Two and Three View Geometry

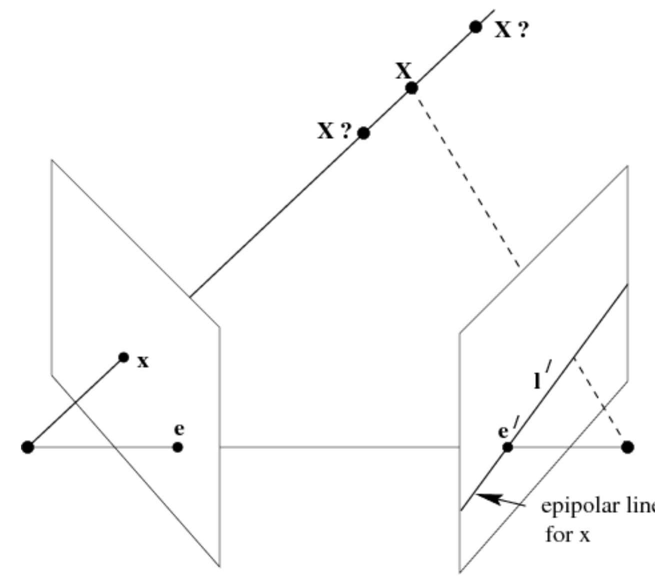
P J Narayanan

CVIT, IIIT, Hyderabad

CS5765. Computer Vision. Spring 2013

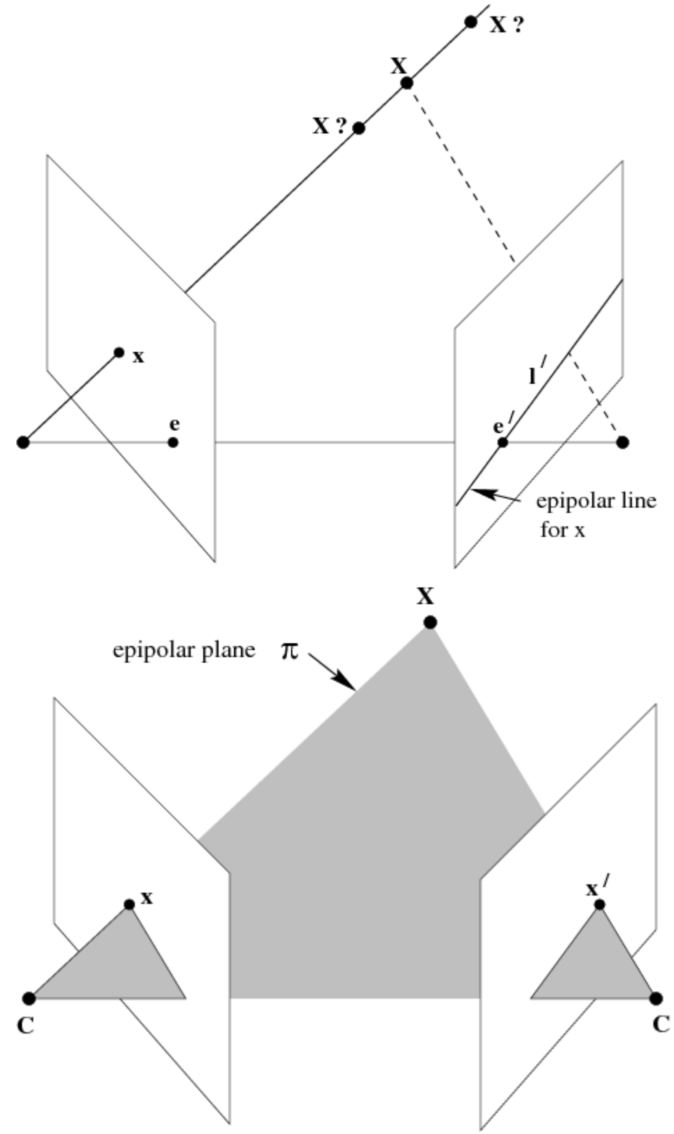
An Image Point in Another View

- Given an image point $\mathbf{x}$ of a world point X in one camera, where does the corresponding point lie in another view?
- The ray corresponding to $\mathbf{x}$ appears as a line



Epipolar Geometry

- Given two cameras and an image $\mathbf{x}$ of a world point X in one, where does the image in the other lie? On a line called the **epipolar line**.
- **Epipole** is the image of one camera on the other. Also the vanishing point of the baseline between cameras.
- **Epipolar plane** contains the baseline and the ray due to $\mathbf{x}$.
- Restricts the *corresponding* point to a line in other view.



