STEREOPSIS

PART III - RECONSTRUCTION

computer vision

STEREO VISION PIPELINE

- extract interesting points from an image pair (eg corners)
- determine a set of matching points
- estimate the epipolar geometry
- rectify the two images
- compute dense stereo matching on the rectified pair
- for each point match derive the 3D point X by triangulation (**reconstruction)**

SPARSE CORRESPONDENCES

- fewer points (~ hundreds)
- often we do not have priors on their position (although if we are using image patches we are implicitly assuming that the corresponding patches haven't changed so much... no large rotations or large perspective variations...)

FEATURES ADJACENCY MATRIX

- given a set of features the two images

$$\mathcal{F}_{I_1} = \{f_i\}_{i=1}^{N} = \{(x1_i, y1_i, N1_i)\}_{i=1}^{N}$$

$$\mathcal{G}_{I_2} = \{g_j\}_{j=1}^{M} = \{(x2_j, y2_j, N2_j)\}_{j=1}^{M}$$

positions

image patch centered at ($x2_j$,$y2_j$)

- we may compute an adjacency matrix as

$$E(i,j) = e^{-||\mathbf{x1}_i - \mathbf{x2}_j||^2 / 2\sigma^2}$$

in terms of euclidean distance

- E is a NxM matrix where each entry measures how *close* feature **fi** in image I**1** is to feature g**j** in image I**2**

- σ controls the maximum distance among features we are willing to consider

FEATURES AFFINITY MATRIX

- E has values in [0,1]   identical features correspond to 1s

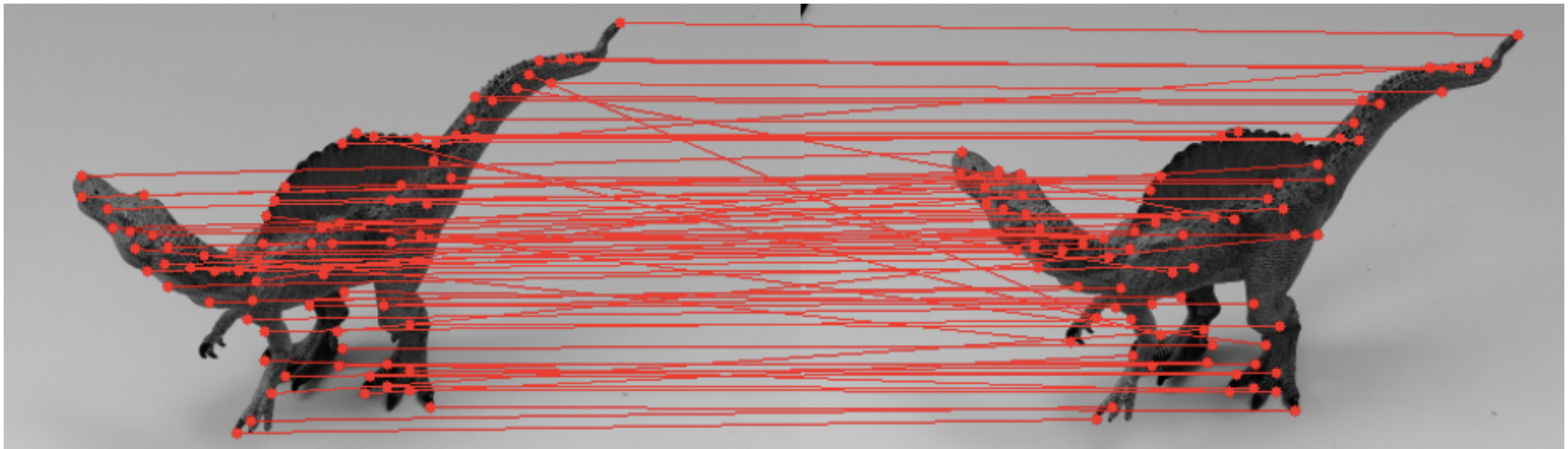
- to take into account appearance similarity too we combine the adjacency matrix E with a measure of the patches similarity, eg

$$A(i,j) = E(i,j) * \frac{1}{2}(\phi_{NCC}(N1_i, N2_j) + 1)$$

- The affinity matrix A has values in [0,1] too

MATCHING ON THE AFFINITY MATRIX

- we look for one-to-one matches according to the following :
- look for the maximum of each row of A
- check that the element found is also the maximum for the corresponding column of A
  - if so you have found a match!



