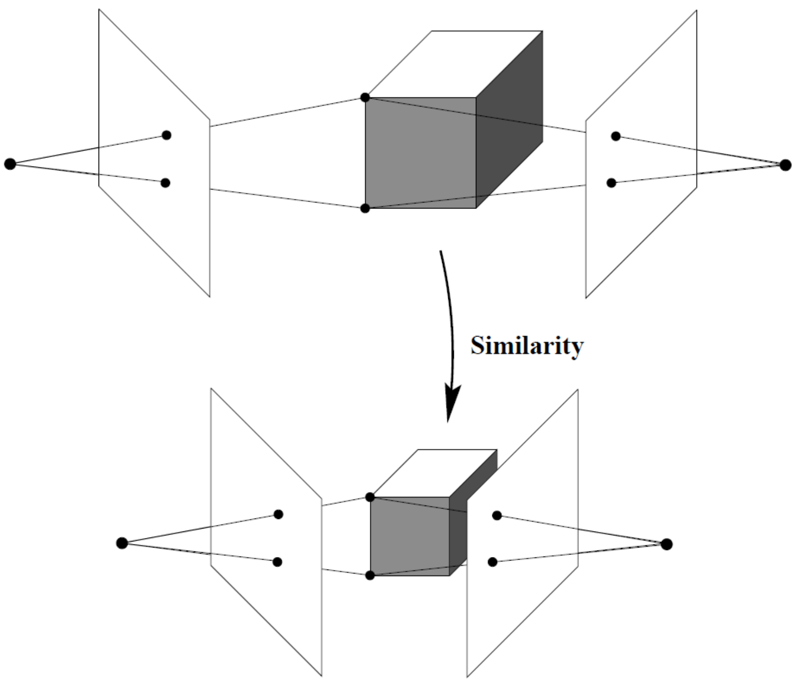



Image Source: R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

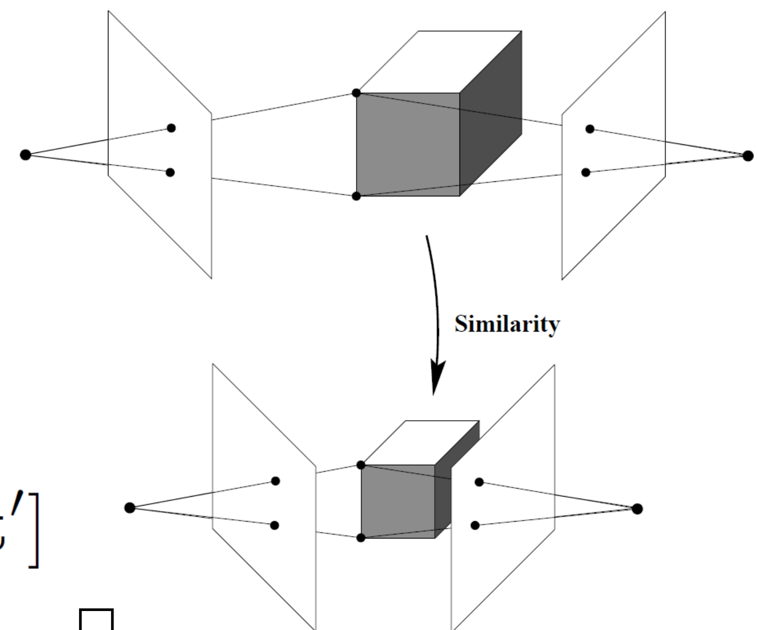
# Reconstruction (Similarity) Ambiguity

**Proof sketch:**

Let $\mathrm{H}_S$ be any similarity transformation: $\mathrm{H}_\mathrm{S} = \begin{bmatrix} \mathrm{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & \lambda \end{bmatrix}$.

We can see that the projection on $\mathrm{X}_i$ is the same under $\mathrm{P}$ and $\mathrm{PH}_S^{-1}$:

$$\mathrm{P}\mathbf{X}_i = (\mathrm{PH}_\mathrm{S}^{-1})(\mathrm{H}_\mathrm{S}\mathbf{X}_i)$$




And $\mathrm{PH}_S^{-1}$ is still a valid projection matrix:

$$\mathrm{P} = \mathrm{K}[\mathrm{R}_\mathrm{P} \mid \mathbf{t}_\mathrm{P}], \ \mathrm{PH}_\mathrm{S}^{-1} = \mathrm{K}[\mathrm{R}_\mathrm{P}\mathrm{R}^{-1} \mid \mathbf{t}']$$

$\square$

Image Source: R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# Reconstruction (Projective) Ambiguity

- **Unknown Calibration:** We saw earlier that the fundamental matrix can be decomposed into $\mathrm{P}$ and $\mathrm{P}'$ or $\mathrm{PH}^{-1}$ and $\mathrm{P}'\mathrm{H}^{-1}$.

