Multi-linear Systems and Invariant Theory

in the Context of Computer Vision and Graphics

Class 4: Mutli-View 3D-from-2D

CS329
Stanford University

Amnon Shashua

Material We Will Cover Today

- Epipolar Geometry and Fundamental Matrix
- The plane+parallax model and relative affine structure
- Why 3 views?
- Trifocal Tensor

Reminder (from class 1):

$$p \cong [I;0]\overline{P}$$

$$p = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \qquad \overline{P} = \begin{pmatrix} p \\ \mu' \end{pmatrix}$$

$$p' \cong [H;e']\overline{P}$$

$$p' \cong \left[\lambda H + e'n^T; \quad e'\right]P \qquad P = \begin{pmatrix} p \\ \mu \end{pmatrix}$$

$$\Longrightarrow \qquad \boxed{p' \cong H_\pi p + \mu e'}$$

$$H_\pi = \lambda H + e'n^T$$

H_π Stands for the family of 2D projective transformations between two fixed images induced by a plane in space

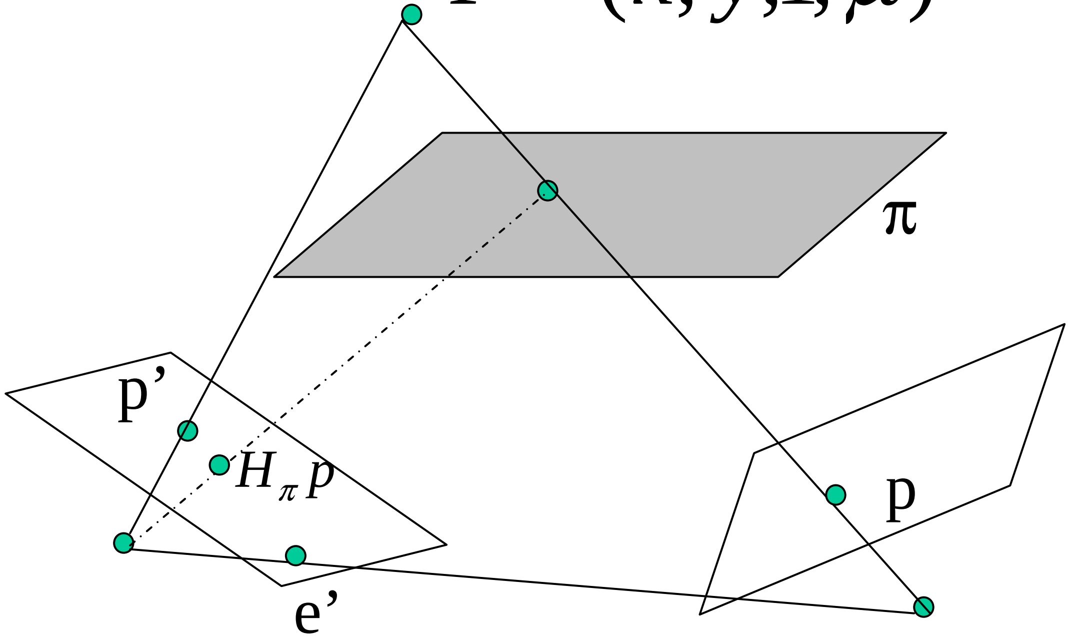
Plane + Parallax

$$p \cong [I,0]P_\pi$$

$$p' \cong \begin{bmatrix} H_\pi & e' \end{bmatrix} P$$

$$\boxed{p' \cong H_\pi p + \mu e'}$$

$$P = (x, y, 1, \mu)$$



- what does μ stand for?
- what would we obtain after eliminating μ

Reminder (from class 1):

$$p \cong [K; \quad 0]P$$

$$p' \cong K'[R \quad t]P$$

$$P = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

$$p' \cong K'RK^{-1}p + \frac{1}{Z}K't$$

$$H_{\infty} = K'RK^{-1}$$

$$e' = K't,$$

$$H_{\pi} \cong K'(R + \frac{1}{d_{\pi}}tn^{T})K^{-1}$$

$$p' \cong K'RK^{-1}p + \frac{1}{Z}K't \qquad H_\pi \cong K'(R + \frac{1}{d_\pi}tn^T)K^{-1}$$

$\Rightarrow$

Recall:

$$p = \frac{1}{Z}K\begin{pmatrix}X\\Y\\Z\end{pmatrix}$$

$$p' \cong H_\pi p - \frac{1}{d_\pi}e'n^TK^{-1}p + \frac{1}{Z}e'$$

$$= H_\pi p + \left[\frac{d_\pi - n^T(ZK^{-1}p)}{Zd_\pi}\right]e'$$

Let: $d = d_\pi - n^T(ZK^{-1}p)$

$$\boxed{p' \cong H_\pi p + \left[\frac{d}{Zd_\pi}\right]e'} \qquad \boxed{p' \cong H_\pi p + \mu e'}$$

Note that e', H_π are determined (each) up to a scale.

⇒

Let $p_0, p_0^{'}$ Be any “reference” point not arising from π

$$p_0^{'} \cong H_\pi p_0 + \mu_0 e'$$

$$p_0^{'} \cong \frac{1}{\mu_0} H_\pi p_0 + e'$$

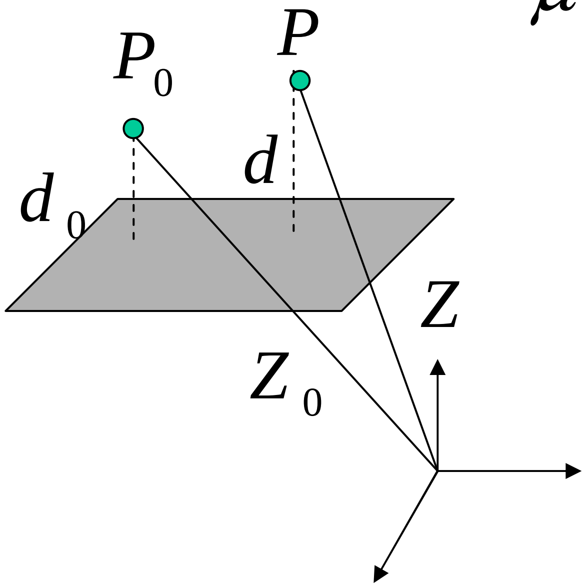
Let $\frac{1}{\mu_0} H_\pi$ be the homography we will use

$$p' \cong \left(\frac{1}{\mu_0} H_\pi \right) p + \frac{\mu}{\mu_0} e'$$

Recall:

$$\mu = \frac{d}{Zd_\pi}$$

$$\frac{\mu}{\mu_0} = \frac{Z_0}{Z} \frac{d}{d_0}$$

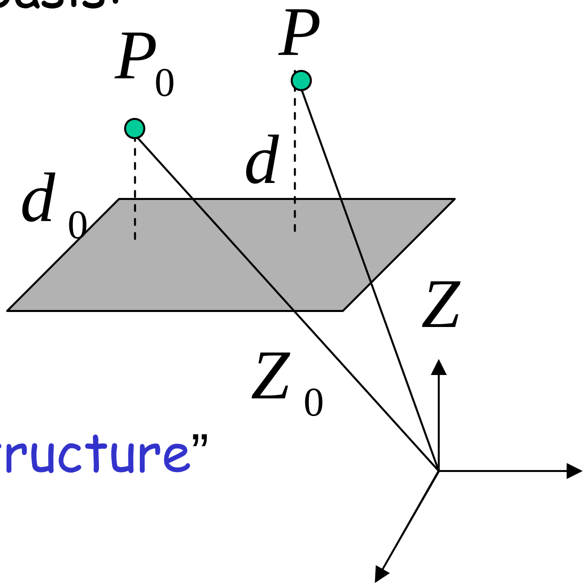


Plane + Parallax

$$\boxed{p' \cong H_{\pi} p + \mu e'}$$

$$\mu = \frac{Z_0}{Z} \frac{d}{d_0}$$

We have used 4 space points for a basis:
3 for the reference plane
1 for the reference point (scaling)