RECONSTRUCTION AND PRIOR KNOWLEDGE

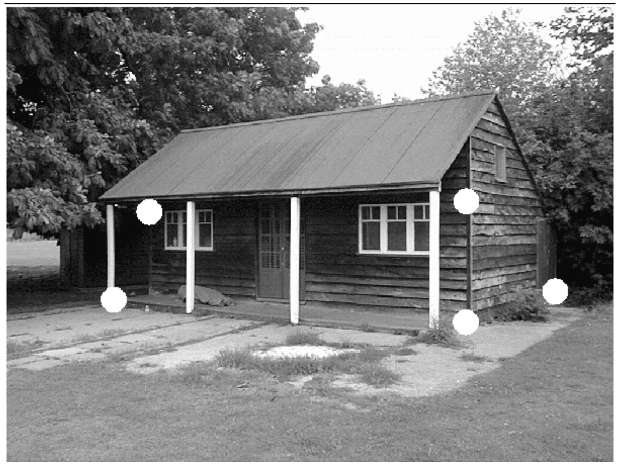
- if we have intrinsic and extrinsic parameters (obtained during a calibration process)
  - **unambiguous (metric) reconstruction**
- *if we have the intrinsic parameters*
  - **up to an unknown scaling factor**

from the correspondences we may obtain the essential matrix

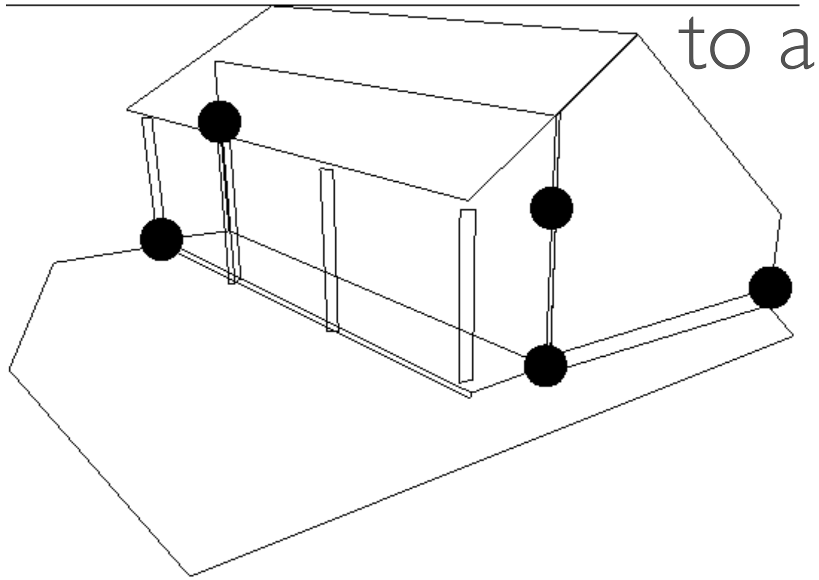
- no information
  - **up to an unknown global projective transformation**