- The point $\mathbf{X}$ is reconstructed as $\mathrm{H}\mathbf{X}$ under $\mathrm{PH}^{-1}$ and $\mathrm{P}'\mathrm{H}^{-1}$ since:

$$\mathrm{P}\mathbf{X} = \mathrm{PH}^{-1}\mathrm{H}\mathbf{X}.$$

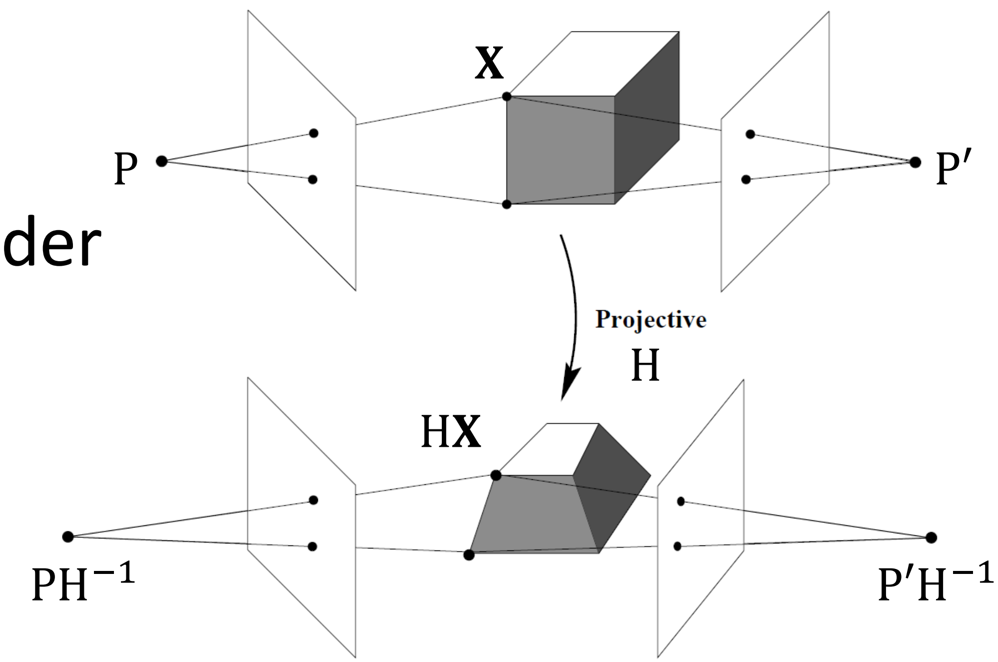



Image Source: R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision"

# Reconstruction (Projective) Ambiguity

- Original image pair


- Two different views of the reconstruction by $\mathrm{P}$ and $\mathrm{P}'$ decomposed from the $\mathrm{F}$ matrix obtained with the image pair.

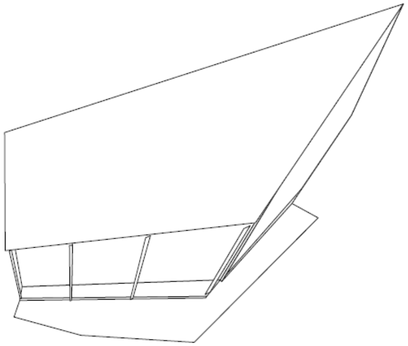
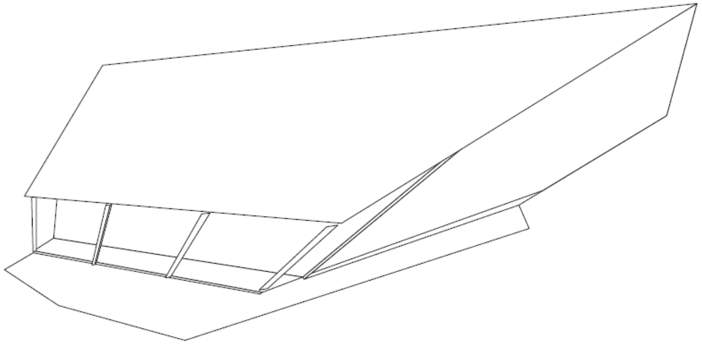



Image Source: R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# Stratified Reconstruction

- The "stratified" approach to reconstruction:

1. Begin with a projective reconstruction.
2. And then refine it progressively to an affine.
3. Finally a metric reconstruction.

