Epipolar geometry

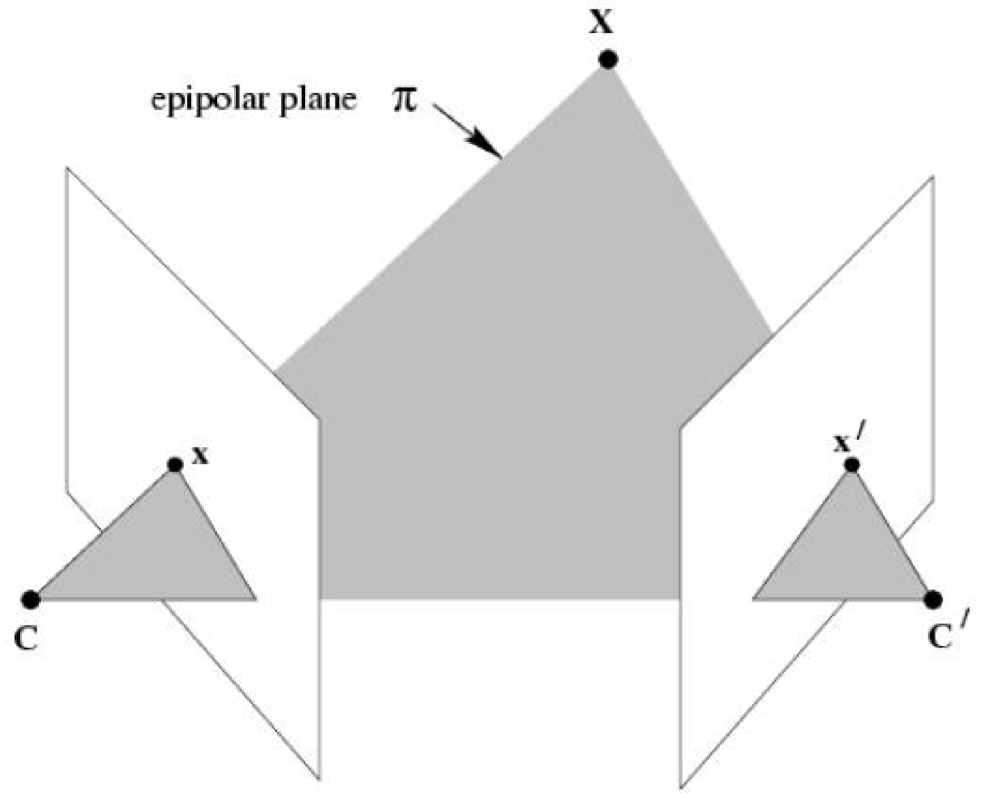
Three questions:

(i) **Correspondence geometry:** Given an image point x in the first view, how does this constrain the position of the corresponding point x' in the second image?

(ii) **Camera geometry (motion):** Given a set of corresponding image points $\{x_i \leftrightarrow x'_i\}$, $i=1,\ldots,n$, what are the cameras P and P' for the two views? Or what is the geometric transformation between the views?

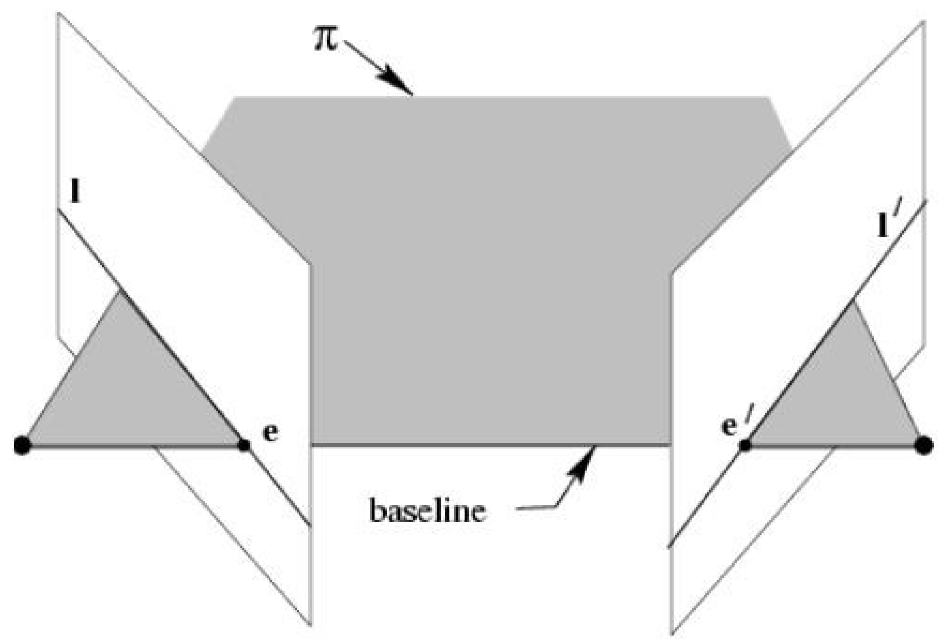
(iii) **Scene geometry (structure):** Given corresponding image points $x_i \leftrightarrow x'_i$ and cameras P, P', what is the position of the point X in space?

The epipolar geometry



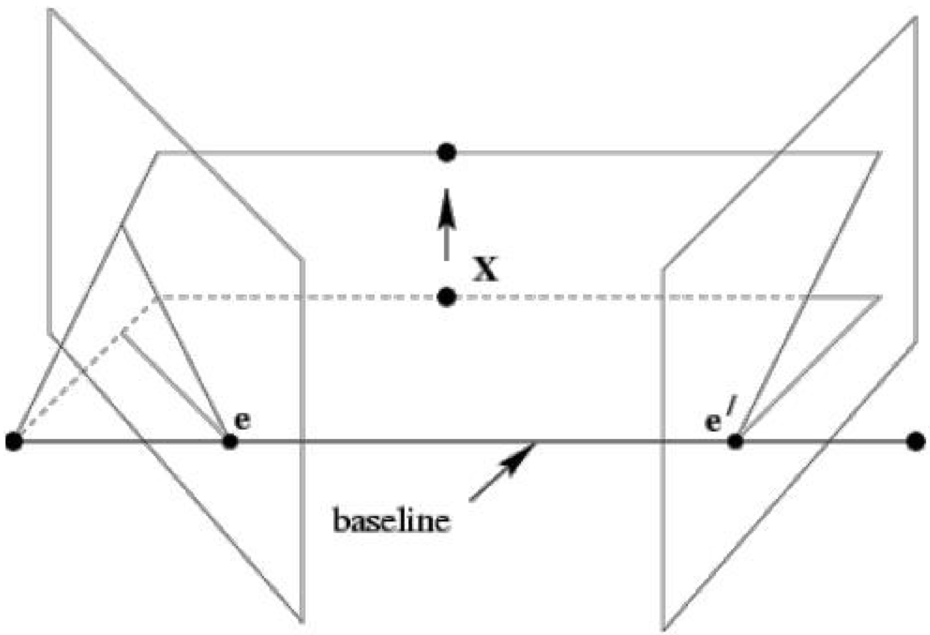
C,C',x,x' and X are coplanar

The epipolar geometry



All points on π project on l and l'