from the correspondences
we may estimate the fundamental matrix

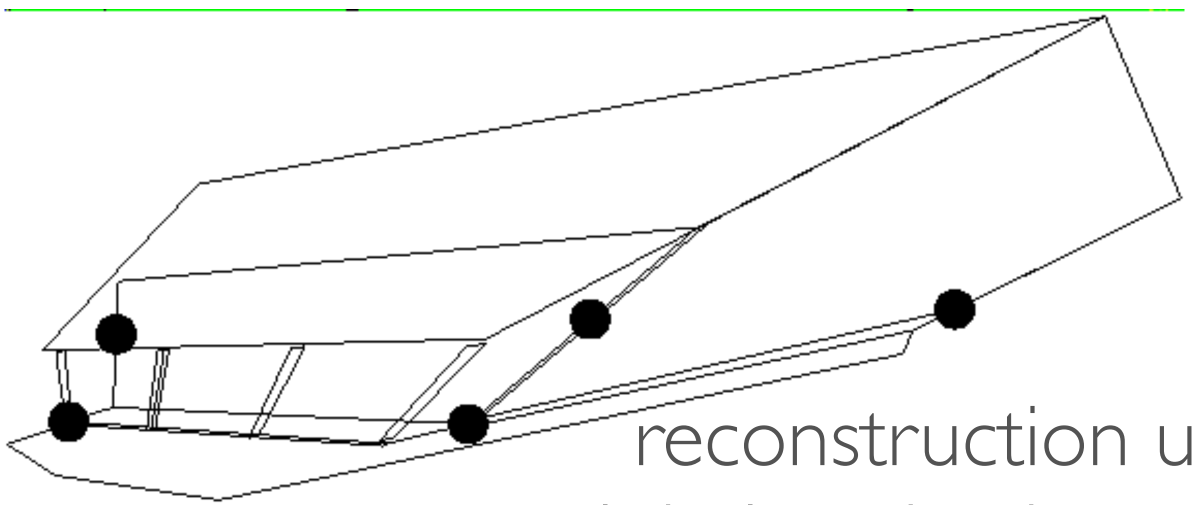
EXAMPLE



the real thing



reconstruction up to a scale



reconstruction up to a global projective transform

RECONSTRUCTION AND PRIOR KNOWLEDGE

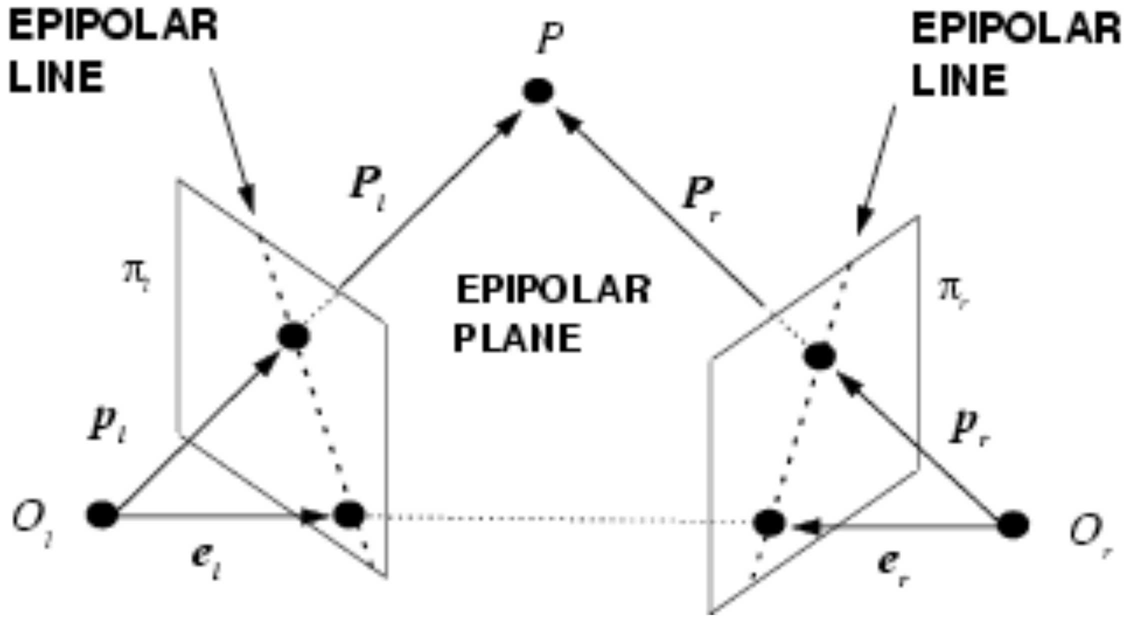
- if we have intrinsic and extrinsic parameters (obtained during a calibration process)
  - unambiguous (metric) reconstruction
- *if we have the intrinsic parameters*
  - *up to an unknown scaling factor*

calibrated reconstruction

- no information
  - up to an unknown global projective transformation

uncalibrated reconstruction

TRIANGULATION: THE IDEAL CASE



TRIANGULATION: OUTLINE OF THE METHOD

- compute the fundamental matrix/essential matrix from point correspondences
- derive the projection matrices from the fundamental matrix/essential matrix
- for each correspondence (x_i, x_i') compute the 3D point that projects to these two image points by triangulating the 2 corresponding optical rays