Fundamental Matrix $\mathbf{F}$

- Let $\mathbf{P_1} = \mathbf{K}[\mathbf{I}|\mathbf{0}]$ and $\mathbf{P_2} = \mathbf{K}'[\mathbf{R}|\mathbf{t}]$ be the two cameras. A point x in image 1 is mapped to a line l' in image 2.
- x maps to direction $\mathbf{K}^{\text{-}1}\mathbf{x}$ and point at infinity is on it.
- It projects to $\mathbf{K}'\mathbf{R}\mathbf{K}^{\text{-}1}\mathbf{x}$ in the second image.
- Epipolar line l' joins this point to epipole e'.
- $\mathbf{l}' = \mathbf{e}' \times (\mathbf{K}'\mathbf{R}\mathbf{K}^{\text{-}1}\mathbf{x}) = [\mathbf{e}']_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{\text{-}1}\mathbf{x} = \mathbf{F}\mathbf{x}$, where $\mathbf{F} = [\mathbf{e}']_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{\text{-}1}$ is called the **Fundamental Matrix**. For general cameras: $\mathbf{F} = [\mathbf{e}']_\times \mathbf{P}'\mathbf{P}^+$
- If $\mathbf{x}, \mathbf{x}'$ are corresponding points, $\mathbf{x}'^{\mathbf{T}}\mathbf{F}\mathbf{x} = \mathbf{0}$.

$[\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$. $\mathbf{t} \times \mathbf{v} = [\mathbf{t}]_\times \mathbf{v}$. $[\mathbf{t}]_\times$ is skew-symmetric, singular.

Alternate Derivation

- Assume a plane π not passing through C, C', with a homography $\mathbf{H}_\pi$.
- For a point on π that projects to x, x' in the two images, we have $\mathbf{x}' = \mathbf{H}_\pi \mathbf{x}$.
- x' is a point on the epipolar line of x. So is e'.
- Epipolar line $\mathbf{l}' = [\mathbf{e}']_\times \mathbf{H}_\pi \mathbf{x} = \mathbf{F}\mathbf{x}$, where $\mathbf{F} = [\mathbf{e}']_\times \mathbf{H}_\pi$.
- If the plane is π_∞, we have $\mathbf{x} = \mathbf{K}\mathbf{d}$ and $\mathbf{x}' = \mathbf{K}'\mathbf{R}\mathbf{d}$. $\mathbf{x}' = \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}\mathbf{x}$, or $\mathbf{H}_\infty = \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}$.
- Thus, $\mathbf{F} = [\mathbf{e}']_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{-1}$, as before.
- Any planar homography $\mathbf{H}$ will work, but given 2 cameras, the fundamental matrix $\mathbf{F}$ is fixed!!

Properties of $\mathbf{F}$

- $\mathbf{F}^\mathbf{T}$ is the fundamental matrix from $\mathbf{I}'$ to $\mathbf{I}$.
  $\mathbf{l}' = \mathbf{F}\mathbf{x}$ and $\mathbf{l} = \mathbf{F}^\mathbf{T}\mathbf{x}'$.
- For epipolar line pairs, $\mathbf{l}' = \mathbf{F}[\mathbf{e}]_\times \mathbf{l}$ and $\mathbf{l} = \mathbf{F}^\mathbf{T}[\mathbf{e}']_\times \mathbf{l}'$.
  $\mathbf{e}$ is a line not passing through the epipole. For an epipolar line $\mathbf{l}$, $[\mathbf{e}]_\times \mathbf{l}$ is its point of intersection with $\mathbf{e}$. $\mathbf{F}[\mathbf{e}]_\times \mathbf{l}$ is its epipolar line, namely, $\mathbf{l}'$. Thus $\mathbf{l}' = \mathbf{F}[\mathbf{e}]_\times \mathbf{l}$
- Epipole $\mathbf{e}'$ is on $\mathbf{l}'$. Hence, $\mathbf{e}'^\mathbf{T}\mathbf{F}\mathbf{x} = (\mathbf{e}'^\mathbf{T}\mathbf{F})\mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$. Thus, $\mathbf{e}'^\mathbf{T}\mathbf{F} = \mathbf{0}$. Similarly, $\mathbf{F}\mathbf{e} = \mathbf{0}$. Epipoles are the left and right null spaces of $\mathbf{F}$.
- $\mathbf{F}$ has rank 2 (as $[\mathbf{e}]_\times$ has rank 2).
- $\mathbf{F}$ is a property of the camera pair only and is independent of the scene.
- It has 7 degrees of freedom; overall defined upto scale, with a 0 determinant. The fundamental matrix can be computed from 7 point correspondences.