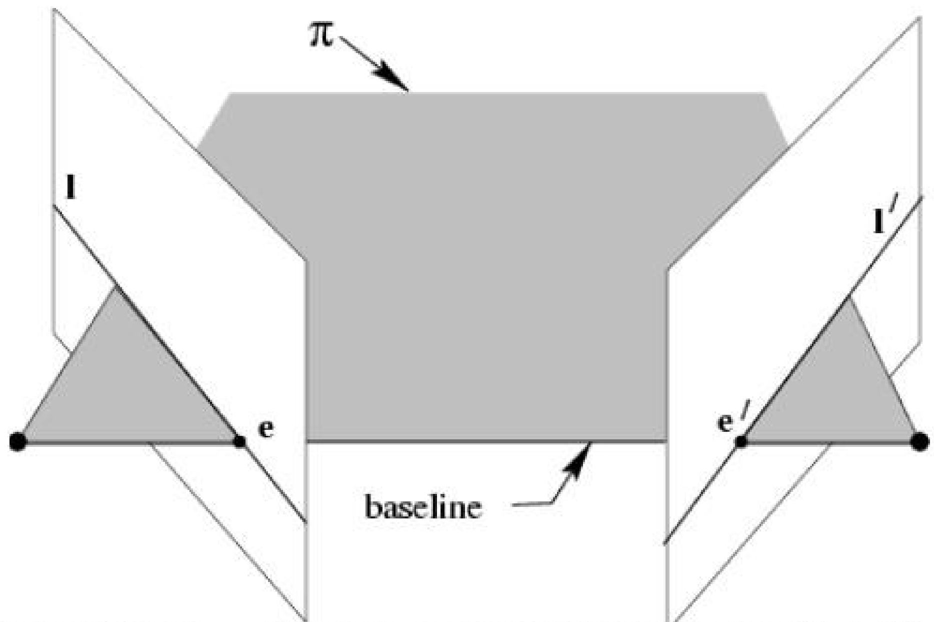
The epipolar geometry



Family of planes π and lines l and l'
Intersection in e and e'

The epipolar geometry

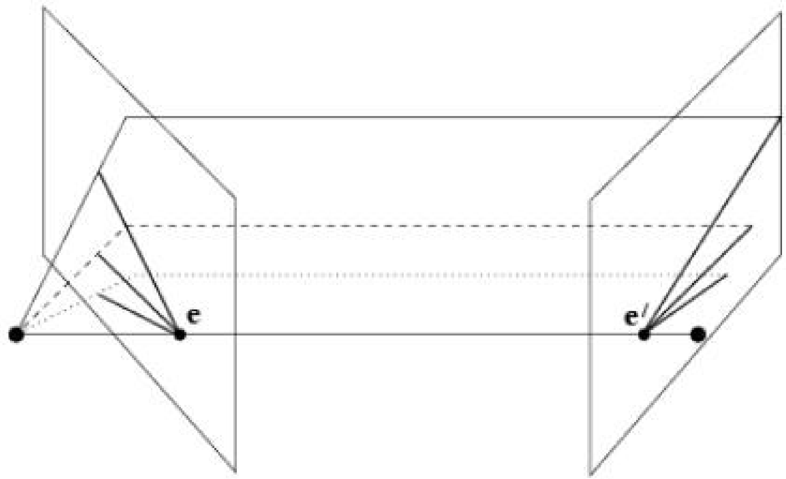
epipoles e,e'
= intersection of baseline with image plane
= projection of projection center in other image
= vanishing point of camera motion direction

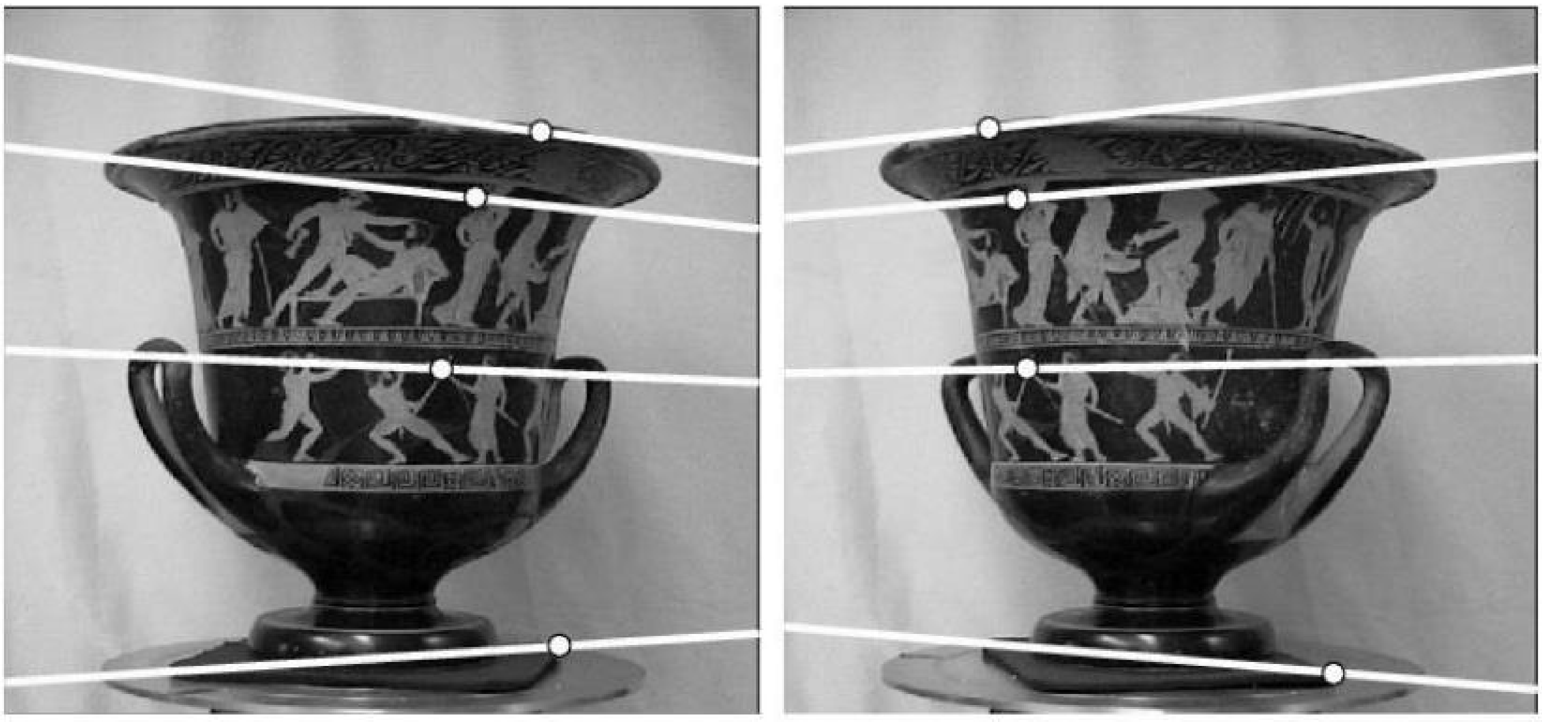


an epipolar plane = plane containing baseline (1-D family)

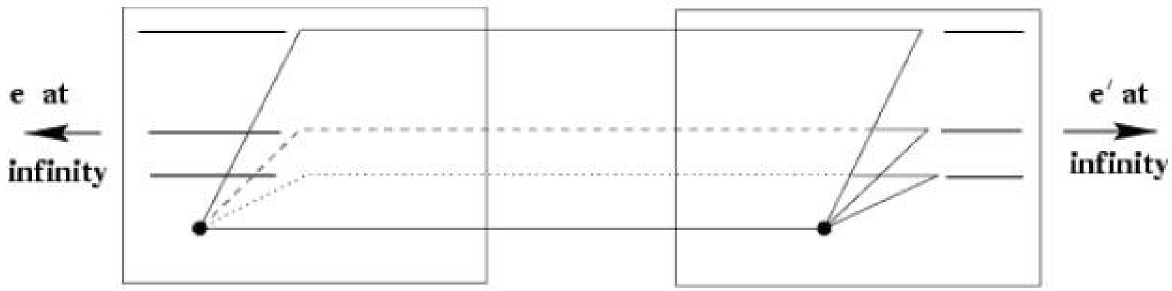
an epipolar line = intersection of epipolar plane with image
(always come in corresponding pairs)