- We will see that affine and metric reconstruction are not possible without further information either about the scene, the motion or the camera calibration.

# The Step to Affine Reconstruction

- The essence of affine reconstruction is to locate the plane at infinity.

- Let the 4-vector $\boldsymbol{\pi}$ be the plane at infinity under projective distortion; the goal is to find the projective transformation $\mathrm{H}$ that maps $\boldsymbol{\pi}$ to $(0,0,0,1)^\top$, i.e. $\boldsymbol{\pi}_\infty = \mathrm{H}^{-\top}\boldsymbol{\pi}$.

- H can be easily obtained as: $\mathrm{H} = \begin{bmatrix} \mathrm{I} \mid \mathbf{0} \\ \boldsymbol{\pi}^\top \end{bmatrix}.$

- Map all 3D reconstruction points $\mathrm{X}$ using $\mathrm{H}$ to remove the projective distortion (get an affine reconstruction).

# The Step to Affine Reconstruction

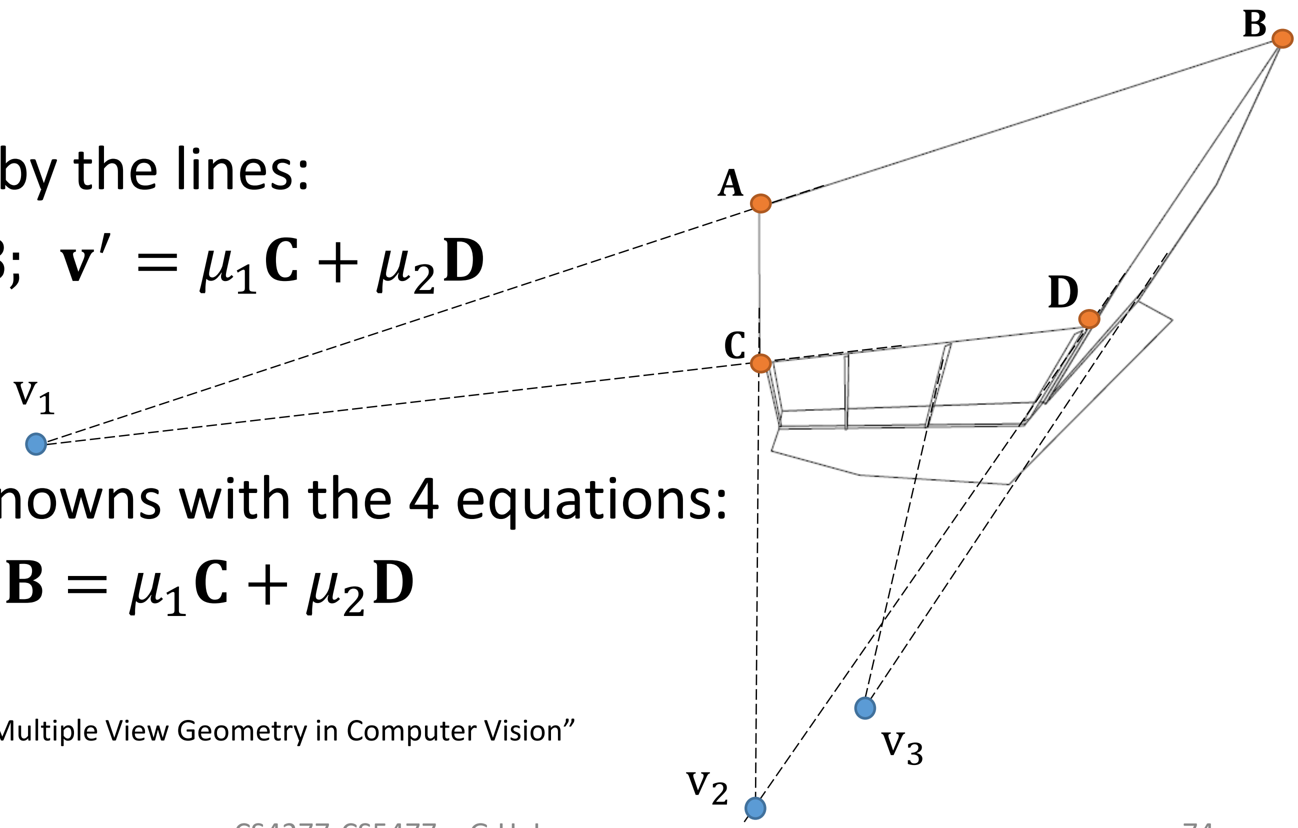
- Let $\mathrm{v}_1, \mathrm{v}_2, \mathrm{v}_3$ be the intersection points of a pair of parallel lines in three different directions, i.e. vanishing points.