➡ Since 4 points determine an affine basis:

μ is called “relative affine structure”

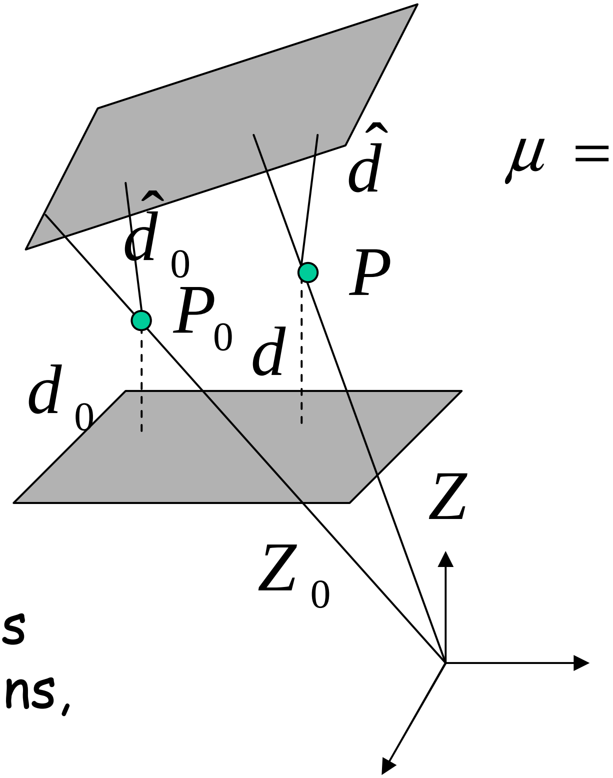
Note: we need 5 points for a projective basis. The 5th point is the first camera center.

Note: A projective invariant

$$p' \cong H_{\pi} p + \mu e'$$

$$p' \cong H_{\hat{\pi}} p + \hat{\mu} e'$$

$$\frac{\mu}{\hat{\mu}} = \frac{\hat{d}_0}{\hat{d}} \frac{d}{d_0}$$



$$\mu = \frac{Z_0}{Z} \frac{d}{d_0}$$

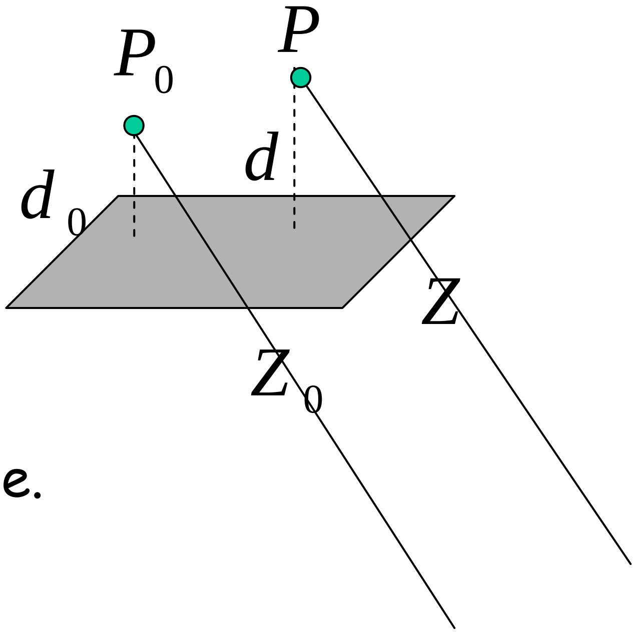
This invariant (“projective depth”) is independent of both camera positions, therefore is projective.

5 basis points: 4 non-coplanar defines two planes, and A 5th point for scaling.

Note: An Affine Invariant

What happens when camera center is at infinity? (parallel projection)

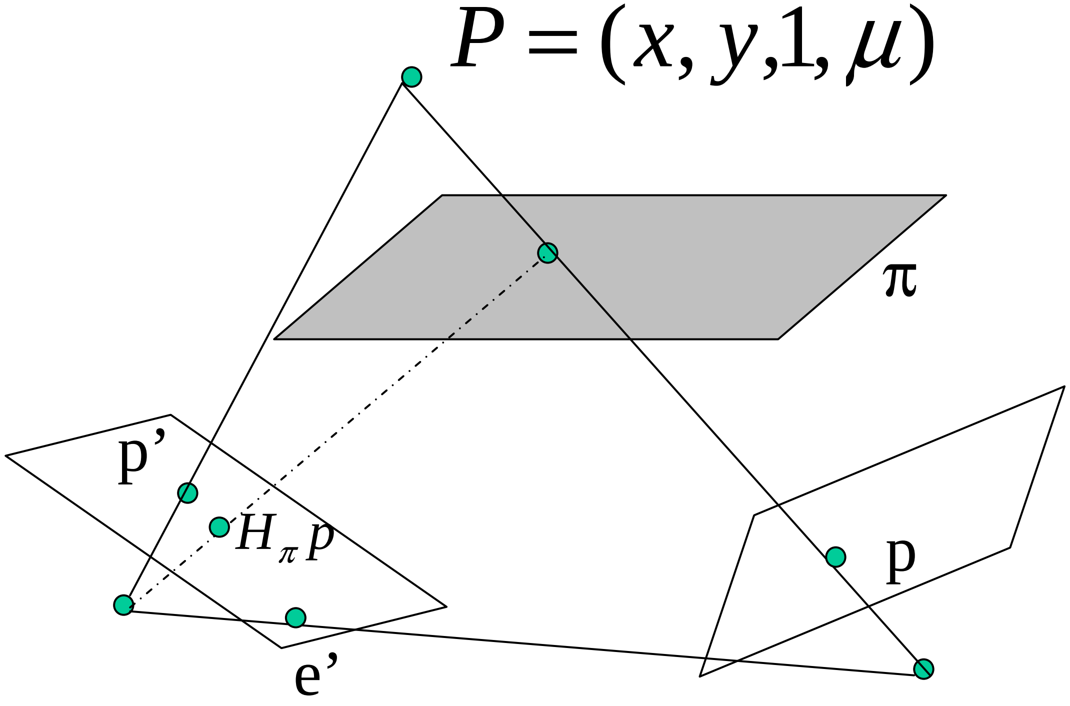
$$\mu = \frac{Z_0}{Z}\frac{d}{d_0} \xrightarrow[Z, Z_0 \to \infty]{} \frac{d}{d_0}$$



This invariant is independent of both camera positions, and is Affine.