Example: converging cameras

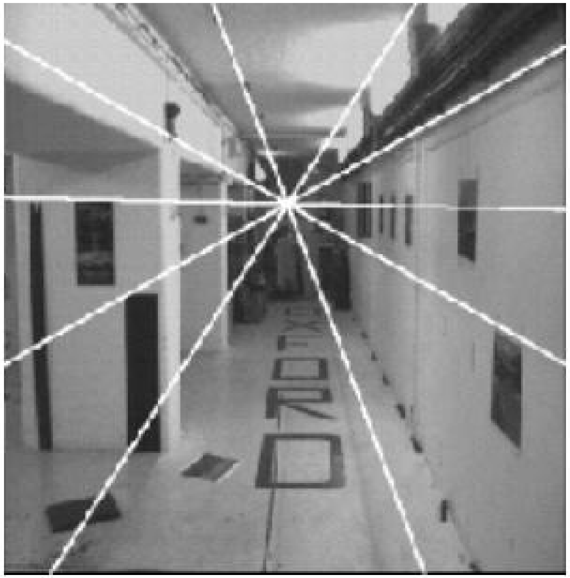


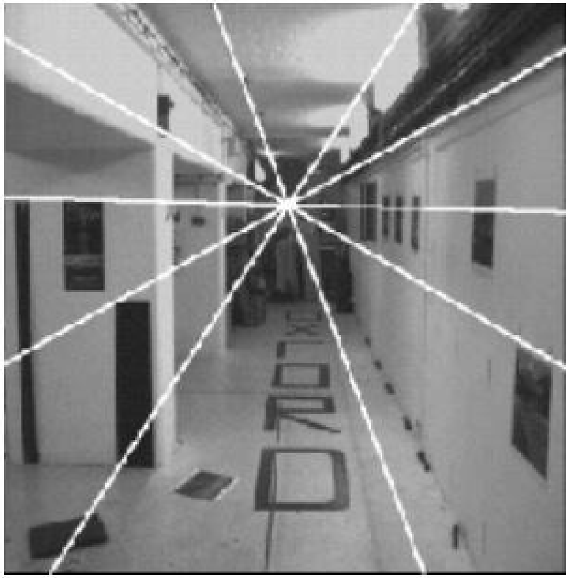


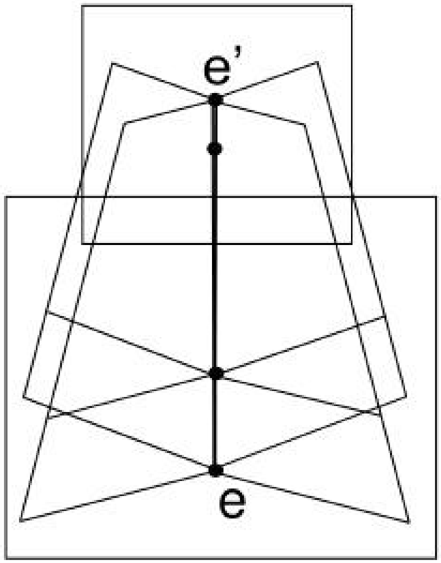
Example: motion parallel with image plane



Example: forward motion







Matrix form of cross product

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

$$a \cdot (a \times b) = 0$$
$$b \cdot (a \times b) = 0$$

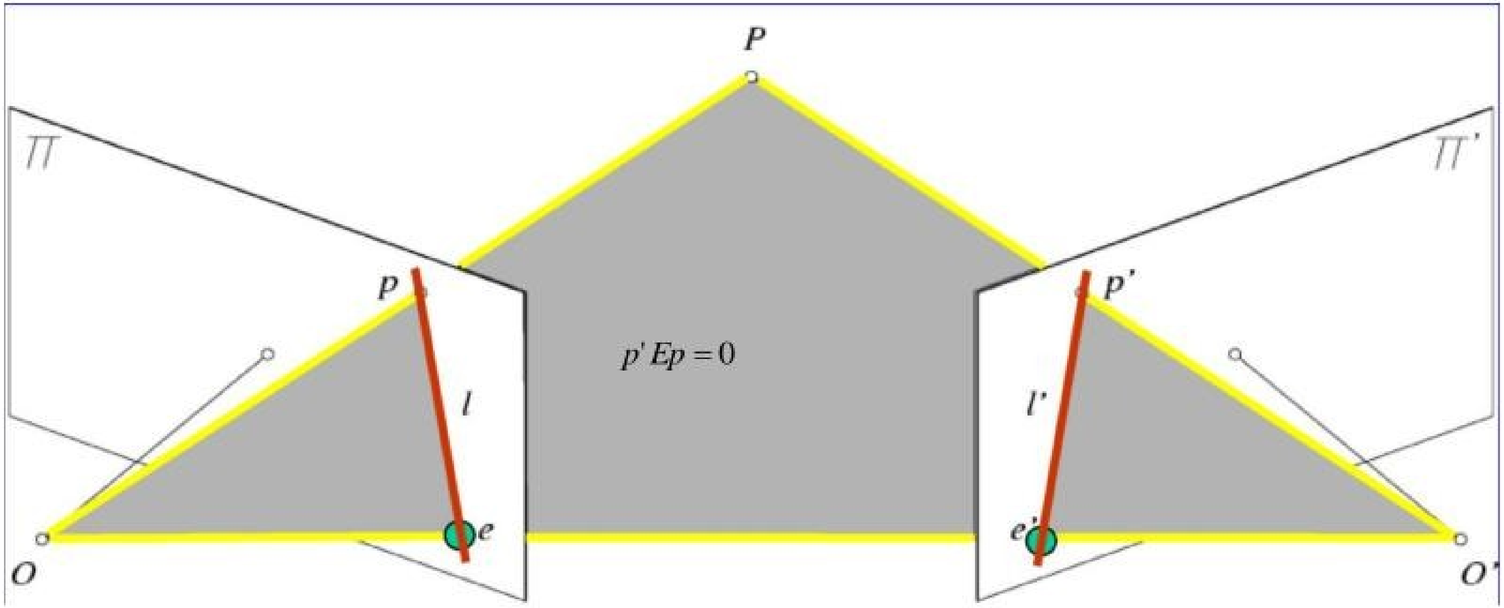
Geometric transformation

$$P' = RP + t$$

$$p = MP \quad \text{with} \quad M = [I \mid 0]$$

$$p' = M'P' \text{ with } M' = [R \mid t]$$

Calibrated Camera

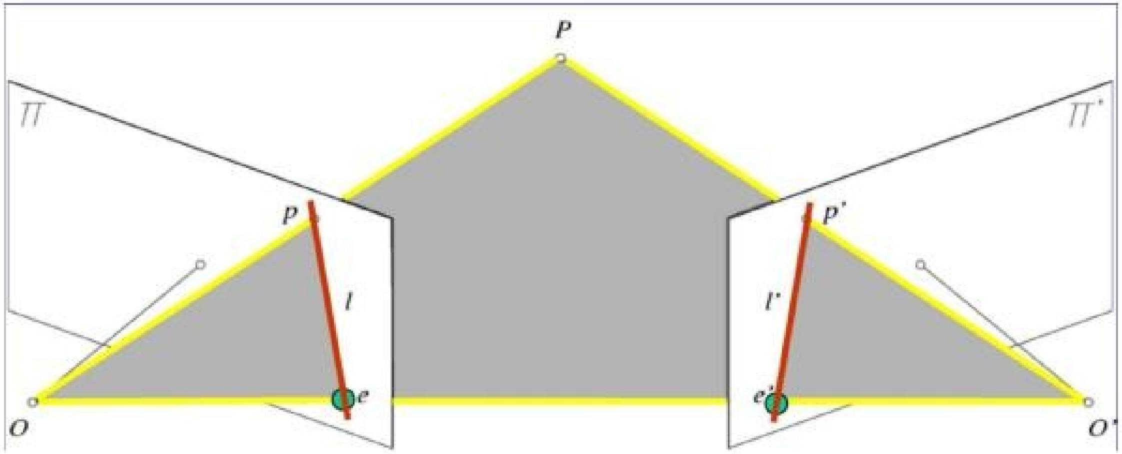


$\vec{Op}, \vec{OO'}, \vec{O'p'}$ are co-planar $\Rightarrow$ $p' \cdot [t \times (Rp)] = 0$ with $\begin{cases} p = (u, v, 1)^T \\ p' = (u', v', 1)^T \end{cases}$