CALIBRATED CASE: RECOVERY OF THE PROJECTION MATRICES FROM E

- in the calibrated case we may consider points in normalized (mm) coordinates

$$\hat{\mathbf{x}} = K^{-1}\mathbf{x}$$

$$\hat{\mathbf{x}}' = K'^{-1}\mathbf{x}'$$

- then we may formulate the projection matrices as

$$P = [I; \mathbf{0}] \quad P' = [R; \mathbf{t}]$$

- starting from points correspondences $(\hat{\mathbf{x}}, \hat{\mathbf{x}}')$ we may derive the essential matrix E (eg, from the 8 points algorithm)

CALIBRATED CASE: RECOVERY OF THE PROJECTION MATRICES FROM E

- Now we discuss how to factorize E so to recover R and t (and then P')

$$\mathrm{svd}(E) = U\mathrm{diag}(1,1,0)V^\top$$

- then there are two possible factorizations $E = [t]_\times R = SR$ as follows

$$S = UZU^\top$$

$$R = UWV^\top \quad \text{or} \quad R = UW^\top V^\top$$

with

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

CALIBRATED CASE: RECOVERY OF THE PROJECTION MATRICES FROM E

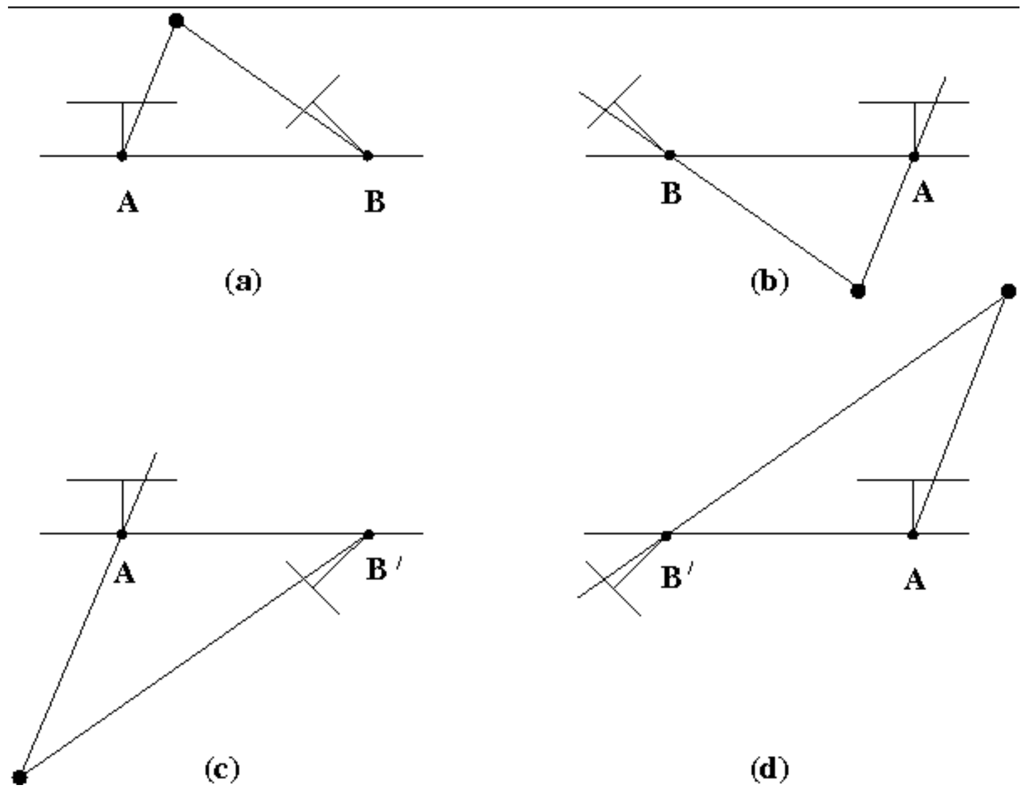
- As for the translation let $[t]_\times = S$ with $||t|| = 1$
- Since $S\mathbf{t} = \mathbf{0}$ then $\mathbf{t} = \mathbf{u}_3$
  but the sign of $\mathbf{t}$ cannot be determined (since the sign of E can't)
- *Summing up:* for a given essential matrix E and a first camera matrix $P = [I; \mathbf{0}]$ there are four possible choices for P':

$$[UWV^\top; \mathbf{u}_3] \quad [UWV^\top; -\mathbf{u}_3] \quad [UW^\top V^\top; \mathbf{u}_3] \quad [UW^\top V^\top; -\mathbf{u}_3]$$