Restricted Situations

- Pure translation: $\mathbf{F} = [\mathbf{e}']_\times$ with 2 dof. $\mathbf{K}' = \mathbf{K}, \mathbf{R} = \mathbf{I}$
- Translation in X only: $\mathbf{P}' = \mathbf{K}[\mathbf{I}| -\mathbf{C}]$, where $\mathbf{C} = [a \;\; 0 \;\; 0]^{\mathbf{T}}$. Thus, $\mathbf{e}' = -\mathbf{KC} = [-fa \;\; 0 \;\; 0]^{\mathbf{T}}$ and

$$\mathbf{F} = [\mathbf{e}']_\times = -\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -fa \\ 0 & fa & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

  $\mathbf{x}'^{\mathbf{T}}\mathbf{F}\mathbf{x} = \mathbf{0}$ reduces to $y' = y$.
- Pure rotation or No translation: No epipolar geometry as $\mathbf{x}' = \mathbf{H}\mathbf{x}$ for all $\mathbf{x}$.
- General motion: $\mathbf{F} = [\mathbf{e}']_\times \mathbf{H}_\infty = [\mathbf{e}']_\times \mathbf{K}'\mathbf{R}\mathbf{K}^{\textbf{-1}}$.

Three Cases of Translation

In X only:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$x'^{\mathbf{T}} F x = 0$
reduces to:

$$v = v'$$

Horizontal
epipolar lines

In Y only:

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

$x'^{\mathbf{T}} F x = 0$
reduces to:

$$u = u'$$

Vertical
epipolar lines

In Z only:

$$\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$x'^{\mathbf{T}} F x = 0$
reduces to:

$$\frac{u}{v} = \frac{u'}{v'}$$

Radial
epipolar lines

Fundamental Matrix Estimation

Estimating F

- Each point match gives one equation $\mathbf{x'^T F x = 0}$ in the unknown entries of $\mathbf{F}$.
- $\mathbf{F}$ has 8 degrees of freedom upto scale factor. Hence the **8-point algorithm** can estimate it.
- $\mathbf{F}$ is singular and $|\mathbf{F}| = \mathbf{0}$. Thus, really only 7 degrees of freedom and 7 points should suffice. However, it is hard to enforce this constraint. Reduces to a cubic polynomial and non-linear optimization required.
- In practice, solve using DLT with $n > 8$ matches.
    - Normalize each image. Centroid is origin, mean distance is $\sqrt{2}$.
    - $\mathbf{A}$ is an $n \times 9$ matrix from equations for each match.
    - SVD of $\mathbf{AA^T = UDV^T}$. Solution: last column of $\mathbf{V}$.

Enforcing Rank-2 Constraint

- Due to noise, $\mathbf{F}$ may have a full rank of 3. Epipolar lines may not meet if the fundamental matrix is not of rank 2
- To enforce singularity, find $\mathbf{F} = \mathbf{UDV^T}$. Modify $\mathbf{D} =$ *diag*$(\mathbf{a}, \mathbf{b}, \mathbf{c})$ to $\mathbf{D}' =$ *diag*$(\mathbf{a}, \mathbf{b}, \mathbf{0})$ by setting last singular value to 0.
- Use $\mathbf{F}' = \mathbf{UD'V^T}$ as the fundamental matrix. This minimizes the Frobenius norm between original $\mathbf{F}$ and $\mathbf{F}'$ with rank of 2.
- Overall procedure to estimate the fundamental matrix:
  - Normalize points in each image independently
  - Find $\mathbf{F}$ from SVD of $\mathbf{AA^T}$ using the DLT algorithm
  - Enforce rank 2 constraint
  - Denormalize values to get $\mathbf{F}$ in original coordinates