$\Downarrow$

Essential matrix $p'Ep = 0$ with $E = [t_\times]R = SR$

Uncalibrated Camera



p and p' points in pixel coordinates corresponding to $\widehat{p}$ and $\widehat{p}'$ in camera coordinates

$$\widehat{p} = M_{\text{int}}^{-1}\; p \quad \text{and} \quad \widehat{p}' = M'^{-1}_{\text{int}}\; p'$$

$$\widehat{p}'\, E\, \widehat{p} = 0$$

$$p'^{T}\, F\, p = 0$$

$$\text{with} \quad F = M'^{-T}_{\text{int}}\; E\, M_{\text{int}}^{-1}$$

$\boxed{\text{Fundamental matrix}}$

Properties of fundamental and essential matrix

- Matrix is 3 x 3
- **Transpose :** If F is essential matrix of cameras (P, P').
  F^T is essential matrix of camera (P',P)

- **Epipolar lines:** Think of *p* and *p'* as points in the projective plane then *F p* is projective line in the right image.
  That is $l' = F\,p \quad l = F^T p'$

- **Epipole:** Since for any p the epipolar line $l' = F\,p$ contains the epipole *e'*. Thus $(e'^T F)\,p = 0$ for a all *p* .
  Thus $e'^T F = 0$ *and* $F\,e = 0$

Fundamental matrix

- Encodes information of the intrinsic and extrinisic parameters
- F is of rank 2, since S has rank 2 (R and M and M' have full rank)
- Has 7 degrees of freedom
  There are 9 elements, but scaling is not significant and det F = 0

Essential matrix

- Encodes information of the extrinisic parameters only
- E is of rank 2, since S has rank 2 (and R has full rank)
- Its two nonzero singular values are equal
- Has only 5 degrees of freedom, 3 for rotation, 2 for translation

Scaling ambiguity

$$P' = RP + t$$

$$p = \frac{P}{\hat{z}^T P} \qquad\qquad p' = \frac{RP + t}{\hat{z}^T (RP + t)}$$

Depth Z and Z' and t can only be recovered up to a scale factor
Only the direction of translation can be obtained

Least square approach

$$\text{Minimize} \sum_{i=1}^{n} (p_i' F p_i)^2$$

$$\text{under the constraint} \mid F \mid^2 = 1$$

We have a homogeneous system A f =0
The least square solution is smallest singular value of A,
i.e. the last column of V in SVD of A = U D V^T

Computing Fundamental Matrix from Point Correspondences

- The fundamental matrix is defined by the equation $\mathbf{x'}_\mathbf{i}^\mathrm{T}\, \mathbf{F}\, \mathbf{x}_\mathbf{i} = 0$ for any pair of corresponding points $\mathbf{x}_\mathrm{i}$ and $\mathbf{x'}_\mathrm{i}$ in the 2 images
- The equation for a pair of points $(x, y, 1)$ and $(x', y', 1)$ is: $x'x\, f_{11} + x'y\, f_{12} + x' f_{13} +$
  $+ y'x\, f_{21} + y'y\, f_{22} + y' f_{23} +$
  $+ x\, f_{31} + y\, f_{32} + f_{33} = 0$
- For n point matches:

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

Computing Fundamental Matrix from Point Correspondences

- We have a homogeneous set of equations $\mathbf{A}\,\mathbf{f} = 0$
- $\mathbf{f}$ can be determined only up to a scale, so there are 8 unknowns, and at least 8 point matchings are needed
  - hence the name "8 point algorithm"
- The least square solution is the singular vector corresponding the smallest singular value of $\mathbf{A}$, i.e. the last column of $\mathbf{V}$ in the SVD $\mathbf{A} = \mathbf{U}\,\mathbf{D}\,\mathbf{V}^T$

The Normalized Eight-Point Algorithm (Hartley, 1995)

- Center the image data at the origin, and scale it so the mean squared distance between the origin and the data points is 2 pixels: $q_i = T p_i$, $q_i' = T' p_i'$.
- Use the eight-point algorithm to compute F from the points q_i and q_i' .
- Enforce the rank-2 constraint.
- Output $T^T \mathsf{F} T'$.

Non-Linear Least Squares Approach

Minimize

$$\sum_{i=1}^{n}(d^2(p_i^{'}Fp_i)+d^2(p_iFp'_i))$$

with respect to the coefficients of F
Using an appropriate rank 2 parameterization

Locating the epipoles

$$p'^T Fe = 0$$

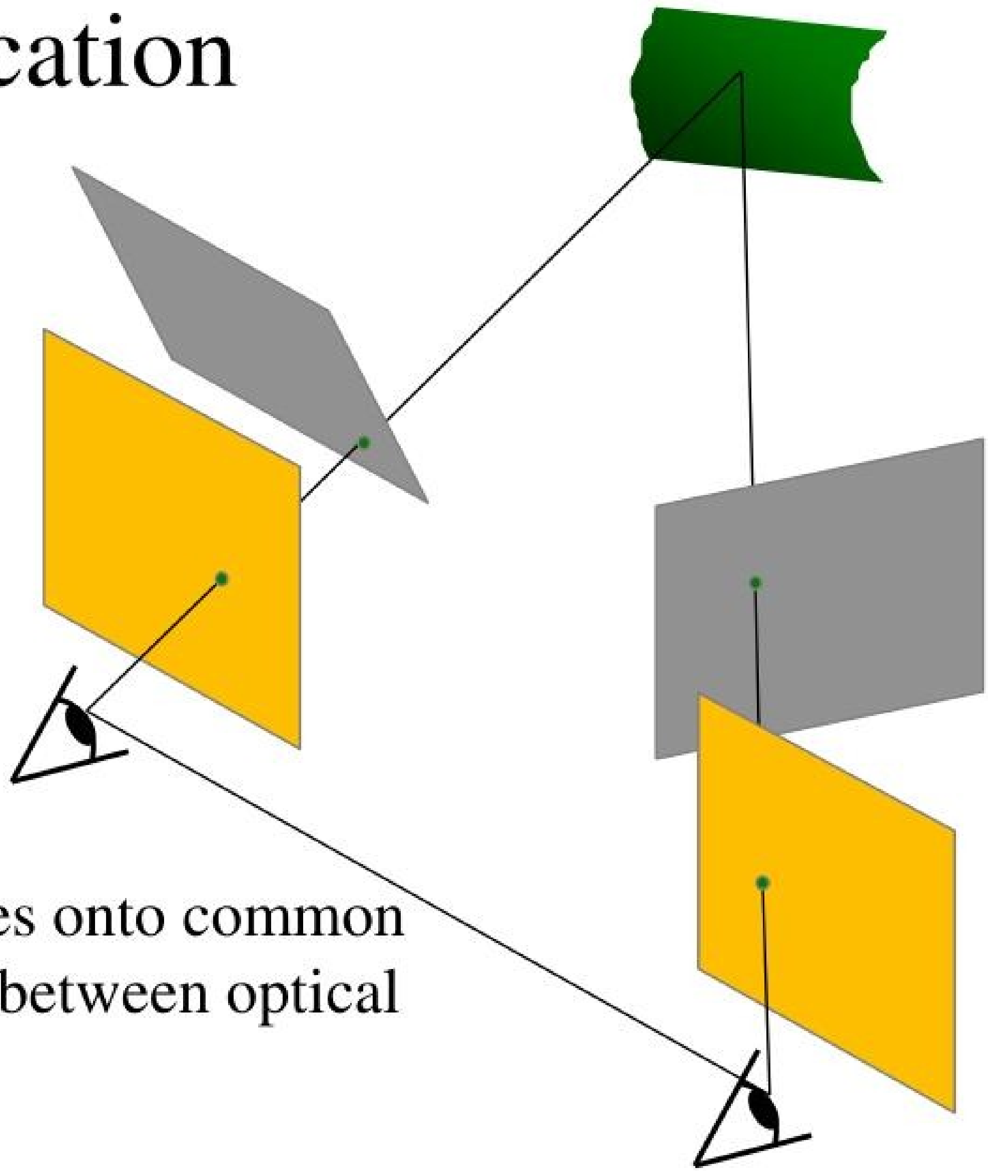
$$Fe = 0$$

e is the nullspace of F;

e' is the nullspace of F^T

SVD of F = UDV^T.