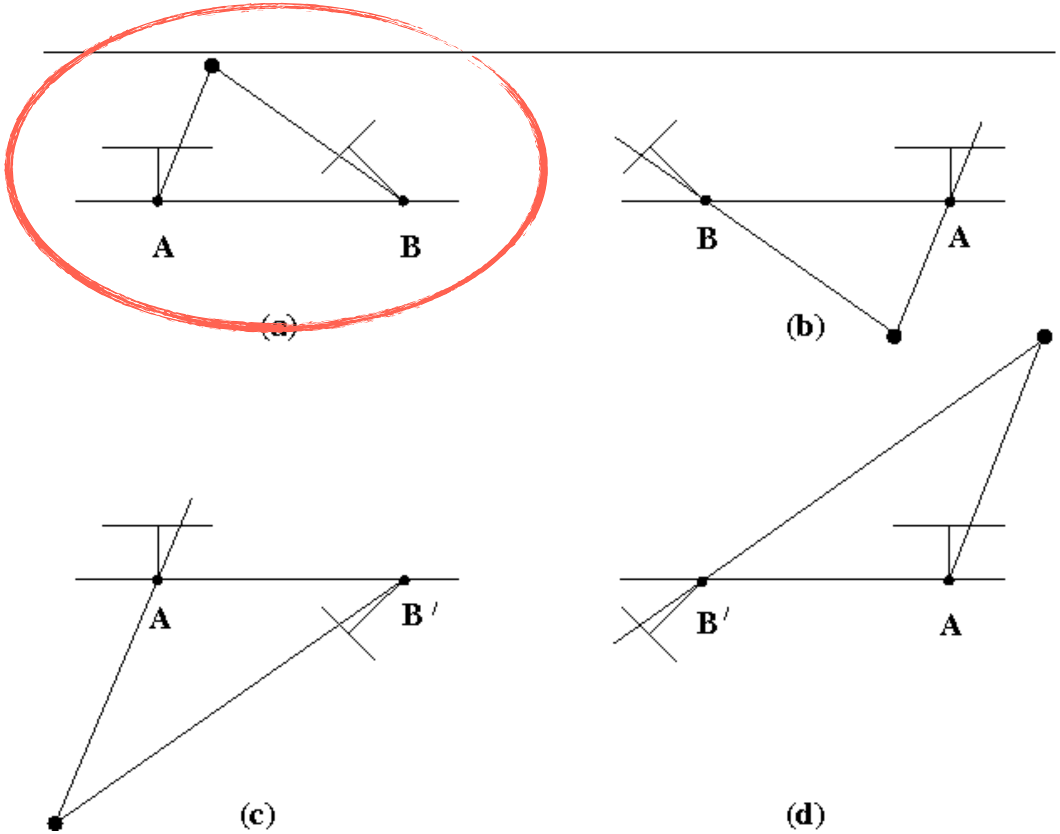
CALIBRATED CASE: RECOVERY OF THE PROJECTION MATRICES FROM E

- *Summing up:* for a given essential matrix E and a first camera matrix P=[I; **0**] there are four possible choices for P':

$$[UWV^\top; \mathbf{u}_3] \quad [UWV^\top; -\mathbf{u}_3] \quad [UW^\top V^\top; \mathbf{u}_3] \quad [UW^\top V^\top; -\mathbf{u}_3]$$