Fundamental Matrix

$$\boxed{p' \cong H_\pi p + \mu e'}$$



$$rank\begin{bmatrix} p' & H_\pi p & e' \end{bmatrix} = 2$$

$$\Rightarrow \quad p'^T (e' \times H_\pi p) = 0$$

$$p'^T ([e']_\times H_\pi) p = 0 \qquad\qquad p'^T F p = 0$$

Fundamental Matrix

$$p'^T ([e']_\times H_\pi) p = 0$$

$$p'^T F p = 0$$

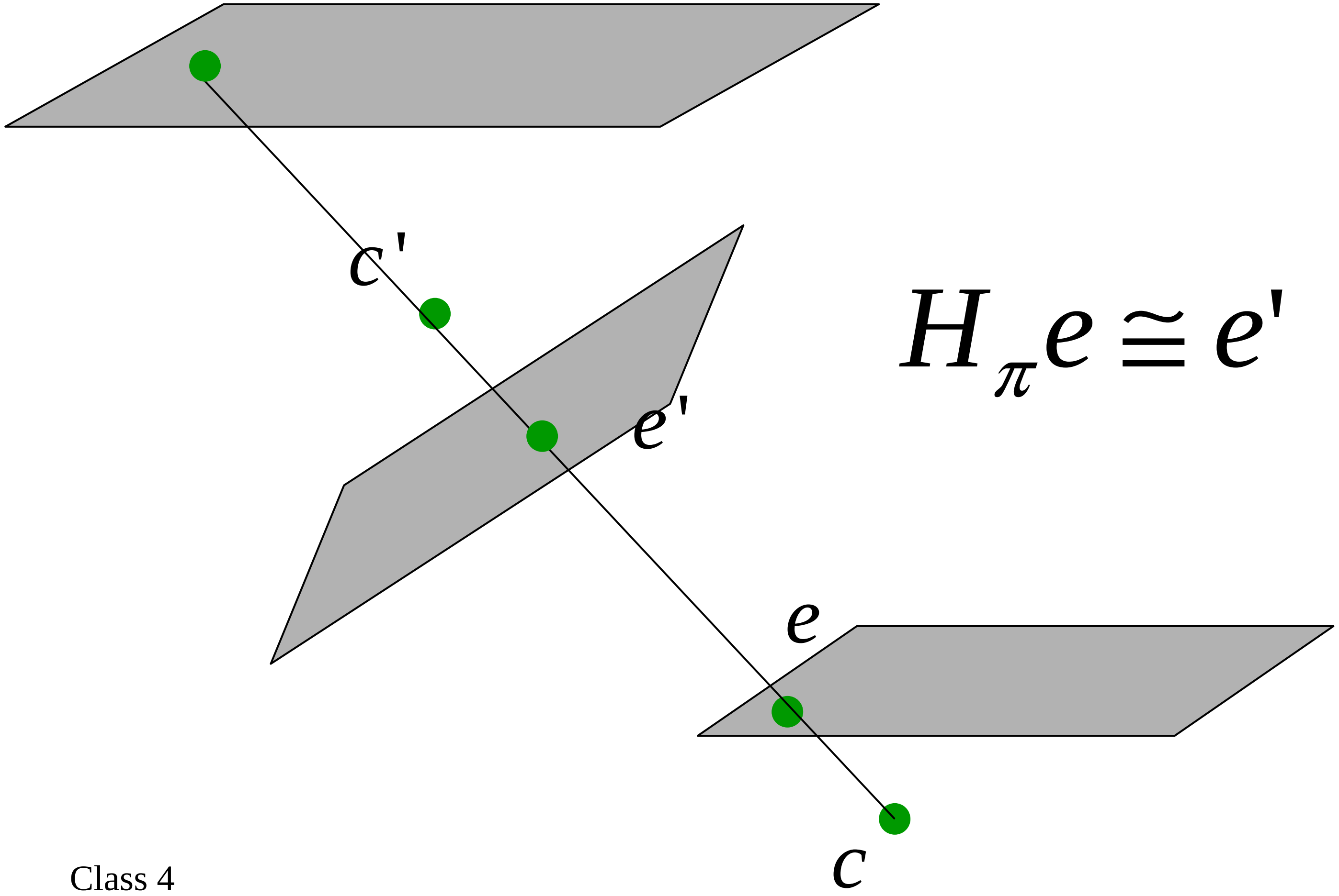
Defines a bilinear matching constraint whose coefficients depend only on the camera geometry (shape was eliminated)

- F does not depend on the choice of the reference plane

$$[e']_\times H_\pi = [e']_\times (\lambda H_\infty + e' n^T) \cong [e']_\times H_\infty$$

Epipoles from F

Note: any homography matrix maps between epipoles:



$$H_\pi e \cong e'$$

Epipoles from F

$$Fe = 0$$

$$[e']_\times H_\infty e \cong [e']_\times e' = 0$$

$$F^T e' = 0$$

$$-H_\infty{}^T [e']_\times e' = 0$$

Estimating F from matching points

$p_i'^T F p_i = 0$ $\quad i = 1,\ldots,8$ Linear solution

$p_i'^T F p_i = 0$ $\quad i = 1,\ldots,7$ N on-linear solution

$\det(F) = 0$

$\det(F) = 0$ is cubic in the elements of F, thus we should expect 3 solutions.

Estimating F from Homographies

$H_\pi^T F$ is skew-symmetric (i.e. provides 6 constraints on F)

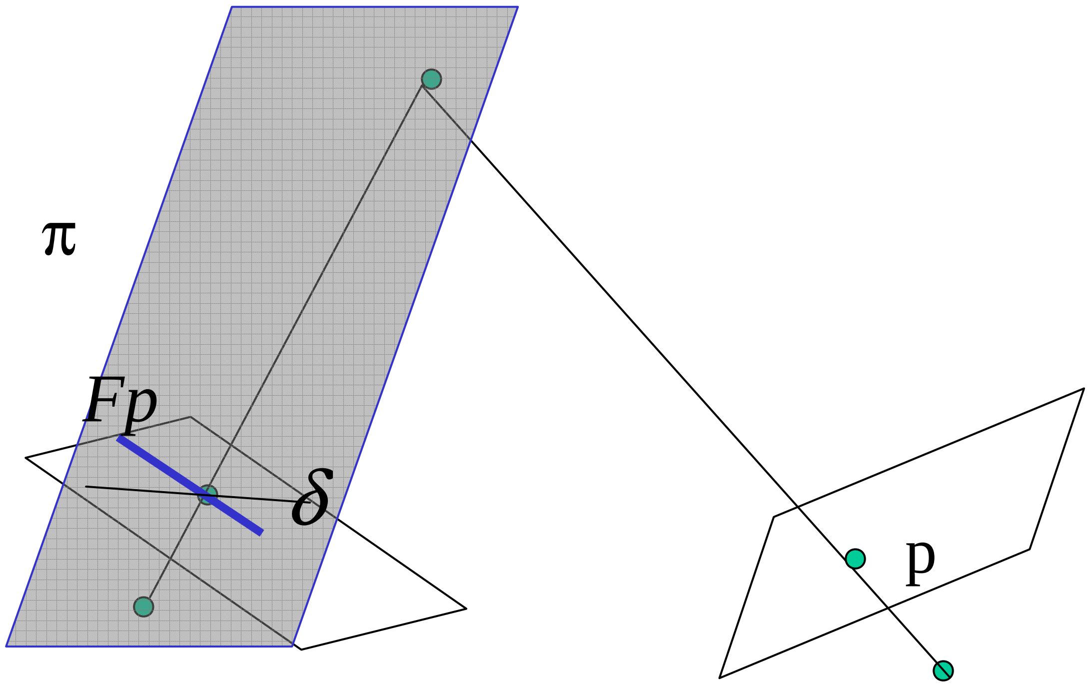
$$H_\pi^T F = (\lambda H_\infty + e' n^T)^T [e']_\times H_\infty = \lambda H_\infty^T [e']_\times H_\infty$$

$$F^T H_\pi = -H_\infty^T [e']_\times (\lambda H_\infty + e' n^T) = -\lambda H_\infty^T [e']_\times H_\infty$$

⇒ $$H_\pi^T F = -F^T H_\pi$$

⇒ 2 homography matrices are required for a solution for F

F Induces a Homography



$[\delta]_\times F$ is a homography matrix induced by the plane defined by the join of the image line δ and the camera center

Projective Reconstruction

1. Solve for F via the system $p_i'^T F p_i = 0$ (8 points or 7 points)

2. Solve for e’ via the system $F^T e' = 0$

3. Select an arbitrary vector δ $\delta^T e' \neq 0$

4. $[I \quad 0]$ and $[[\delta]_\times F \quad e']$ are a pair of camera matrices.