Rectification



- Image Reprojection
  - reproject image planes onto common plane parallel to line between optical centers

Rectification

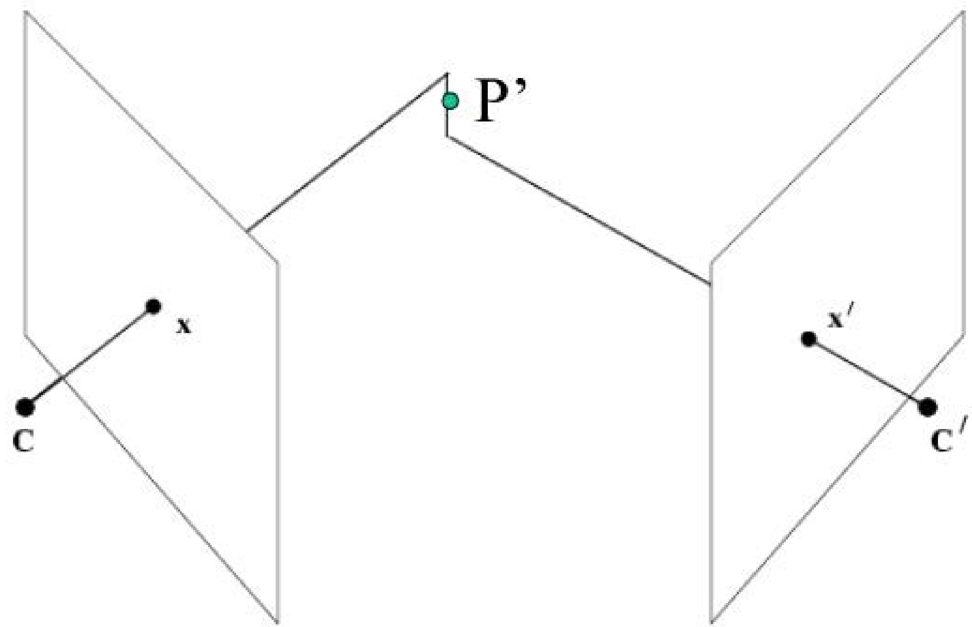
- Rotate the left camera so epipole goes to infinity along the horizontal axis
- Apply the same rotation to the right camera
- Rotate the right camera by R
- Adjust the scale

3D Reconstruction

- **Stereo**: we know the viewing geometry (extrinsic parameters) and the intrinsic parameters: Find correspondences exploiting epipolar geometry, then reconstruct
- **Structure from motion** (with calibrated cameras): Find correspondences, then estimate extrinsic parameters (rotation and direction of translation), then reconstruct.
- **Uncalibrated cameras**: Find correspondences,

  Compute projection matrices (up to a projective transformation), then reconstruct up to a projective transformation.

Reconstruction by triangulation



If cameras are intrinsically and extrinsically calibrated, find P as the midpoint of the common perpendicular to the two rays in space.

Triangulation

*a*p’ ray through C’ and p’,
*b*Rp + T ray though C and p expressed in right coordinate system

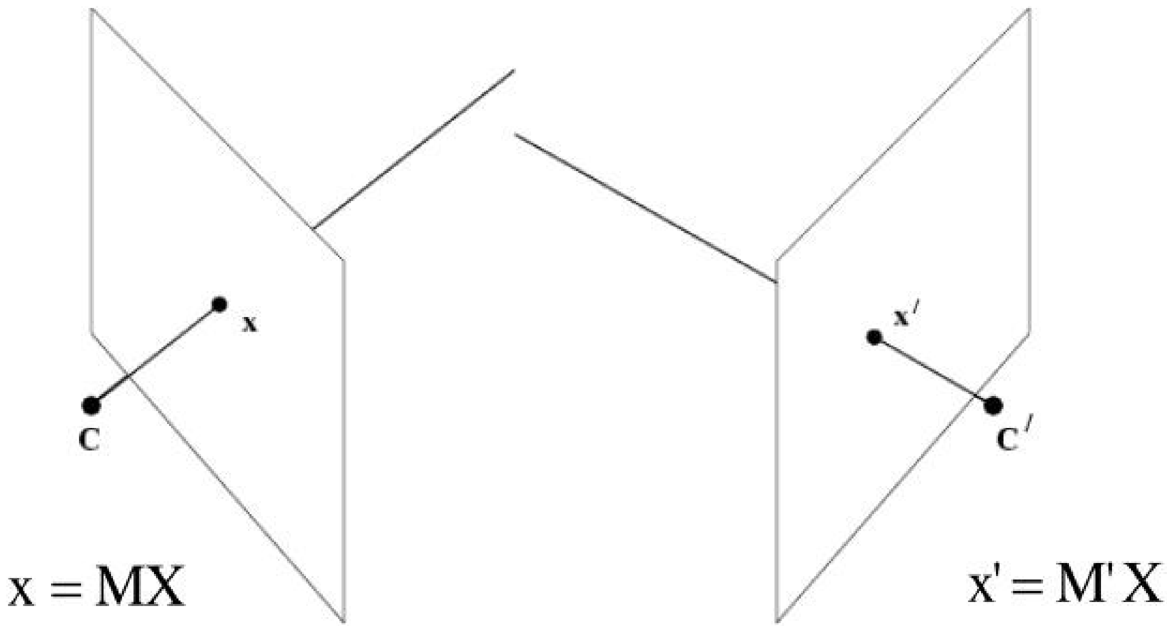
$$ap' - bRp + c(p' \times Rp) = T$$

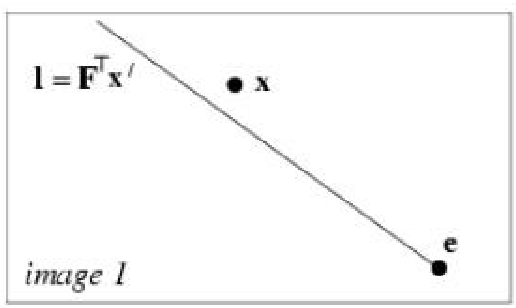
R = ?
T = ?

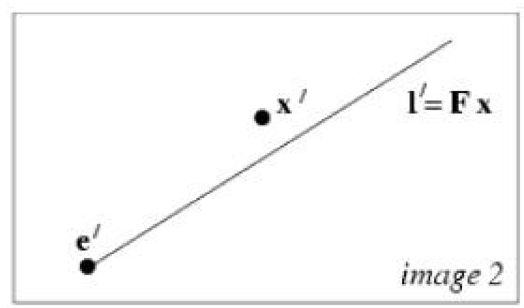
$$R = R_r R_l^T$$

$$T = -T_r + RT_l$$

Point reconstruction







Linear triangulation

$$\mathrm{x} = \mathrm{MX} \quad \mathrm{x}' = \mathrm{M}'\mathrm{X}$$

$$\mathrm{x} \times \mathrm{MX} = 0$$

$$\mathrm{x}' \times \mathrm{M}'\mathrm{X}' = 0$$

$$\longrightarrow \quad \mathrm{AX} = 0$$

$$x\left(\mathrm{m}_3{}^{\mathrm{T}}\mathrm{X}\right) - \left(\mathrm{m}_1{}^{\mathrm{T}}\mathrm{X}\right) = 0$$

$$y\left(\mathrm{m}_3{}^{\mathrm{T}}\mathrm{X}\right) - \left(\mathrm{m}_2{}^{\mathrm{T}}\mathrm{X}\right) = 0$$