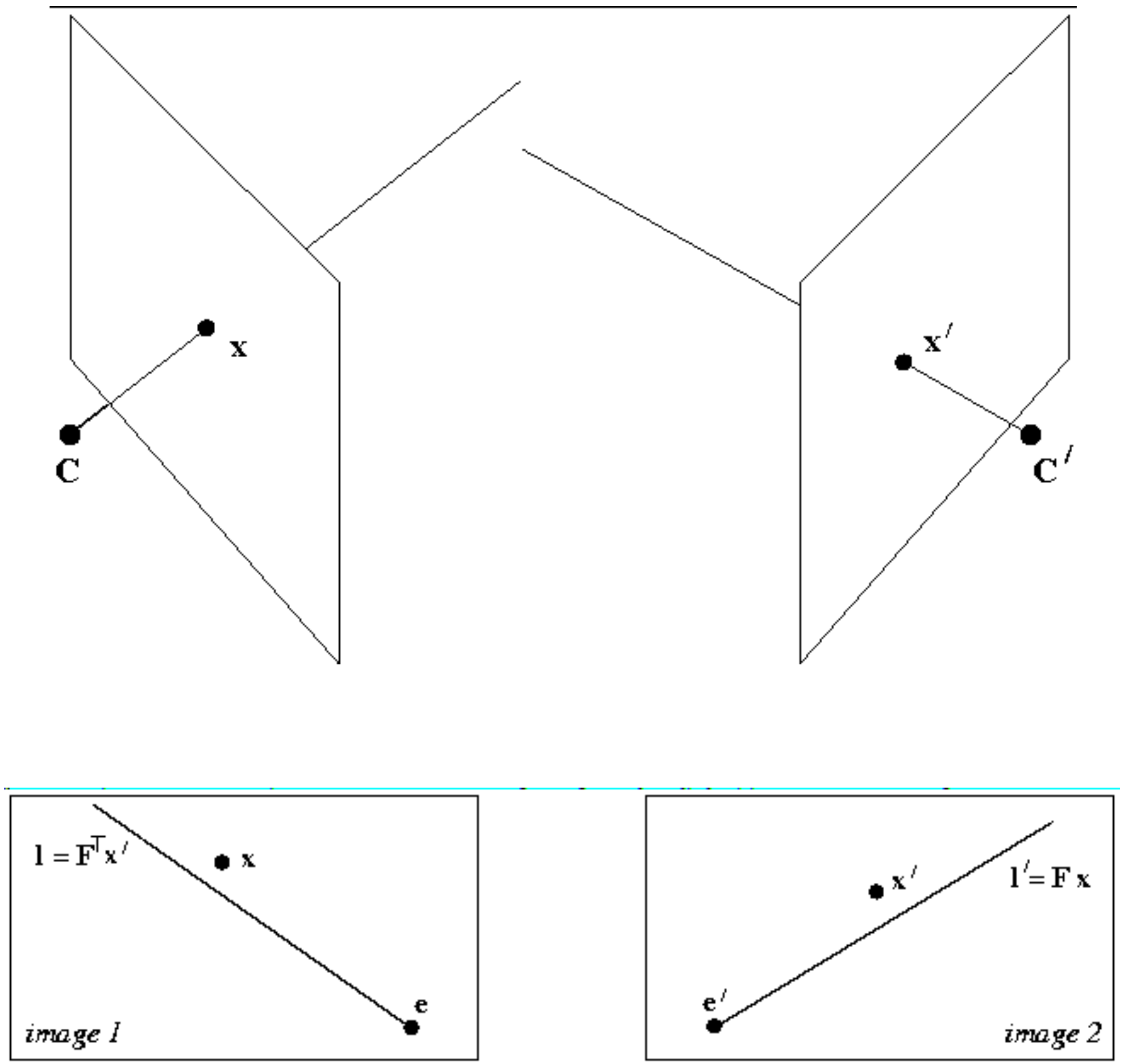
CALIBRATED CASE: RECOVERY OF THE PROJECTION MATRICES FROM E



or most

- to choose one solution you may check the reconstructed points and choose the solution that brings all points in the right place --- that is, in front of the camera