- $\boldsymbol{\pi}$ can be identified from: $[\mathrm{v}_1 \quad \mathrm{v}_2 \quad \mathrm{v}_3]^\top \boldsymbol{\pi} = \mathbf{0}$.

**Example:**

Points on span by the lines:

$$\mathbf{v} = \lambda_1 \mathbf{A} + \lambda_2 \mathbf{B}; \;\; \mathbf{v}' = \mu_1 \mathbf{C} + \mu_2 \mathbf{D}$$



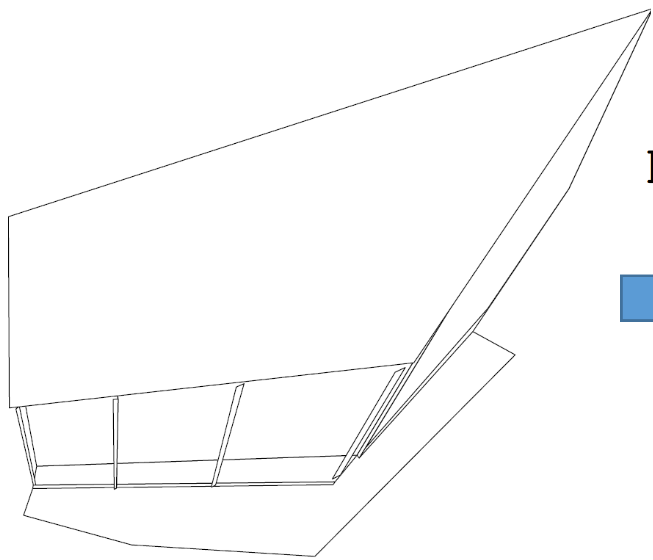

Solve for 4 unknowns with the 4 equations:

$$\mathbf{v}_1 = \lambda_1 \mathbf{A} + \lambda_2 \mathbf{B} = \mu_1 \mathbf{C} + \mu_2 \mathbf{D}$$

Image Source:
R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# The Step to Affine Reconstruction

Projective distortion



$$\mathrm{H} = \begin{bmatrix} \mathrm{I} \mid \mathbf{0} \\ \boldsymbol{\pi}^{\mathsf{T}} \end{bmatrix}$$

Affine distortion

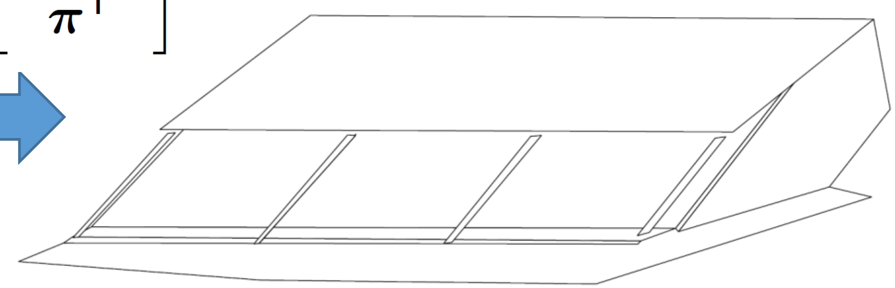


Image Source: R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# The Step to Metric Reconstruction

- The key to metric reconstruction is the identification of the image of absolute conic $\omega$ (IAC).