Linear combination of 2 other equations

$$x\left(\mathrm{m}_2{}^{\mathrm{T}}\mathrm{X}\right) - y\left(\mathrm{m}_1{}^{\mathrm{T}}\mathrm{X}\right) = 0$$

$$\mathrm{A} = \begin{bmatrix} x\mathrm{m}_3{}^{\mathrm{T}} - \mathrm{m}_1{}^{\mathrm{T}} \\ y\mathrm{m}_3{}^{\mathrm{T}} - \mathrm{m}_2{}^{\mathrm{T}} \\ x'\mathrm{m}_3{}'^{\mathrm{T}} - \mathrm{m}_1{}'^{\mathrm{T}} \\ y'\mathrm{m}'_3{}^{\mathrm{T}} - \mathrm{m}'_2{}^{\mathrm{T}} \end{bmatrix}$$

homogeneous $\quad \|\mathrm{X}\| = 1$

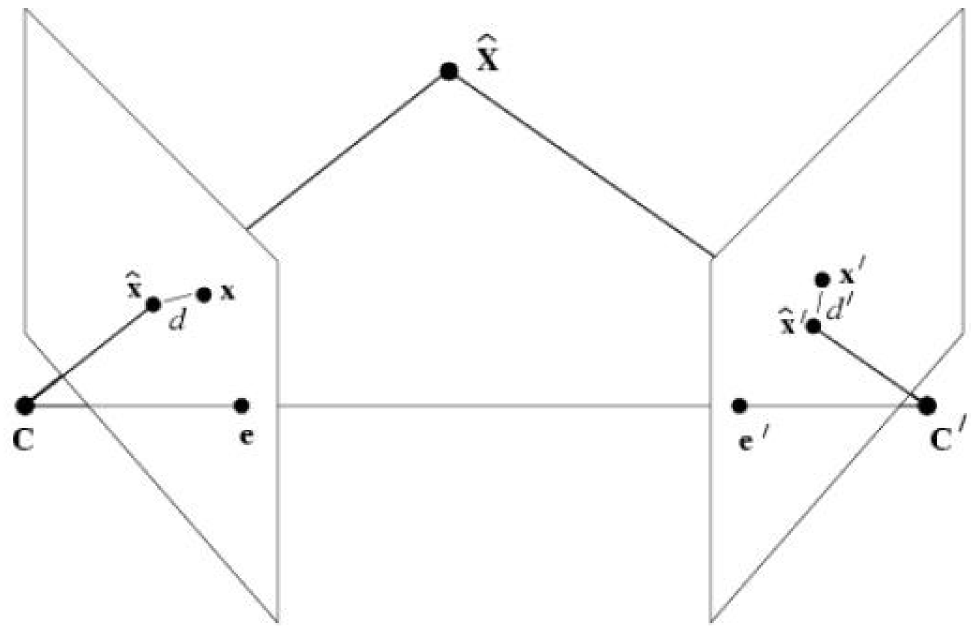
$\mathrm{AX} = 0$ Homogenous system:

X is last column of V in the SVD of A= $U\Sigma V^T$

geometric error

$$d(\mathrm{x}, \hat{\mathrm{x}})^2 + d(\mathrm{x}', \hat{\mathrm{x}}')^2 \text{ subject to } \hat{\mathrm{x}}'^{\mathrm{T}} \mathrm{F} \hat{\mathrm{x}} = 0$$

$$\text{or equivalently subject to } \hat{\mathrm{x}} = \mathrm{M}\hat{\mathrm{X}} \text{ and } \hat{\mathrm{x}}' = \mathrm{M}'\hat{\mathrm{X}}$$

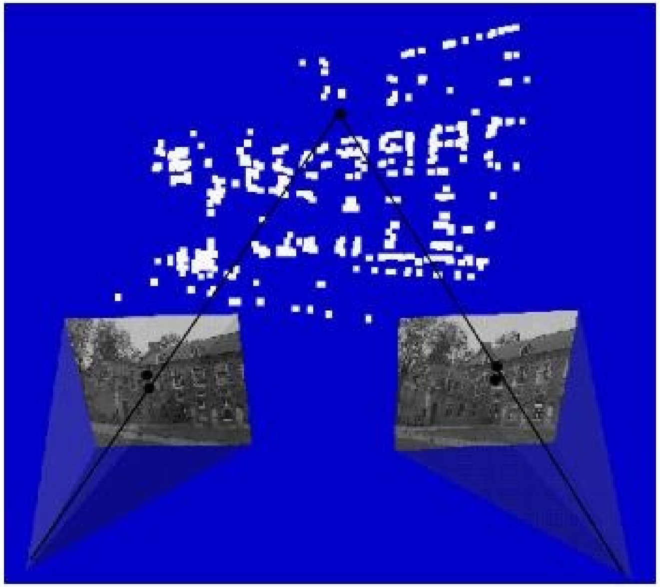


Geometric error

Reconstruct matches in projective frame
by minimizing the reprojection error

$$d(\mathrm{x}, \mathrm{MX})^2 + d(\mathrm{x}', \mathrm{M}'\mathrm{X})^2$$

Non-iterative optimal solution



Reconstruction for intrinsically calibrated cameras

- Compute the essential matrix E using normalized points.

- Select M=[I|0]   M'=[R|T] then $E=[T_x]R$
- Find T and R using SVD of E

Decomposition of E

$E = [T_x]R$ E can be computed up to scale factor

$$EE^T = [T_x]RR^T[T_x]^T = \begin{bmatrix} T_y^2 + T_z^2 & -T_xT_y & -T_xT_z \\ -T_xT_y & T_x^2 + T_z^2 & -T_yT_z \\ -T_xT_z & -T_yT_z & T_x^2 + T_y^2 \end{bmatrix}$$

$$Tr(EE^T) = 2||T||$$

T can be computed up to sign (EE^T is quadratic)

$\longrightarrow$ Four solutions for the decomposition, Correct one corresponds to positive depth values

SVD decomposition of E

- $E = U\Sigma V^T$

$$[T_\times] = UZU^T \quad R = UWV^T \quad \text{or} \quad R = UW^TV^T$$

$$\text{with} \quad W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Reconstruction from uncalibrated cameras

Reconstruction problem:

given $x_i \leftrightarrow x'_i$, compute M,M' and X_i

$$\mathrm{x}_i = \mathrm{M}\mathrm{X}_i \qquad \mathrm{x}'_i = \mathrm{M}\mathrm{X}'_i \qquad \text{for all } i$$

without additional information possible
only up to projective ambiguity

Projective Reconstruction Theorem

- Assume we determine matching points $\mathbf{x_i}$ and $\mathbf{x_i}$'. Then we can compute a unique Fundamental matrix $\mathbf{F}$.
- The camera matrices **M**, **M'** cannot be recovered uniquely
- Thus the reconstruction ($\mathbf{X_i}$) is not unique
- There exists a projective transformation **H** such that