TRIANGULATION: REAL CASE



(DLT) LINEAR METHOD FOR TRIANGULATING POINTS

- for each point correspondence we have two constraints

$$\mathbf{x} = P\mathbf{X} \quad \mathbf{x}' = P'\mathbf{X}$$

- as usual the homogeneous scale factor is eliminated by considering cross products

$$\mathbf{x} \times P\mathbf{X} = \mathbf{0} \quad \mathbf{x}' \times P'\mathbf{X}' = \mathbf{0}$$

$$P = \begin{bmatrix} p^1 \\ p^2 \\ p^3 \end{bmatrix}$$

- in details (for point $\mathbf{x}$)

$$x(\mathbf{p}^{3\top}\mathbf{X}) - (\mathbf{p}^{1\top}\mathbf{X}) = 0$$

$$y(\mathbf{p}^{3\top}\mathbf{X}) - (\mathbf{p}^{2\top}\mathbf{X}) = 0$$

$$x(\mathbf{p}^{2\top}\mathbf{X}) - y(\mathbf{p}^{1\top}\mathbf{X}) = 0$$

(DLT) LINEAR METHOD FOR TRIANGULATING POINTS

- after discarding linearly dependent equations, for each point pair we obtain 4 equations which form the homogeneous system A**X**=**0** with

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

- In the next slides additional material on the uncalibrated case (internal parameters unknown)

UNCALIBRATED CASE: PROJECTION MATRICES FROM F

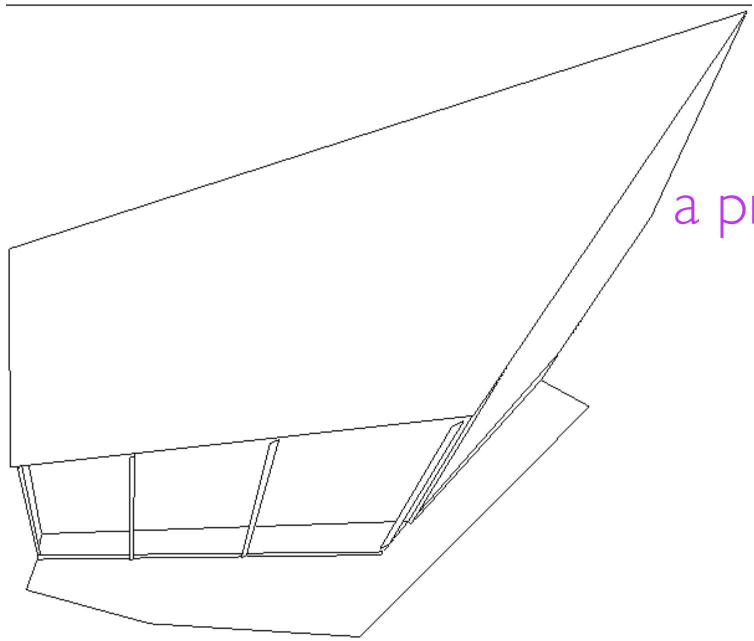
- From the correspondences we may estimate F
- the camera matrices corresponding to a fundamental matrix F may be chosen as

$$P = [I; \mathbf{0}]$$
$$P' = [[\mathbf{e}_R]_\times F; \mathbf{e}_R]$$

UNCALIBRATED CASE
DLT - COMMENTS

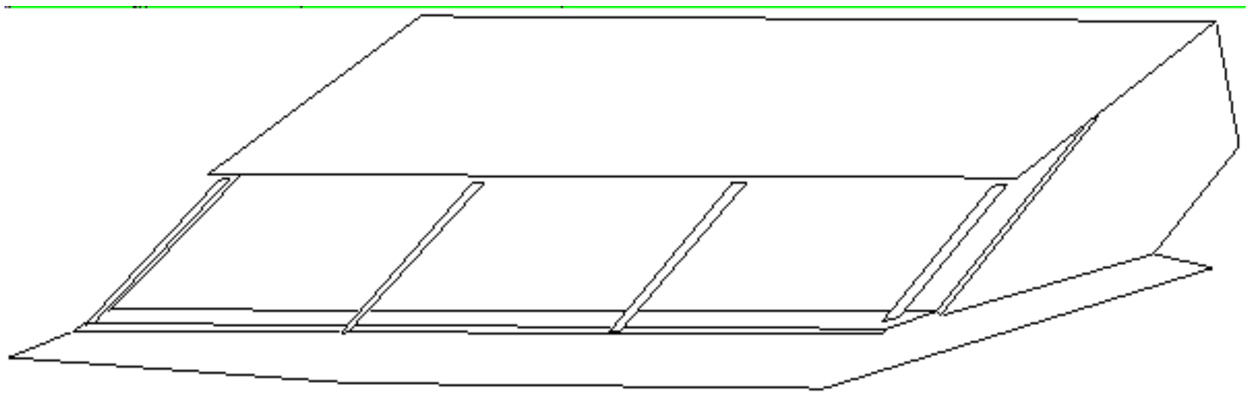
- as all DLT algorithms the solution we obtain is not invariant to projective transformations
  - for this reason the algorithm is not suitable for projective reconstructions
- in this case a better approach is through the minimization of the geometric error

