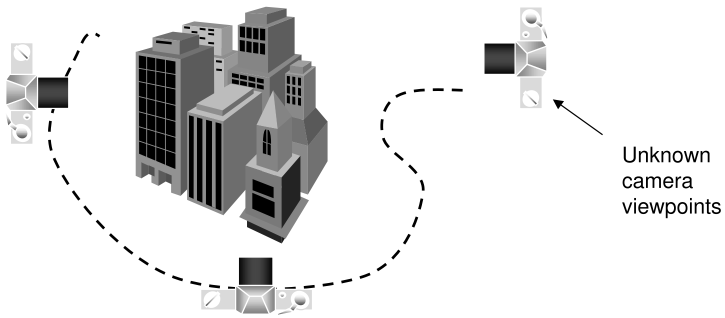
Structure and Motion

Prof A Zisserman Michaelmas 2003



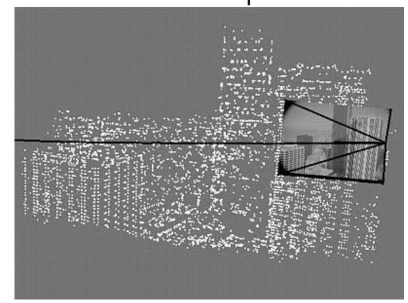
Reconstruct

- Scene geometry
- Camera motion

Example

image sequence

cameras and points



The story so far … stereo reconstruction from 2 views

Given cameras $P = K[I \mid \mathbf{0}]$ $P' = K'[R \mid \mathbf{t}]$

- Epipolar geometry: compute fundamental matrix $F = K'^{-\top}[\mathbf{t}]_{\times}RK^{-1}$
- Correspondence search: 1D search for corresponding points $\mathbf{x} \leftrightarrow \mathbf{x}'$ along epipolar line $\mathbf{l}' = F\,\mathbf{x}$
- Triangulation: compute 3D point $\mathbf{X}$ from $\mathbf{x} \leftrightarrow \mathbf{x}'$, and P, P'

Now, structure and motion …

Structure and Motion: Problem statement

Given 2 (or more) images of a scene, compute the scene structure and the camera motion

- Assume internal calibration (K, K') is known
- Assume scene is rigid
- Start with 2 views only
- NB epipolar geometry is not known

Outline

- Image point motion
- Computing the fundamental matrix
  - 8-point algorithm
  - automation
  - motion from the fundamental matrix
- More than two views
  - matching
  - estimation
- Applications

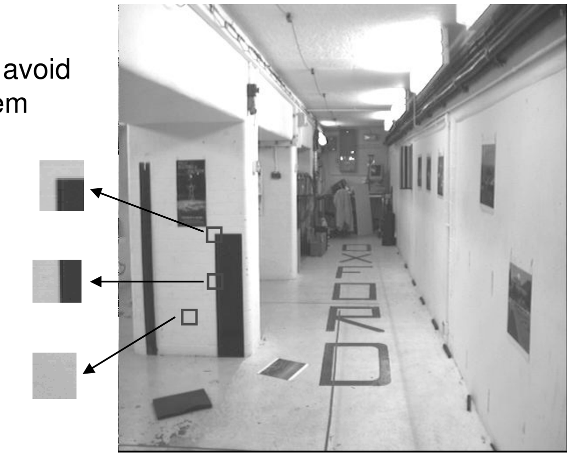
Why use interest points ?

Compute points to avoid the aperture problem

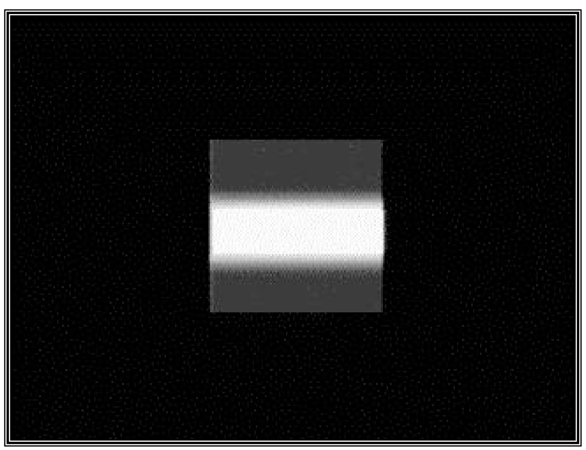
2D feature

1D feature (edge)

uniform



The aperture problem



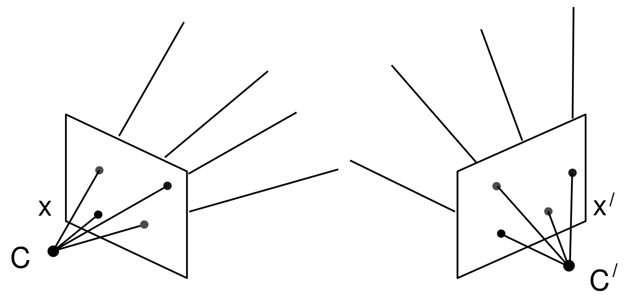
Only the component of motion perpendicular to the line can be determined from local image measurements

interest points computed for each frame

- Harris corner detector

The geometric motion problem

Given image point correspondences, $\mathbf{x}_i \leftrightarrow \mathbf{x}_i'$, determine R and $\mathbf{t}$



Rotate and translate camera until stars of rays intersect

Outline of structure and motion computation

1. Compute the fundamental matrix F from point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}_i'$

2. Compute the cameras (motion) from the fundamental matrix (recall $\mathtt{F} = \mathtt{K}'^{-\top}[\mathbf{t}]_\times \mathtt{R}\mathtt{K}^{-1}$).
   Obtain

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

3. Compute the 3D structure $\mathbf{X}_i$ from the cameras P, P' and point correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}_i'$ (triangulation)

What can be computed from point correspondences?