- The affine reconstruction may be transformed to a metric reconstruction by applying a 3D transformation of the form:

$$\mathtt{H} = \begin{bmatrix} \mathtt{A}^{-1} & \\ & 1 \end{bmatrix},$$

where

- A is obtained by Cholesky factorization of $\mathtt{A}\mathtt{A}^{\mathsf{T}} = (\mathtt{M}^{\mathsf{T}}\boldsymbol{\omega}\mathtt{M})^{-1}$.
- The affine reconstruction is from the camera matrix $\mathrm{P}' = [\mathrm{M} \mid \mathrm{m}]$.

# The Step to Metric Reconstruction

**Proof:**

- We have seen earlier that under known calibration $K'$, the camera matrix $P'_M = K'[R \mid \mathbf{t}]$ is subjected to similarity distortion.

- The affinely distorted camera matrix $P' = [M \mid m]$ is transformed to $P'_M$ as $P'_M = P'H^{-1}$, where

$$H^{-1} = \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{bmatrix} \quad \Rightarrow \quad [\, K'R \mid K'\mathbf{t}\,] = [MA \mid m]$$

# The Step to Metric Reconstruction

**Proof (cont.):**

- Hence, we get $\mathrm{MA} = \mathrm{K'R}$, which can be written as:

$$\mathrm{MA(MA)^\top = K'R(K'R)^\top} \;\Rightarrow\; \mathrm{MAA^\top M^\top = K'K'^\top}$$

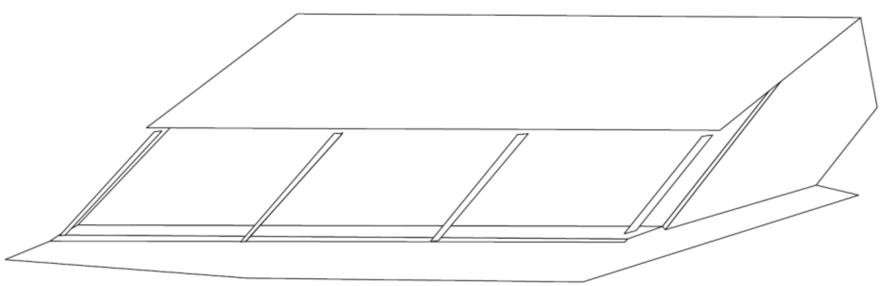
$$\Rightarrow\; \mathrm{AA^\top = M^{-1}\underbrace{K'K'^\top}_{\boldsymbol{\omega}^* = \boldsymbol{\omega}^{-1}}M^{-\top}}$$

$$\Rightarrow\; \mathrm{AA^\top = (M^\top \boldsymbol{\omega} M)^{-1}}. \qquad \square$$

- Refer to Lecture 5 for the various methods to get the Image of absolute conic $\boldsymbol{\omega}$ (IAC).

# The Step to Metric Reconstruction

Affine distortion



$$\mathtt{H} = \begin{bmatrix} \mathtt{A}^{-1} & \\ & 1 \end{bmatrix}$$

Similarity distortion

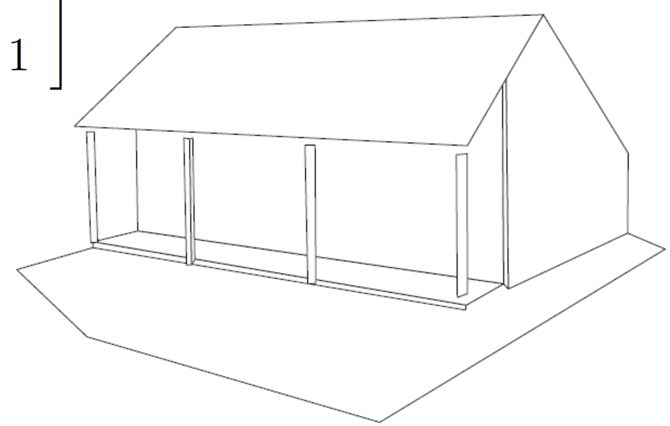


Image Source: R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”

# Summary

- We have looked at how to:

1. Describe the epipolar geometry between two views.
2. Estimate fundamental / essential matrix with 8 point correspondences.
3. Decompose fundamental matrix into the camera matrices of two views.
4. Find rotation and translation between two views from the essential matrix.
5. Recover 3D structures with linear triangulation, and do stratified reconstruction from uncalibrated reconstruction.