$$X_{2,i} = HX_{1,i}, \quad M_2 = M_1 H^{-1} \quad M'_2 = M'_1 H^{-1}$$

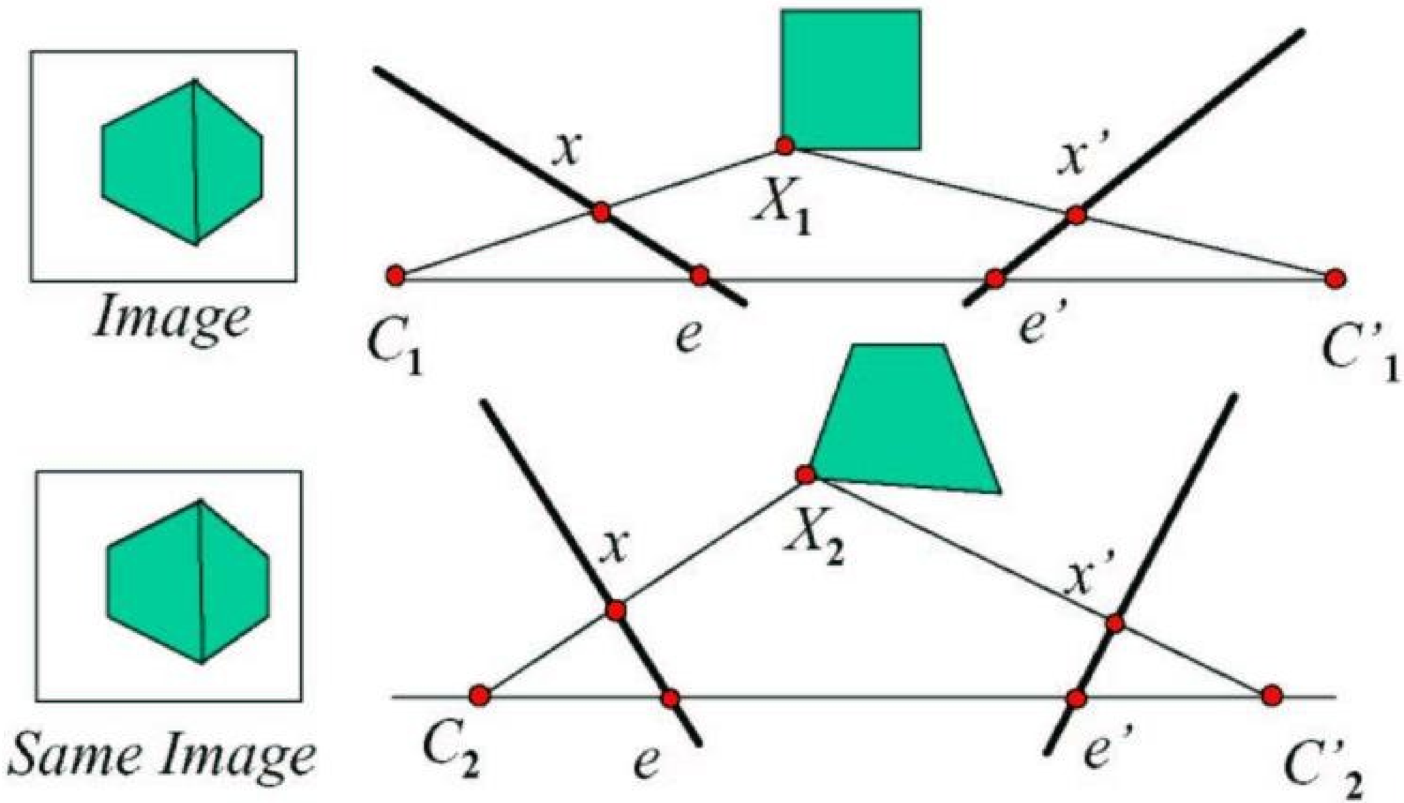
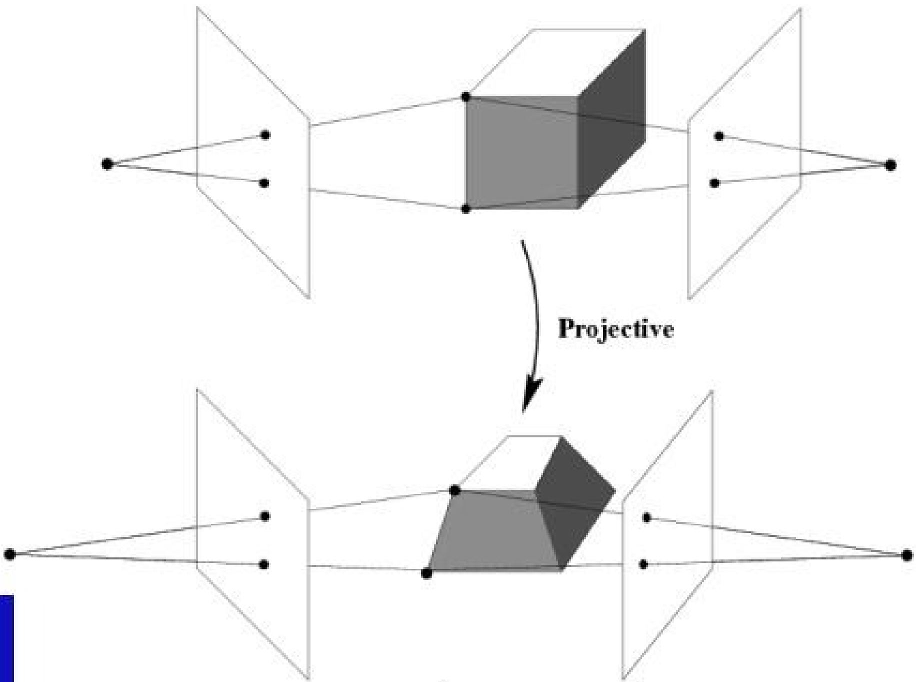
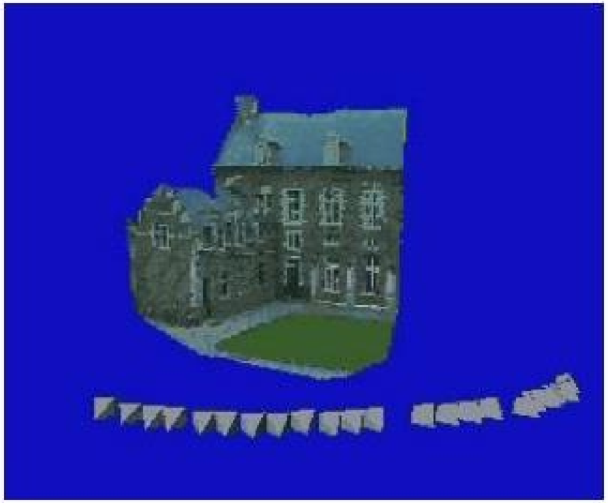


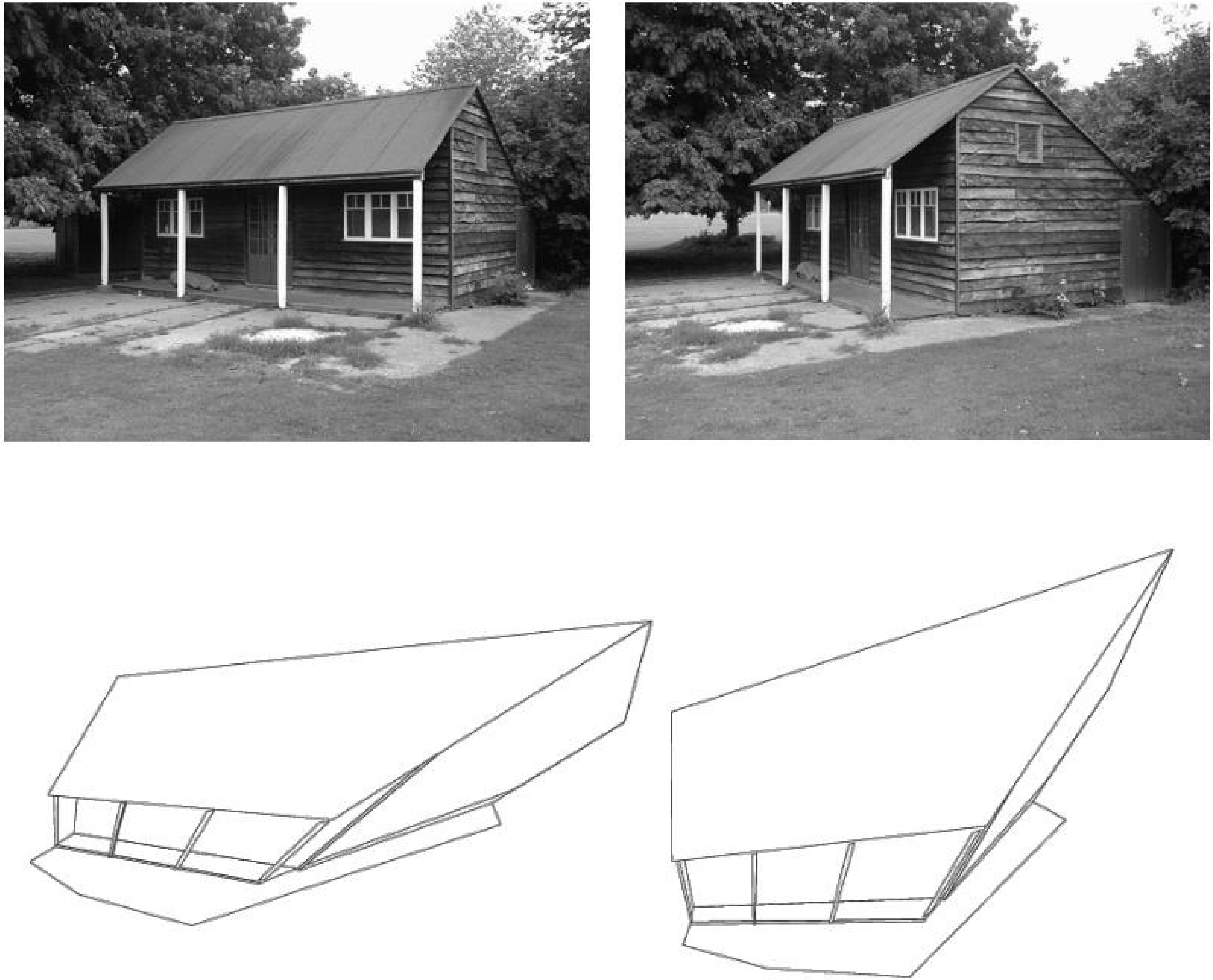
Image
x
X1
x'
C1
e
e'
C'1
Same Image
x
X2
x'
C2
e
e'
C'2