RANSAC to Estimate F

- Find interest points in each image
- Find matches using similarity, proximity, etc
- Estimate $\mathbf{F}$ using RANSAC, using N samples
  - Select a random sample of 8 points, estimate $\mathbf{F}$
  - Compute number of inliers among matches to $\mathbf{F}$
  - Keep the solution with maximum number of inliers
- Estimate $\mathbf{F}$ from all matches classified as inliers. Can use a more accurate, non-linear optimization of geometric distance for this step
- Use $\mathbf{F}$ matrix to find better interest point matches
- Repeat last 2 steps until matches converge

Projective Properties of Fundamental Matrix

Projective Ambiguity of $\mathbf{F}$

- $\mathbf{x}'^{\mathbf{T}}\mathbf{F}\mathbf{x} = \mathbf{0}$ for a point $\mathbf{X}$ for camera pair $(\mathbf{P}, \mathbf{P}')$. Consider a camera pair $(\mathbf{PH}, \mathbf{P}'\mathbf{H})$, where $\mathbf{H}$ is a general, non-singular, 4×4 homography in 3-space.
- Point $\mathbf{H}^{-1}\mathbf{X}$ projects to $(\mathbf{PH})\ (\mathbf{H}^{-1}\mathbf{X})$ and $(\mathbf{P}'\mathbf{H})\ (\mathbf{H}^{-1}\mathbf{X})$. $\mathbf{F}$ is the fundamental matrix of $(\mathbf{PH}, \mathbf{P}'\mathbf{H})$, for any $\mathbf{H}$

  Any property based on projections $\mathbf{x}$ will have a projective ambiguity
- Theorem: **That is the only ambiguity**! If two camera configurations have the same fundamental matrix, there is a homography that relate the cameras.
- This suggests we can reduce every camera pair to their canonical form of $[\mathbf{I}|\mathbf{0}]$ and $[\mathbf{M}|\mathbf{m}]$. The fundamental matrix for this is $[\mathbf{m}]_\times \mathbf{M}$.

  Let $\mathbf{P}^*$ be a non-singular matrix built by adding a row to $\mathbf{P}$.

  Set $\mathbf{H} = (\mathbf{P}^*)^{-1}$. $\mathbf{PH} = [\mathbf{I}|\mathbf{0}]$ and $\mathbf{P}'\mathbf{H} = [\mathbf{M}|\mathbf{m}]$

Decomposing $\mathbf{F}$ to Cameras

- Same $\mathbf{F}$ can be given by several camera pairs. We can decompose a known $\mathbf{F}$ into camera pairs.
- The cameras $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ and $\mathbf{P}' = [[\mathbf{e}']_\times \mathbf{F}|\mathbf{e}']$ have the fundamental matrix $\mathbf{F}$.
  $\mathbf{F} = [\mathbf{e}']_\times [\mathbf{e}']_\times \mathbf{F} = (\mathbf{e}'\mathbf{e}'^{\mathbf{T}} - \mathbf{I})\mathbf{F} = -\mathbf{F} = \mathbf{F}$
- Epipole $\mathbf{e}'$ can be obtained from $\mathbf{F}$ as, $\mathbf{e}'^{\mathbf{T}}\mathbf{F} = \mathbf{0}$.
- Given $\mathbf{F}$, we can arrive at a pair of cameras for it! This will be off from the real camera pair only be an unknown projective transformation $\mathbf{H}$ of the cameras/world.
- Generally, $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ and $\mathbf{P}' = [[\mathbf{e}']_\times \mathbf{F} + \mathbf{e}'\mathbf{v}^{\mathbf{T}}|\lambda\mathbf{e}']$ for any 3-vector $\mathbf{v}$ and non-zero constant λ will also have the same $\mathbf{F}$. Many canonical camera pairs with same $\mathbf{F}$.