$$\boxed{p' \cong [\delta]_\times Fp + \mu e'}$$

Trifocal Geometry

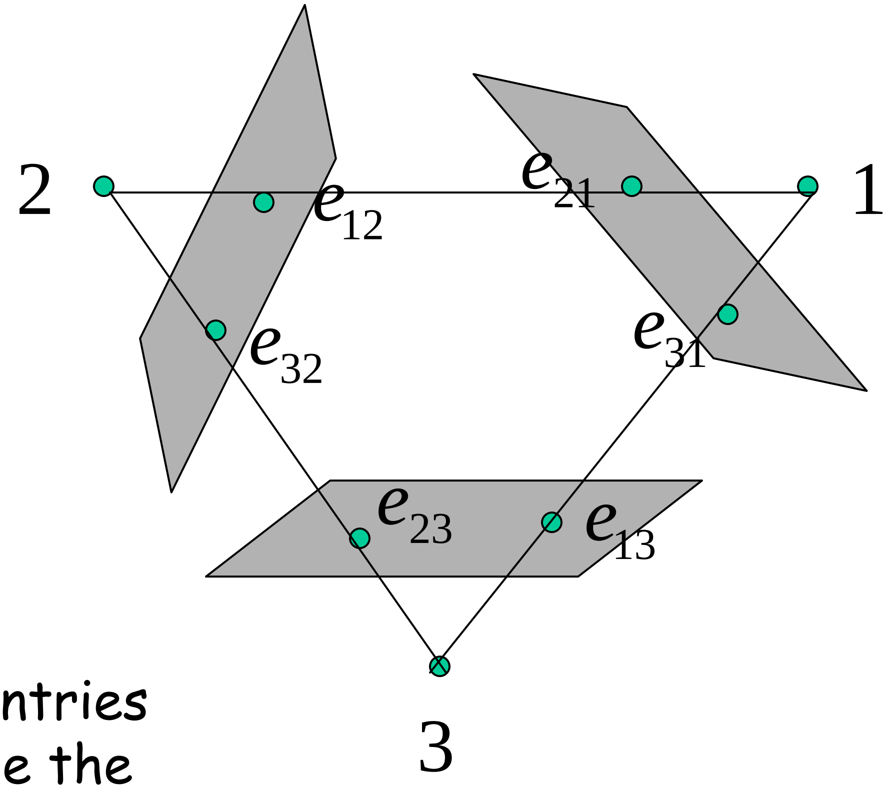
The three fundamental matrices completely describe the trifocal geometry (as long as the three camera centers are not collinear)

$$F_{12}e_{31} = e_{12} \times e_{32}$$

$$\Rightarrow \quad e_{32}^T F_{12} e_{31} = 0$$

Likewise: $$e_{13}^T F_{23} e_{12} = 0$$

$$e_{23}^T F_{13} e_{21} = 0$$



Each constraint is non-linear in the entries of the fundamental matrices (because the epipoles are the respective null spaces)

Trifocal Geometry

$$e_{32}^{T} F_{12} e_{31} = 0$$

$$e_{13}^{T} F_{23} e_{12} = 0$$

$$e_{23}^{T} F_{13} e_{21} = 0$$

3 fundamental matrices provide 21 parameters. Subtract 3 constraints, Thus we have that the trifocal geometry is determined by 18 parameters.

This is consistent with the straight-forward counting:

3x11 – 15 = 18

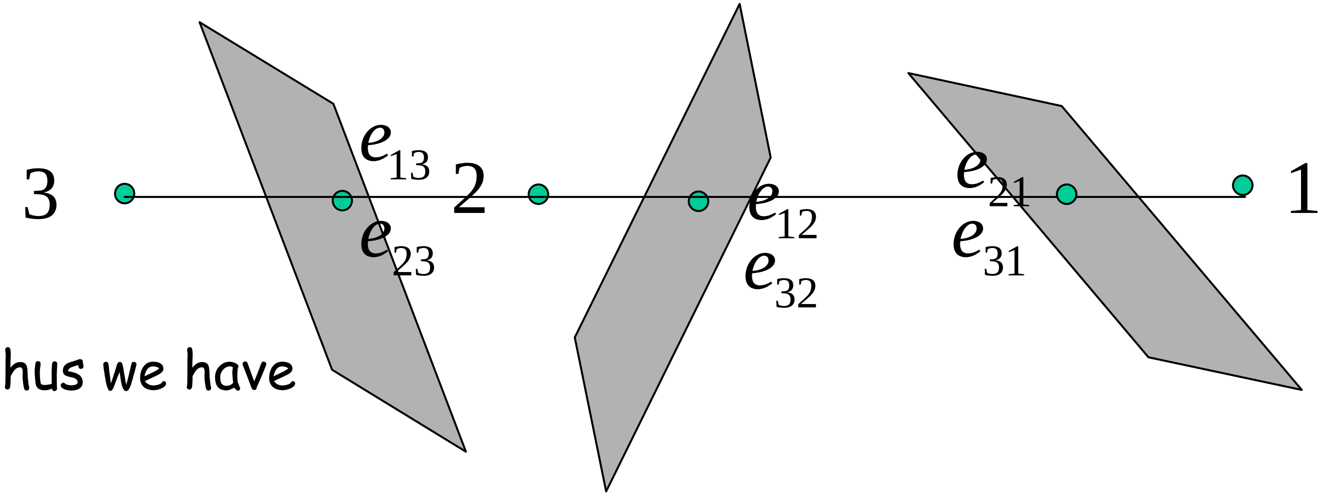
(3 camera matrices provide 33 parameters, minus the projective basis)

What Goes Wrong with 3 views?