Reconstruction ambiguity: projective





$$x_i = MX_i = \left(MH_P^{-1}\right)\left(H_P X_i\right)$$



Projective Reconstruction Theorem (Consequences)

- We can compute a projective reconstruction of a scene from 2 views based on image correspondences alone
- We don't have to know anything about the calibration or poses of the cameras
- The true reconstruction is within a projective transformation **H** of the projective reconstruction: $\mathbf{X}_{2i} = \mathbf{H}\,\mathbf{X}_{1i}$

Reconstruction Ambiguities

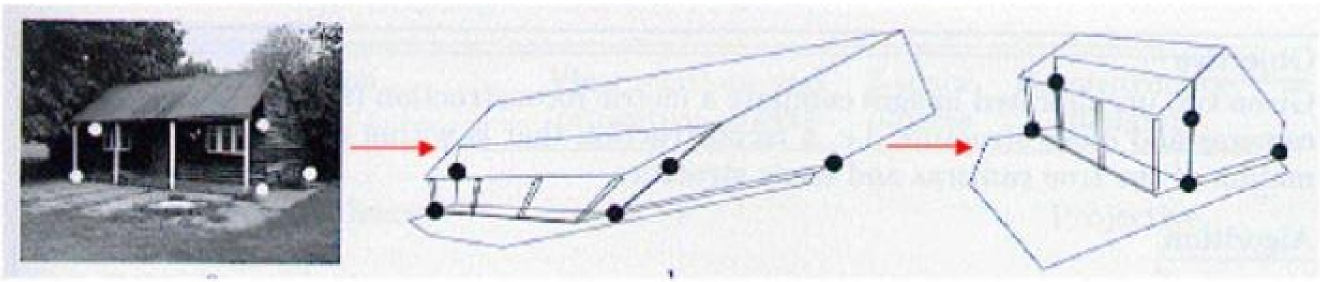
- If the reconstruction is derived from real images, there is a **true** reconstruction that can produce the actual points **X**i of the scene
- Our reconstruction may differ from the actual one
  - If the cameras are calibrated but their relative pose is unknown, then angles between rays are the true angles, and the reconstruction is correct within a similarity (we cannot get the scale)
    - Euclidean or metric reconstruction
  - If we don't use calibration, then we get a projective reconstruction

From Projective to Metric Reconstruction

- Compute homography H such that $X_{Ei}=HX_i$ for 5 or more control points X_{Ei} with known Euclidean position.
- Then the metric reconstruction is

$$M_{\mathrm{M}} = MH^{-1} \quad M'_{\mathrm{M}} = M'H^{-1} \quad X_{M,i} = HX_i$$

Rectification using 5 points



Stratified Reconstruction

- Begin with a projective reconstruction
- Refine it to an affine reconstruction
  - Parallel lines are parallel; ratios along parallel lines are correct
  - Reconstructed scene is then an affine transformation of the actual scene
- Then refine it to a metric reconstruction
  - Angles and ratios are correct
  - Reconstructed scene is then a scaled version of actual scene

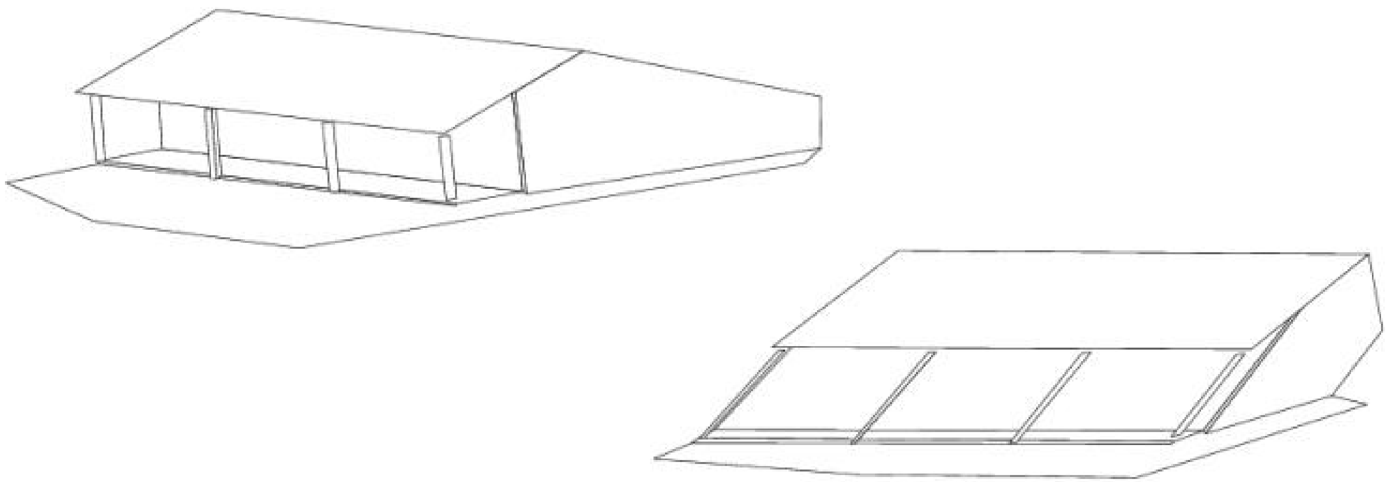
From Projective to Affine Reconstruction

- Find 3 intersections of sets of lines in the scene that are supposed to be parallel
  - These 3 points define a plane $\boldsymbol{\pi}$
- Find a transformation **H** that maps the plane $\boldsymbol{\pi}$ to the plane at infinity $(0, 0, 0, 1)^T$:
  - This plane contains all points at infinity: $(0, 0, 0, 1)\,(x, y, z, 0)^T = 0$
  - $\mathbf{H}^{-T}\boldsymbol{\pi} = (0, 0, 0, 1)^T$, or $\mathbf{H}^T(0, 0, 0, 1)^T = \boldsymbol{\pi}$

$$\begin{bmatrix} 1 & 0 & 0 & \pi_1 \\ 0 & 1 & 0 & \pi_2 \\ 0 & 0 & 1 & \pi_3 \\ 0 & 0 & 0 & \pi_4 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} \Rightarrow \mathbf{H} = \begin{bmatrix} \mathbf{I} \mid \mathbf{0} \\ \mathbf{p}^T \end{bmatrix}$$

Apply **H** to scene points, and to cameras **P** and **P**'

Affine reconstructions



From affine to metric

- Use constraints from scene orthogonal lines
- Use constraints arising from having the same camera in both images

Reconstruction from N Views

- Projective or affine reconstruction from a possible large set of images
- Given a set of camera M^i,
- For each camera M^i a set of image point x_j^i
- Find 3D points X_j and cameras M^i, such that $M^i X_j = x_j^i$

Bundle adjustment

- Solve following minimization problem
- Find M^i and X_j that minimize

$$\sum_{i,j} d(M^i X_j, x_j^i)^2$$

- Levenberg Marquardt algorithm
- Problems many parameters
  11 per camera, 3 per 3d point
- Useful as final adjustment step for bundles of rays