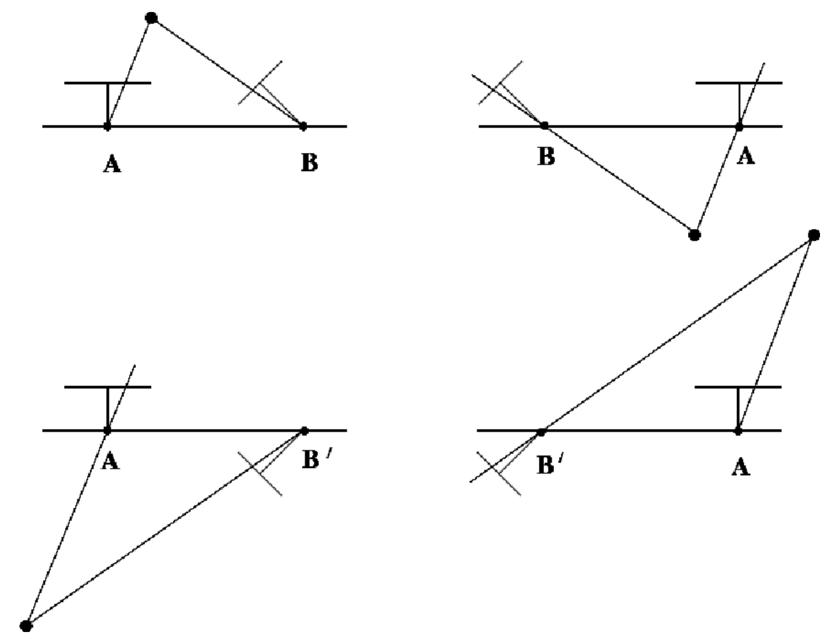
Essential Matrix E

- For calibrated cameras, apply $\mathbf{K}^{-1}$ to $\mathbf{x}$ to get *normalized coordinates* $\mathbf{K}^{-1}\mathbf{x}$. Similarly, $\mathbf{K}'^{-1}\mathbf{x}'$
- The epipolar relation can be expressed in normalized coordinates as: $\mathbf{x}'^{\mathbf{T}}\mathbf{E}\mathbf{x} = \mathbf{0}$ for an **Essential Matrix E**
- $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$ since $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ and $\mathbf{P}' = [\mathbf{R}|\mathbf{t}]$
- Using $\mathbf{F}$, we can write $(\mathbf{K}'^{-1}\mathbf{x}')^{\mathbf{T}}\ \mathbf{E}\ (\mathbf{K}^{-1}\mathbf{x}) = \mathbf{0}$. Thus, $\mathbf{E} = \mathbf{K}'^{\mathbf{T}}\mathbf{F}\mathbf{K}$ in terms of $\mathbf{F}$.
- Properties of Essential Matrix:
  - It has a rank 2, with 2 equal singular values and one 0 singular value. It can be written as (upto scale) $\mathbf{E} = \mathbf{U}\ \ diag(\mathbf{1}, \mathbf{1}, \mathbf{0})\ \mathbf{V}^{\mathbf{T}}$ using SVD.
  - $\mathbf{E}$ only has 5 degrees of freedom: 3 for rotation, 3 for translation, with a scale ambiguity.

Cameras from E

- Given $\mathbf{E} = \mathbf{U}\ diag(\mathbf{1},\mathbf{1},\mathbf{0})\ \mathbf{V}^\mathbf{T}$, the second (normalized) cameras is one of: $[\mathbf{UWV^T} | \pm \mathbf{u_3}]$ or $[\mathbf{UW^TV^T} | \pm \mathbf{u_3}]$