$e_{31} \cong e_{21}$

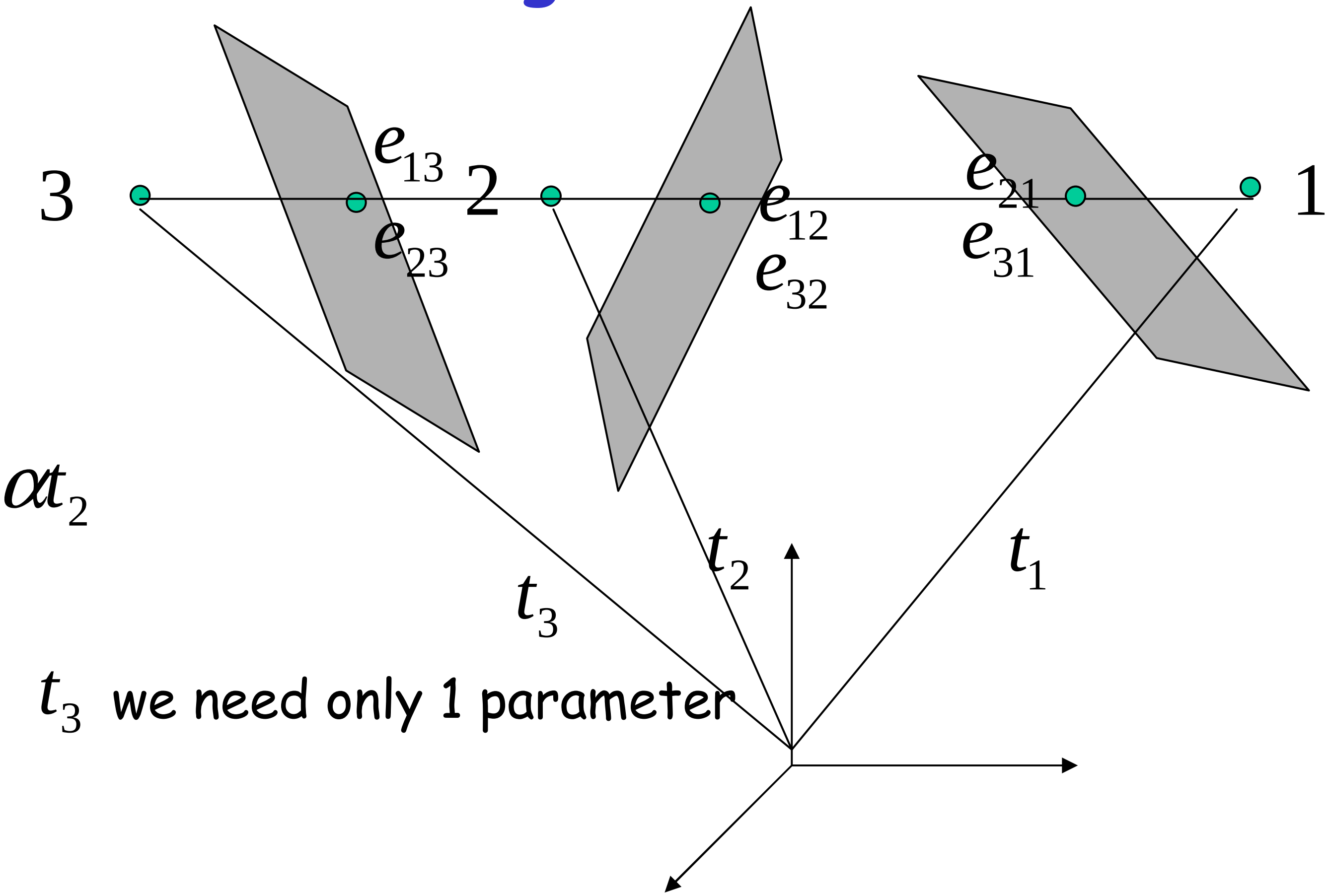
$e_{12} \cong e_{32}$

$e_{13} \cong e_{23}$



2 constraints each, thus we have 21-6=15 parameters

What Goes Wrong with 3 views?



$$t_3 = t_1 + \alpha t_2$$

Thus, to represent t_3 we need only 1 parameter (instead of 3).

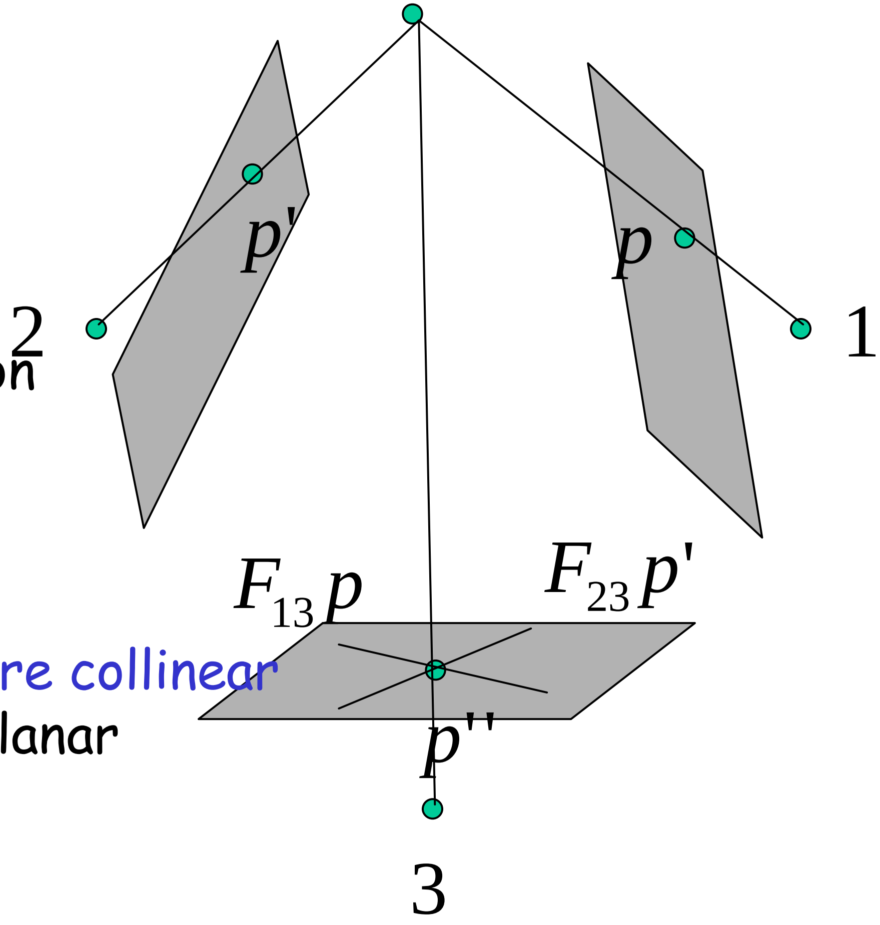
18-2=16 parameters are needed to represent the trifocal geometry in this case.

but the pairwise fundamental matrices can account for only 15!

What Else Goes Wrong: Reprojection

$$p'' \cong F_{13} p \times F_{23} p'$$

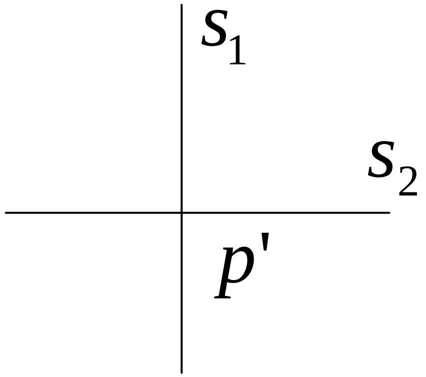
Given p,p' and the pairwise F-mats one can directly determine the position of the matching point p''



This fails when the 3 camera centers are collinear because all three line of sights are coplanar thus there is only one epipolar line!

The Trifocal Constraints

$$p = \begin{bmatrix} I & 0 \end{bmatrix} P \qquad p' \cong \begin{bmatrix} A & e' \end{bmatrix} P \qquad p'' \cong \begin{bmatrix} B & e'' \end{bmatrix} P$$



$$s_1 = \begin{pmatrix} -1 \\ 0 \\ x' \end{pmatrix} \quad s_2 = \begin{pmatrix} 0 \\ -1 \\ y' \end{pmatrix} \Rightarrow \begin{matrix} s_1^T p' = 0 \\ s_2^T p' = 0 \end{matrix}$$

$$r_1 = \begin{pmatrix} -1 \\ 0 \\ x'' \end{pmatrix} \quad r_2 = \begin{pmatrix} 0 \\ -1 \\ y'' \end{pmatrix} \Rightarrow \begin{matrix} r_1^T p'' = 0 \\ r_2^T p'' = 0 \end{matrix}$$

The Trifocal Constraints

$$s_1^T p' = 0$$

$$s_2^T p' = 0$$

$$p' \cong \begin{bmatrix} A & e' \end{bmatrix} P \Rightarrow \begin{matrix} s_1^T \begin{bmatrix} A & e' \end{bmatrix} P = 0 \\ s_2^T \begin{bmatrix} A & e' \end{bmatrix} P = 0 \end{matrix}$$

$$r_1^T p'' = 0$$

$$r_2^T p'' = 0$$

$$p'' \cong \begin{bmatrix} B & e'' \end{bmatrix} P \Rightarrow \begin{matrix} r_1^T \begin{bmatrix} B & e'' \end{bmatrix} P = 0 \\ r_2^T \begin{bmatrix} B & e'' \end{bmatrix} P = 0 \end{matrix}$$

$$p = \begin{bmatrix} I & 0 \end{bmatrix} P \Rightarrow \begin{matrix} \begin{pmatrix} -1 & 0 & x & 0 \end{pmatrix} P = 0 \\ \begin{pmatrix} 0 & -1 & y & 0 \end{pmatrix} P = 0 \end{matrix}$$

The Trifocal Constraints

$$\begin{bmatrix} -1 & 0 & x & 0 \\ 0 & -1 & y & 0 \\ \multicolumn{3}{c}{s_1^T A} & s_1^T e' \\ \multicolumn{3}{c}{s_2^T A} & s_2^T e' \\ \multicolumn{3}{c}{r_1^T B} & r_1^T e'' \\ \multicolumn{3}{c}{r_2^T B} & r_2^T e'' \end{bmatrix}_{6\times 4} P = 0$$

Every 4x4 minor must vanish!

12 of those involve all 3 views, they are arranged in 3 groups
Depending on which view is the reference view.