FINAL COMMENTS ON THE UNCALIBRATED CASE



a projective reconstruction

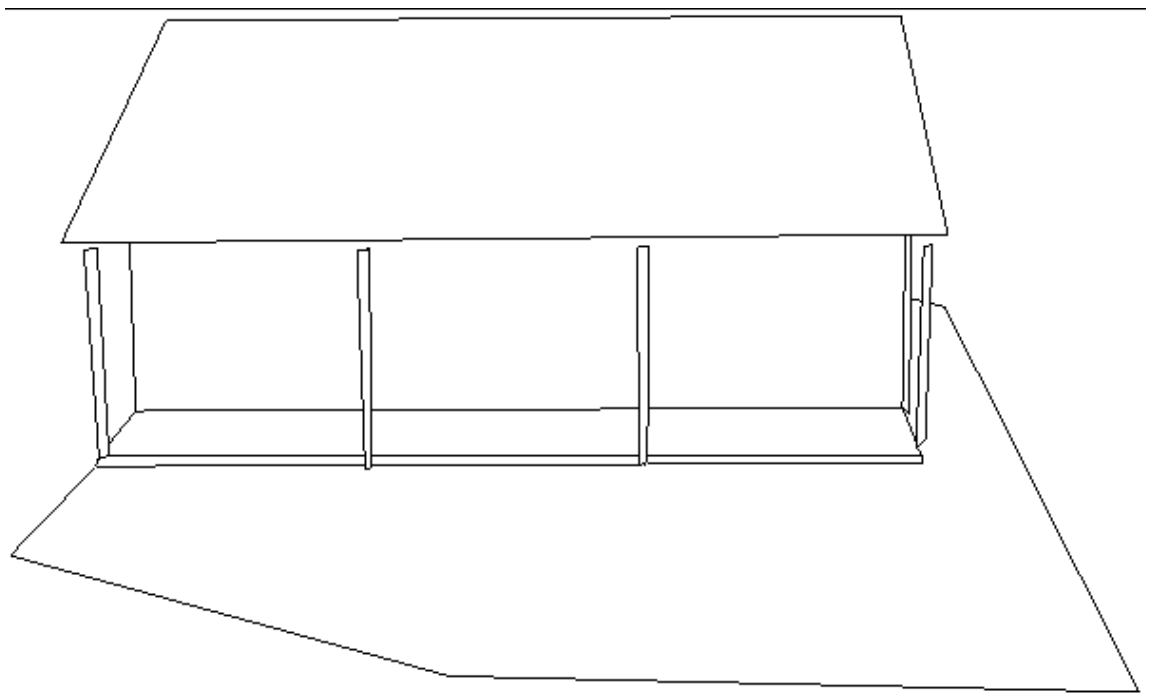
FINAL COMMENTS ON THE UNCALIBRATED CASE



an affine reconstruction

parallel lines are parallel in the reconstruction, but perpendicular lines are not perpendicular in the reconstruction

FINAL COMMENTS ON THE UNCALIBRATED CASE



a metric reconstruction

further constraints (scene orthogonality, square pixels, ...)

DIRECT RECONSTRUCTION (WITH PRIOR KNOWLEDGE)

- if we have a set of *control points* (that is, points with known locations in the 3D world) X_{Ei}
- we may estimate the 3D positions of such points from image correspondences X_i
- and then compute the overall 3D projective transformation H so that $X_{Ei} = HX_i$, $i=1, ..., n$
- each correspondence provides 3 equations, H has 15 d.o.f., thus we need at least 5 control points (no 4 points must be coplanar)

DIRECT RECONSTRUCTION (WITH PRIOR KNOWLEDGE)