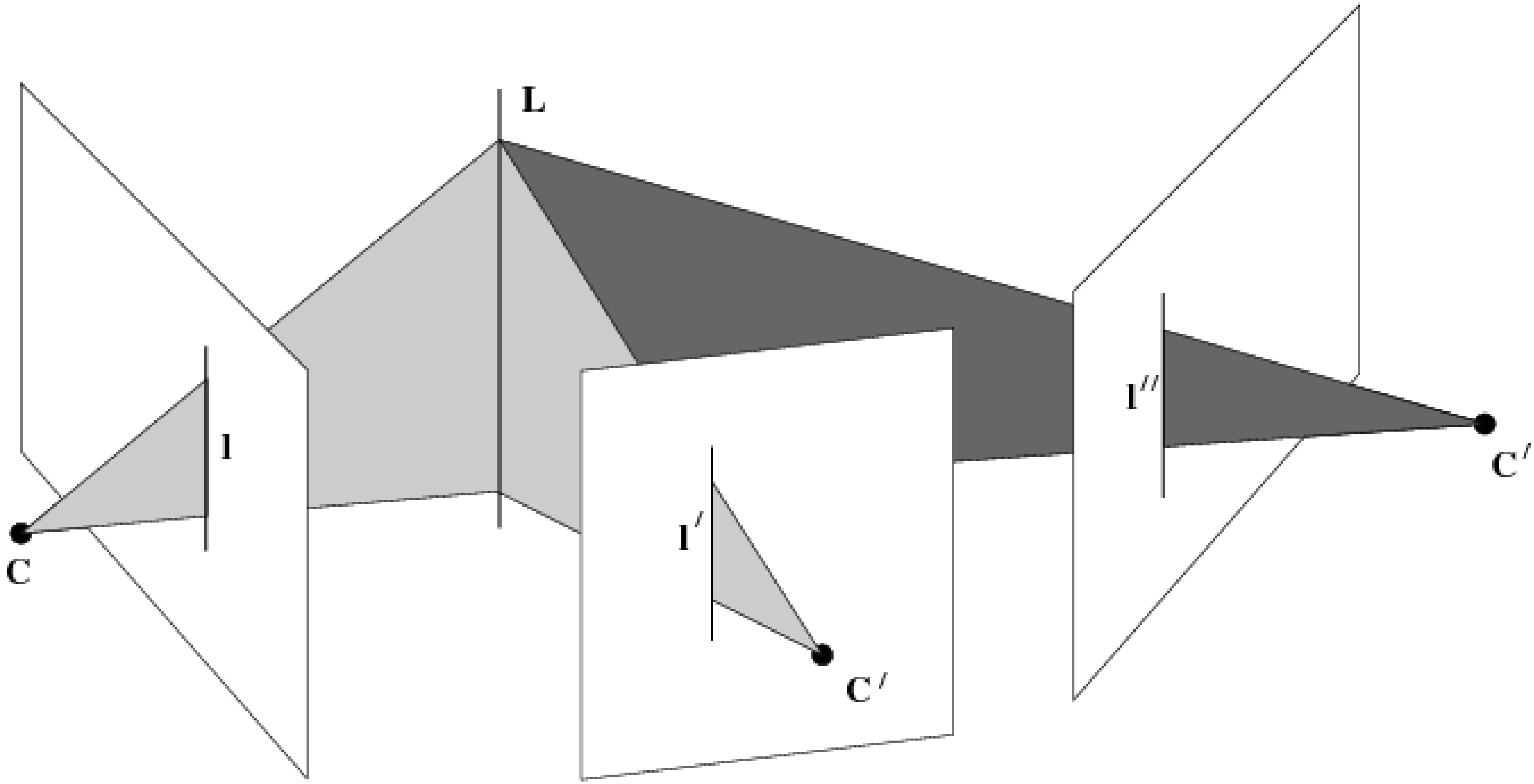
  where $\mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$



- The 4-way ambiguity is resolved using a single point, as a point will be in front of the camera in only one.

Trifocal Tensor of Three Views

A Line in Three Views



Geometry of Three Views

- What happens when 3 cameras are viewing a scene?
- Let the cameras be $\mathbf{P} = [\mathbf{I}|\mathbf{0}], \mathbf{P}' = [\mathbf{A}|\mathbf{a_4}], \mathbf{P}'' = [\mathbf{B}|\mathbf{b_4}]$. We have $\mathbf{a_4} = \mathbf{e}'$ and $\mathbf{b_4} = \mathbf{e}''$ to be the epipoles in views two and three. $\mathbf{A}, \mathbf{B}$ are respectively the $\mathbf{H}_\infty$ to views 2 and 3 respectively from view one
- Conside a world line $\mathbf{L}$ projecting to $\mathbf{l}, \mathbf{l}'$ and $\mathbf{l}''$ in the three views. The planes containing the image lines and their respective camera centres intersect at the world line $\mathbf{L}$
- The planes $\pi = \mathbf{P^T l}$, $\pi' = \mathbf{P'^T l'}$, and $\pi'' = \mathbf{P''^T l''}$ are:

$$\pi = \begin{bmatrix} \mathbf{l} \\ 0 \end{bmatrix}, \quad \pi' = \begin{bmatrix} \mathbf{A^T l'} \\ \mathbf{a_4^T l'} \end{bmatrix}, \quad \pi'' = \begin{bmatrix} \mathbf{B^T l''} \\ \mathbf{b_4^T l''} \end{bmatrix}$$