The Trifocal Constraints

$$\begin{bmatrix} -1 & 0 & x & 0 \\ 0 & -1 & y & 0 \\ & s_1^T A & & s_1^T e' \\ & s_2^T A & & s_2^T e' \\ & r_1^T B & & r_1^T e'' \\ & r_2^T B & & r_2^T e'' \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}0\\0\end{matrix}}\right\} \text{The reference view} \\ \left.\vphantom{\begin{matrix}0\\0\end{matrix}}\right\} \text{Choose 1 row from here} \\ \left.\vphantom{\begin{matrix}0\\0\end{matrix}}\right\} \text{Choose 1 row from here} \end{matrix}$$

We should expect to have 4 matching constraints $f_i(p, p', p'') = 0$

The Trifocal Constraints

Expanding the determinants:

$$p' \cong Ap + \mu e' \Rightarrow s_i^T Ap + \mu s_i^T e' = 0 \qquad i = 1,2$$

$$p'' \cong Bp + \mu e'' \Rightarrow r_j^T Bp + \mu r_j^T e'' = 0 \qquad j = 1,2$$

eliminate μ

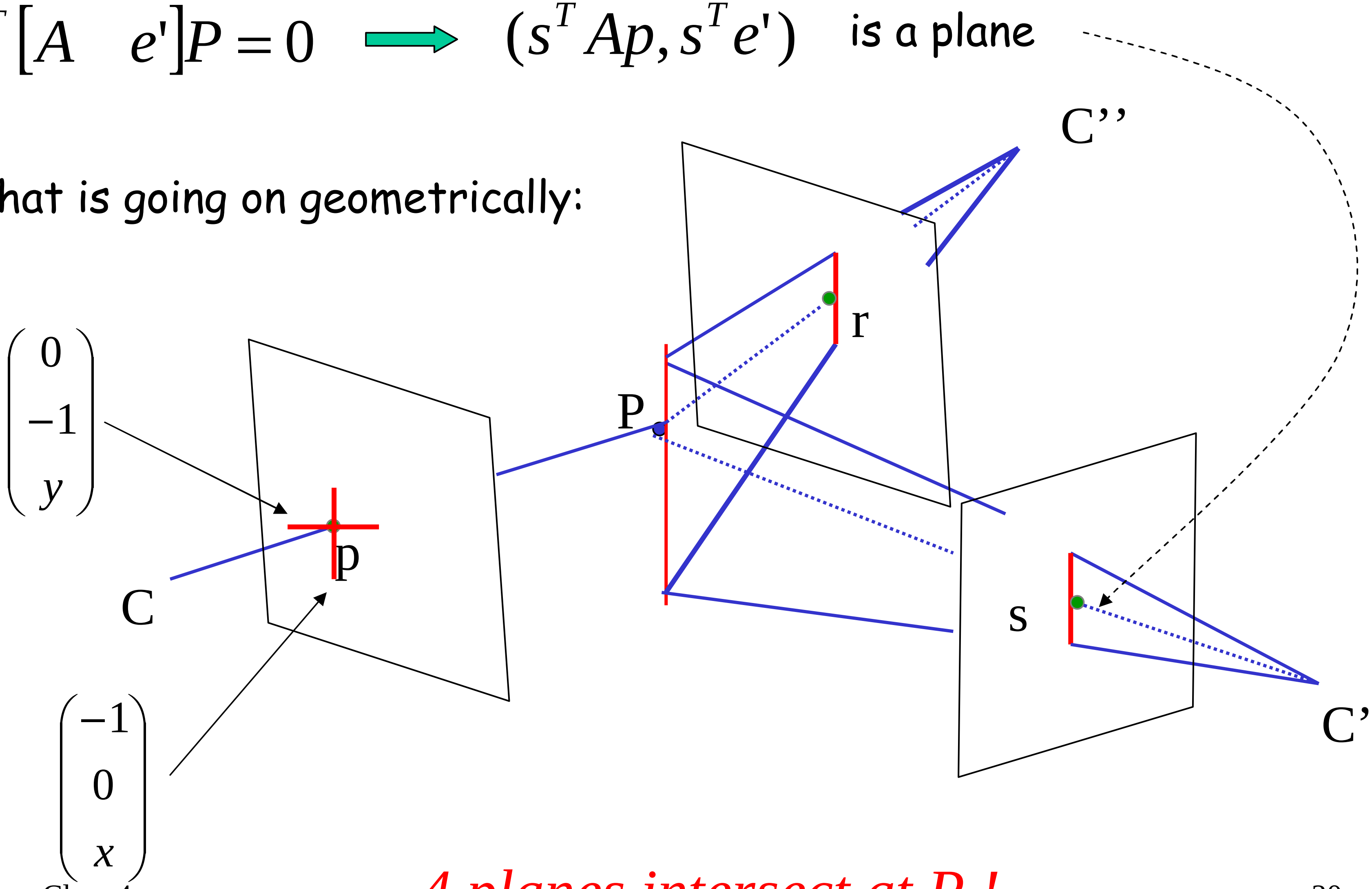
$$\frac{s_i^T Ap}{s_i^T e'} = \frac{r_j^T Bp}{r_j^T e''}$$

$$\Rightarrow \boxed{(r_j^T e'')(s_i^T Ap) = (s_i^T e')(r_j^T Bp)} \quad i, j = 1,2$$

The Trifocal Constraints

$s^T[A \quad e']P = 0 \implies (s^T Ap, s^T e')$ is a plane

What is going on geometrically:



4 planes intersect at P !

The Trifocal Tensor

$$\boxed{(r_j^T e'')(s_i^T Ap) = (s_i^T e')(r_j^T Bp)}$$

New index notations: i-image 1, j-image 2, k-image 3

$$s^T Ap + \mu s^T e' = 0 \quad \Longrightarrow \quad s_j a_i^j p^i + \mu s_j e'^j = 0$$

p^i is a point in image 1

s_j is a line in image 2

e'^j is a point in image 2

The Trifocal Tensor

s_j^l $l = 1,2$ are the two lines coincident with p', i.e. $s_j^l p'^j = 0$

r_k^m $m = 1,2$ are the two lines coincident with p'', i.e. $r_k^m p''^k = 0$

$$s_j^l a_i^j p^i + \mu s_j^l e'^j = 0$$

$$r_k^m b_i^k p^i + \mu r_k^m e''^k = 0$$

Eliminate μ

$$(s_j^l e'^j)(r_k^m b_i^k p^i) - (r_k^m e''^k)(s_j^l a_i^j p^i) = 0$$

The Trifocal Tensor

$$(s_j^l e'^j)(r_k^m b_i^k p^i) - (r_k^m e''^k)(s_j^l a_i^j p^i) = 0$$

Rearrange terms:

$$p^i s_j^l r_k^m (e'^j b_i^k - e''^k a_i^j) = 0 \qquad l, m = 1,2$$

The trifocal tensor is:

$$T_i^{jk} = e'^j b_i^k - e''^k a_i^j$$

The Trifocal Tensor

$$p^i s_j^l r_k^m T_i^{jk} = 0$$

$$s^l{}_j = \begin{bmatrix} -1 & 0 & x' \\ 0 & -1 & y' \end{bmatrix} \qquad r^m{}_k = \begin{bmatrix} -1 & 0 & x'' \\ 0 & -1 & y'' \end{bmatrix}$$

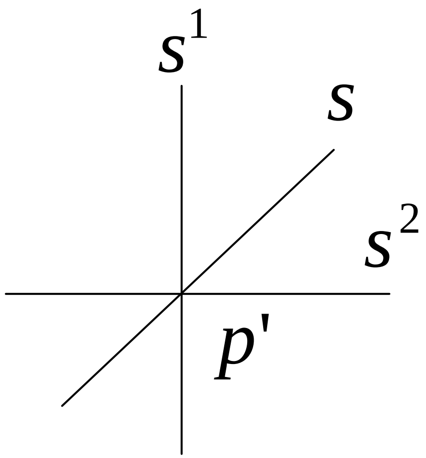
The four "trilinearities":

$$x'' T_i^{13} p^i - x''x' T_i^{33} p^i + x' T_i^{31} p^i - T_i^{11} p^i = 0$$
$$y'' T_i^{13} p^i - y''x' T_i^{33} p^i + x' T_i^{32} p^i - T_i^{12} p^i = 0$$
$$x'' T_i^{23} p^i - x''y' T_i^{33} p^i + y' T_i^{31} p^i - T_i^{21} p^i = 0$$
$$y'' T_i^{23} p^i - y''y' T_i^{33} p^i + x' T_i^{32} p^i - T_i^{22} p^i = 0$$

The Trifocal Tensor

$$s_j = \alpha s_j^1 + \beta s_j^2$$

$$r_k = \gamma r_k^1 + \delta r_k^2$$



$$p^i s_j r_k T_i^{jk} = p^i (\alpha s_j^1 + \beta s_j^2)(\gamma r_k^1 + \delta r_k^2) T_i^{jk} = 0$$

A trilinearity is a contraction with a point-line-line where the lines are coincident with the respective matching points.