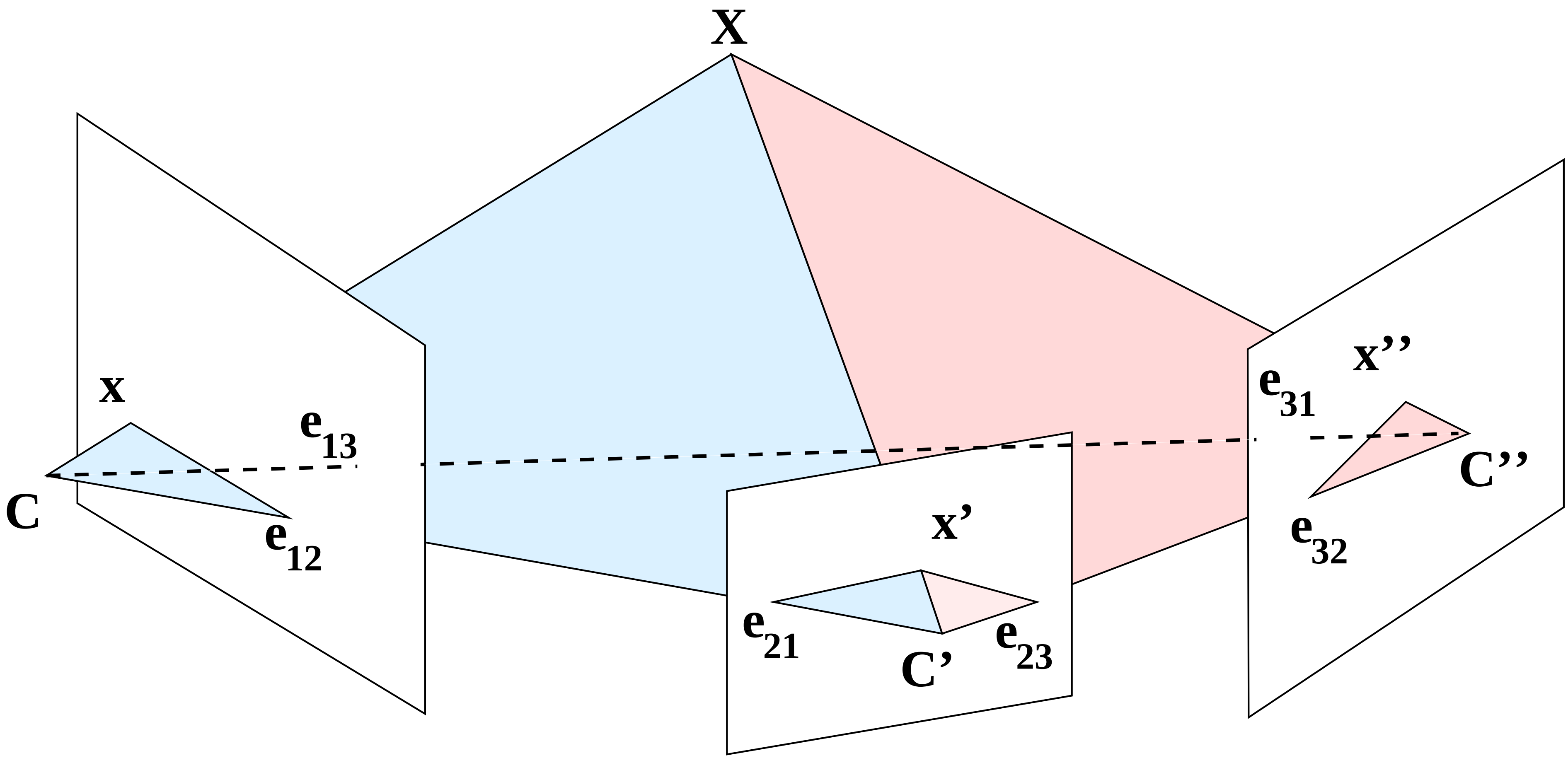
Trifocal Tensor

- Since they meet in a line, the matrix $\mathbf{M} = [\pi \;\; \pi' \;\; \pi'']$ must have rank of 2

$$\mathbf{M} = \begin{bmatrix} \mathbf{l} & \mathbf{A^T l'} & \mathbf{B^T l''} \\ 0 & \mathbf{a_4^T l'} & \mathbf{b_4^T l''} \end{bmatrix} = [\mathbf{m_1} \;\; \mathbf{m_2} \;\; \mathbf{m_3}]$$