Slices of the Trifocal Tensor

Now that we have an explicit form of the tensor, what can we do with it?

$$p^i s_j T_i^{jk} = ?$$

The result must be a contravariant vector (a point). This point is coincident with r for all lines coincident with p''

➡

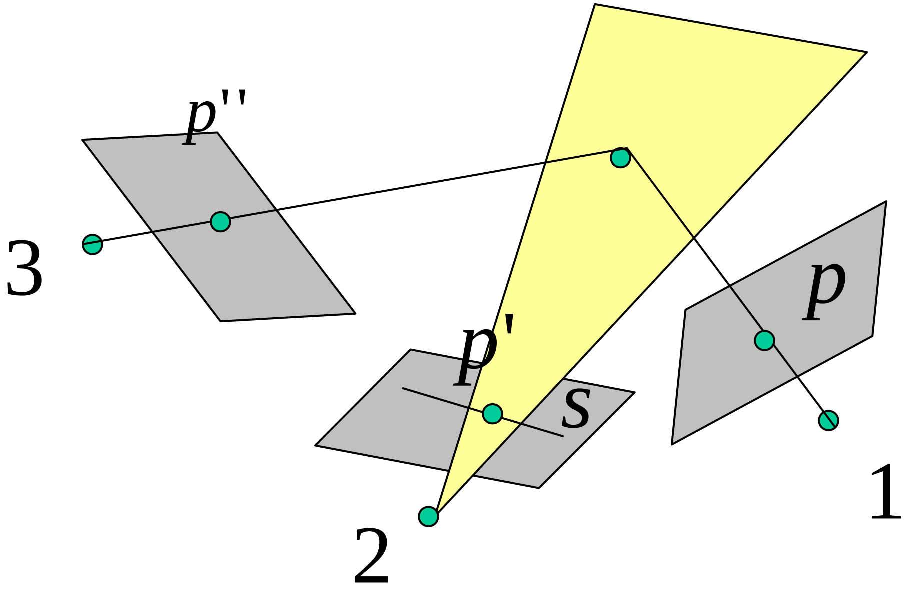
$$p^i s_j T_i^{jk} \cong p''^k \qquad s \neq e' \times p'$$

The point reprojection equation (will work when camera centers are collinear as well).

Note: reprojection is possible after observing 7 matching points, (because one needs 7 matching triplets to solve for the tensor). This is in contrast to reprojection using pairwise fundamental matrices Which requires 8 matching points (in order to solve for the F-mats).

Slices of the Trifocal Tensor

$$p^i s_j T_i^{jk} \cong p''^k$$



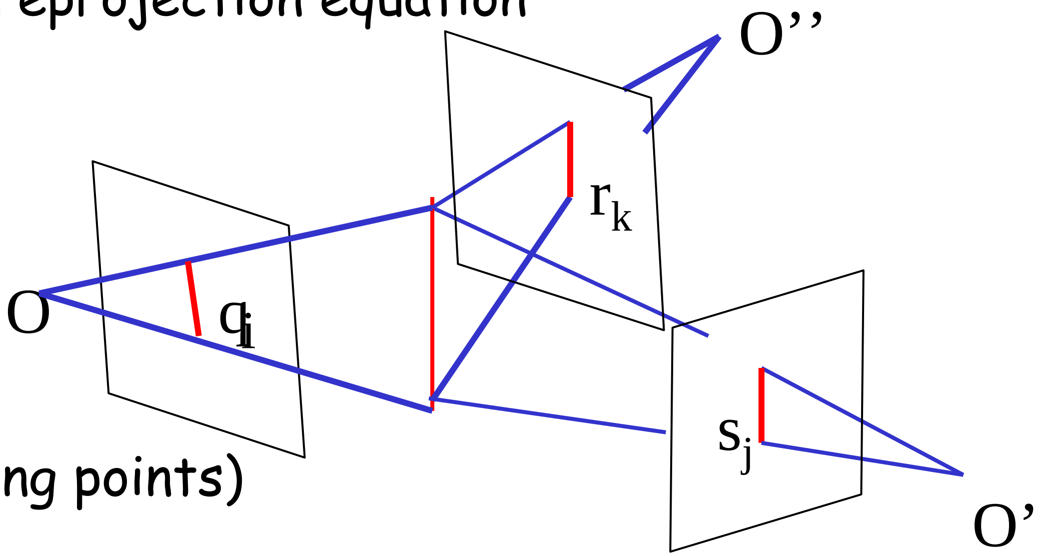
Slices of the Trifocal Tensor

$$s_j r_k T_i^{jk} = ?$$

The result must be a line.

$s_j r_k T_i^{jk} \cong q_i$ Line reprojection equation

13 matching lines are necessary for solving for the tensor (compared to 7 matching points)



Slices of the Trifocal Tensor

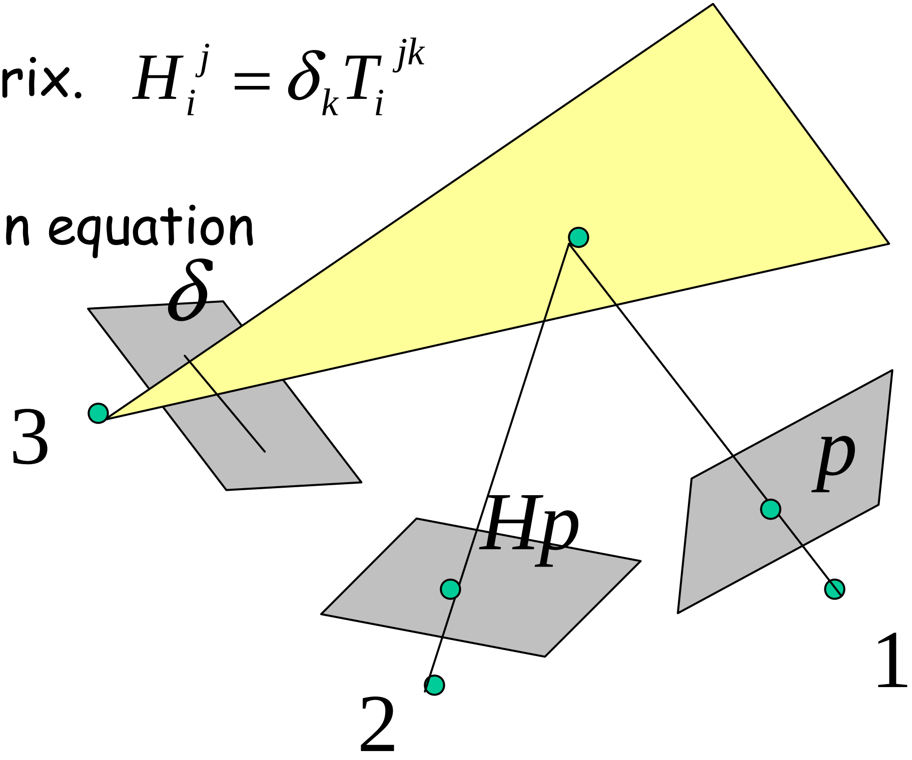
$$\delta_k T_i^{jk} = ?$$

The result must be a matrix. $H_i^j = \delta_k T_i^{jk}$

$p^i \delta_k T_i^{jk}$ is the reprojection equation

→

H is a homography matrix



$$\delta_k T_i^{jk} \quad \forall \delta$$

is a family of homography matrices (from 1 to 2) induced by the family of planes coincidant with the 3rd camera center.

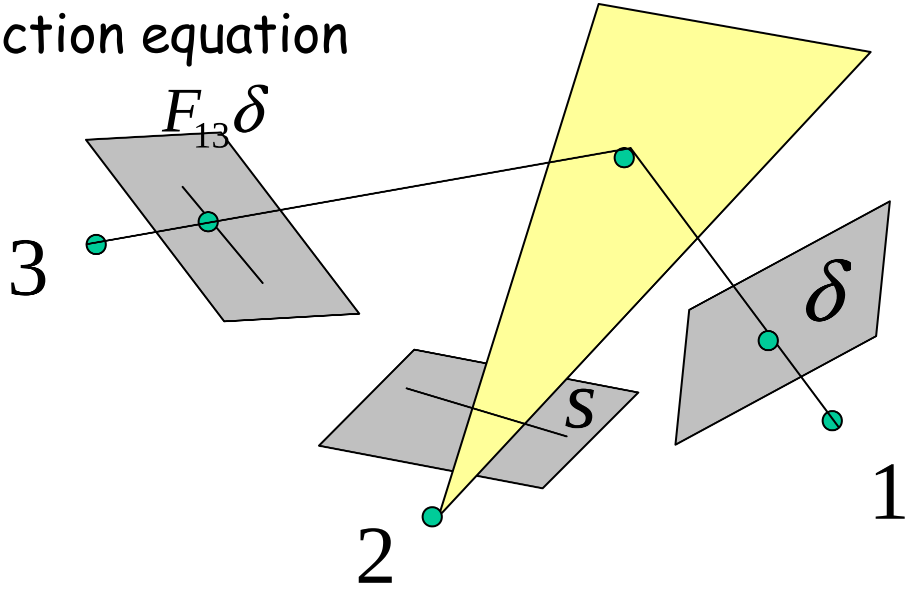
Slices of the Trifocal Tensor

$\delta_j T_i^{jk}$ is the homography matrix from 1 to 3 induced by the plane defined by the image line δ and the second camera center.

$$\delta^i T_i^{jk} = ?$$

$\delta^i s_j T_i^{jk}$ is the reprojection equation

The result is a point on the epipolar line of δ on image 3



Slices of the Trifocal Tensor

$$\delta^i T_i^{jk} = G^{jk}$$

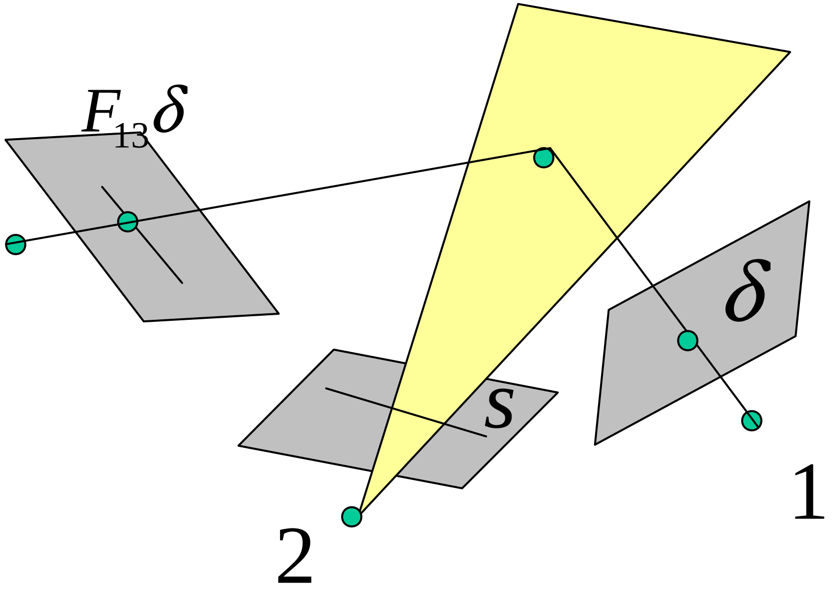
Gs Is a point on the epipolar line $F_{13}\delta$

⇒ $rank(G) = 2$

(because it maps the dual plane onto collinear points)

$$null(G) = F_{13}\delta$$

$$null(G^T) = F_{12}\delta$$



18 Parameters for the Trifocal Tensor

$$T_i^{\ jk} = e'^j b_i^k - e''^k a_i^j$$

$$e'^j (b_i^k + n_i e''^k) - e''^k (a_i^j + n_i e'^j) \qquad \forall n$$

$$= T_i^{\ jk} + n_i e'^j e''^k - n_i e'^j e''^k$$

$$= T_i^{\ jk}$$

$T_i^{\ jk}$ Has 24 parameters (9+9+3+3)
minus 1 for global scale
minus 2 for scaling e',e'' to be unit vectors
minus 3 for setting n_i such that B has a vanishing column

= 18 independent parameters

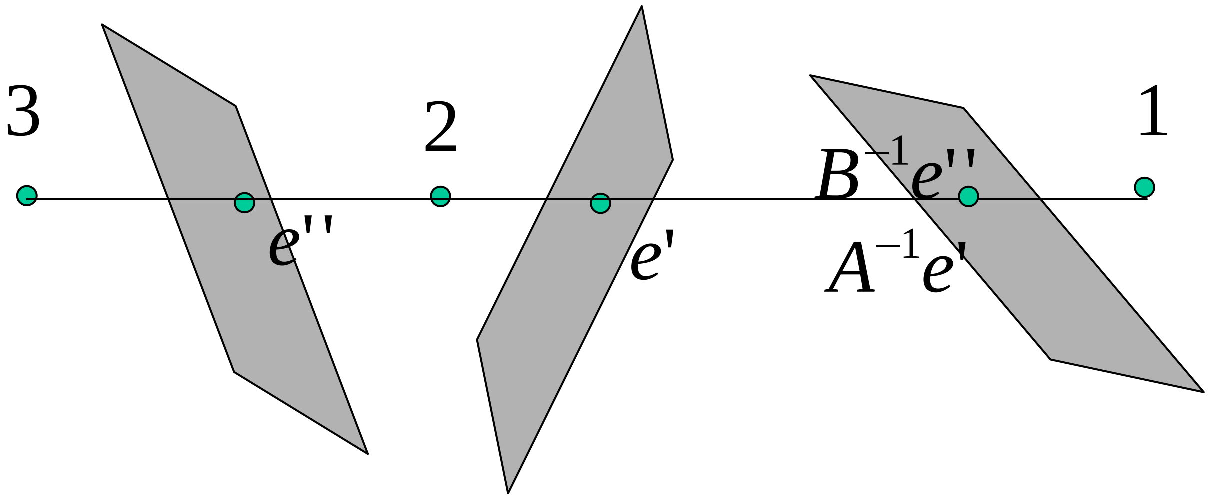
⇒ We should expect to find 9 non-linear constraints among the 27 entries of the tensor (admissibility constraints).

18 Parameters for the Trifocal Tensor

What happens when the 3 camera centers are collinear?

(we saw that pairwise F-mats account for 15 parameters).

$$A^{-1}e' \cong B^{-1}e''$$



This provides two additional (non-linear) constraints, thus 18-2=16.

Items not Covered in Class

- Degenerate configurations (Linear Line Complex, Quartic Curve)
- The source of the 9 admissibility constraints (come from the homography slices).
- Concatenation of trifocal tensors along a sequence
- Quadrifocal tensor (and its relation to the homography tensor)