- We can write $\mathbf{m_1} = \alpha \mathbf{m_2} + \beta \mathbf{m_3}$. Comparing, $\alpha = k(\mathbf{b_4^T l''})$ and $\beta = -k(\mathbf{a_4^T l'})$. Thus, $\mathbf{l} = (\mathbf{l''^T b_4})\mathbf{A^T l'} - (\mathbf{l'^T a_4})\mathbf{B^T l''}$
- Components of $\mathbf{l}$, $l_i, \; i = 1, 2, 3$ can be given by: $l_i = \mathbf{l''^T}(\mathbf{b_4 a_i^T})\mathbf{l'} - \mathbf{l'^T}(\mathbf{a_4 b_i^T})\mathbf{l''} = \mathbf{l'^T}(\mathbf{a}_i\mathbf{b_4^T} - \mathbf{a_4 b}_i^\mathbf{T})\;\mathbf{l''} = \mathbf{l'^T T}_i\mathbf{l''}$
- $\mathbf{T}_i, i = 1, 2, 3$ are 3×3 matrices that contitute the **trifocal tensor** of views 1, 2, and 3, relating lines in two views to a line in the third as: $\mathbf{l^T} = \mathbf{l'^T} \; [\mathbf{T_1} \;\; \mathbf{T_2} \;\; \mathbf{T_3}]\mathbf{l''}$
- The relation can be written for lines, points, and their combinations in three views.

A Point Three Views



Tensor Acting on Points

- For corresponding points $\mathbf{x} \leftrightarrow \mathbf{x}' \leftrightarrow \mathbf{x}''$, the tensor gives:

$$[\mathbf{x}']_\times \left(\sum_i x_i \mathbf{T}_i \right) [\mathbf{x}'']_\times = 0_{3\times 3}$$

- $x_i, i = 1, 2, 3$ are the respective coordinates of the point. $\sum_i x_i \mathbf{T}_i$ is a 3×3 matrix. The right hand side is a 3-vector of zero values
- Defines a *trilinear* relation in coordinates of $\mathbf{x}, \mathbf{x}', \mathbf{x}''$.
- Nine trilinearities can be defined using different choices of views for each matching triplet of points. However, only 4 of them are linearly independent

Trifocal Tensor: Properties

- The trifocal tensor $[\mathbf{T}_i] = [\mathbf{T_1}\ \mathbf{T_2}\ \mathbf{T_3}]$ has 26 independent elements and a common scale factor.
- The tensor only has 18 independent degrees of freedom. This is given by 3×11 for 3 camera matrices **minus** 15 for an arbitrary projective transformation.
- $\mathbf{H} = [\mathbf{T_1^T}\ \ \mathbf{T_2^T}\ \ \mathbf{T_3^T}]\,\mathbf{l}'$ gives the point homography $\mathbf{H_{13}}(\mathbf{l}')$ between views 1 and 3 due to the plane defined by $\mathbf{C}'$ and $\mathbf{l}'$. Similarly, $\mathbf{H_{12}}(\mathbf{l}'') = [\mathbf{T_1}\ \ \mathbf{T_2}\ \ \mathbf{T_3}]\,\mathbf{l}''$
- It represents a projective property of the combination of 3 cameras only, independent of the scene structure.
- 3 matching lines give 2 linear equations in entries of $[\mathbf{T}_i]$. 3 matching points give 4 equations. 26 equations are needed to estimate the tensor. There is no easy way to exploit the reduced real degrees of freedom.

View Transfer

- Given the projections of a point in 2 views, where is it in a third view? Given $\mathbf{x}, \mathbf{x}'$, what is $\mathbf{x}''$?
- Known pairwise fundamental matrices $\mathbf{F_{21}}, \mathbf{F_{32}}, \mathbf{F_{13}}$:
  - $\mathbf{F_{31}x}$ and $\mathbf{F_{32}x}'$ are epipolar lines due to projections of a world point.
  - $\mathbf{x}''$ is the intersection of these lines. Hence, $\mathbf{x}'' = (\mathbf{F_{31}x}) \times (\mathbf{F_{32}x}')$
- Trifocal tensor is known:
  - The point version of the tensor gives 4 independent trilinear equations in the coordinates $\mathbf{x}, \mathbf{x}', \mathbf{x}''$. Given two points, the third can be calculated easily from this relation.

Thank You!

Many figures are from the book

Multiview Geometry in Computer Vision

by **Richard Hartley and Andrew Zisserman**