Suppose we have computed $\mathtt{F} = \mathtt{K}'^{-\top}[\mathbf{t}]_\times \mathtt{R}\mathtt{K}^{-1}$ can the motion be computed?
F is a homogeneous matrix, so

$$\mathtt{F} = \mathtt{K}'^{-\top}[\mathbf{t}]_\times \mathtt{R}\mathtt{K}^{-1} = \mathtt{K}'^{-\top}[\lambda\mathbf{t}]_\times \mathtt{R}\mathtt{K}^{-1}$$

i.e. the translation can only be determined up to scale. This is a consequence of the depth / speed ambiguity: only the ratio of $\mathbf{t}$ and Z can be computed since if

$$\mathbf{t} \rightarrow \lambda\mathbf{t} \text{ and } Z \rightarrow \lambda Z$$

the images are unchanged.

- a large motion of a distant object, and
- a small motion of a nearby object

are indistinguishable (from point motion alone)

Summary: the rotation R (3 dof) can be determined completely, but only the translation direction (2 dof) can be determined, not its magnitude

How many point correspondences are required ?

- for n points there are 3n unknowns (the 3D position of each point)
- for 2 views there are 5 unknowns (that are recoverable)
- each point correspondence gives 4 measurements
- for n points expect a solution if 4n $\geq$ (3n + 5), i.e. n $\geq$ 5
- we will give solutions for n = 7 and n = 8 correspondences

Computing the fundamental matrix

Problem statement

Given: *n* corresponding points $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i, i = 1, \ldots, n\}$
compute the fundamental matrix F such that

$$\mathbf{x}_i'^{\top} \mathrm{F} \mathbf{x}_i = 0 \qquad 1 \leq i \leq n$$

Solution

Each point correspondence $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ generates one constraint on $\mathtt{F}$

$$(\, x'_i \; y'_i \; 1\,) \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = 0$$

which may be written

$$x'xf_1 + x'yf_2 + x'f_3 + y'xf_4 + y'yf_5 + y'f_6 + xf_7 + yf_8 + f_9 = 0$$

$$(x'x,\; x'y,\; x',\; y'x,\; y'y,\; y',\; x,\; y,\; 1) \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \end{pmatrix} = 0$$

For n points

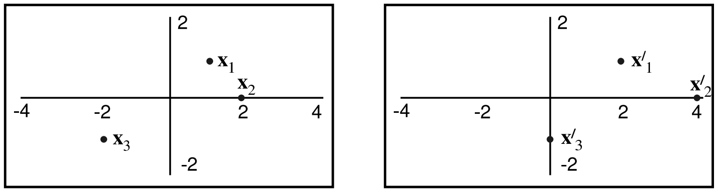
$$\mathtt{A}\mathbf{f} = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \end{pmatrix} = 0$$

A is an n x 9 measurement matrix, and $\mathbf{f}$ is the fundamental matrix written as a 9-vector

- For 8 points, A is an 8 x 9 matrix and $\mathbf{f}$ can be computed as the null-vector of A, i.e. $\mathbf{f}$ is determined up to scale
- Note, this solution (and those following) does not require (K, K$'$)

Example: compute F from 8 point correspondences

Images from a parallel camera stereo rig – epipolar lines $y = y'$



just consider first three points

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \quad \begin{pmatrix} x'_1 \\ y'_1 \end{pmatrix} \qquad \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \quad \begin{pmatrix} x'_2 \\ y'_2 \end{pmatrix} \qquad \begin{pmatrix} x_3 \\ y_3 \end{pmatrix} \quad \begin{pmatrix} x'_3 \\ y'_3 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} 2 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 2 \\ 0 \end{pmatrix} \leftrightarrow \begin{pmatrix} 4 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} -2 \\ -1 \end{pmatrix} \leftrightarrow \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

$$\mathtt{A} = \begin{bmatrix} x'_i x_i & x'_i y_i & x'_i & y'_i x_i & y'_i y_i & y'_i & x_i & y_i & 1 \\ 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 0 & 4 & 0 & 0 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 2 & 1 & 0 & -2 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$$\mathtt{A} = \begin{bmatrix} x'_i x_i & x'_i y_i & x'_i & y'_i x_i & y'_i y_i & y'_i & x_i & y_i & 1 \\ 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 \\ 8 & 0 & 4 & 0 & 0 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 2 & 1 & 0 & -2 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$$\mathbf{f} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}^\top$$

satisfies $\mathtt{A}\mathbf{f} = \mathbf{0}$

write $\mathbf{f}$ in matrix form

$$\mathtt{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

The "8-point" algorithm – Least squares solution

Given n corresponding points (n is typically hundreds) with noise on their measured positions

For $n > 8$ point correspondences, $\mathtt{A}$ is a $n \times 9$ matrix,

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

and in general there will not be an exact solution to $\mathtt{A}\mathbf{f} = \mathbf{0}$.

A (linear) solution which minimises $||\mathtt{A}\mathbf{f}||$, subject to $||\mathbf{f}|| = 1$ is obtained from the eigenvector with least eigenvalue of $\mathtt{A}^\top\mathtt{A}$.

Solution for 7 points

1. Form the 7×9 set of equations $\mathtt{A}\mathbf{f} = \mathbf{0}$
2. The system has a 2-dimensional solution set
3. General solution (use SVD) has the form

$$\mathbf{f} = \lambda \mathbf{f}_0 + \mu \mathbf{f}_1$$

4. In matrix terms

$$\mathtt{F} = \lambda \mathtt{F}_0 + \mu \mathtt{F}_1$$

5. Condition $\det \mathtt{F} = 0$ gives cubic equation in λ and μ
6. Either one or three real solutions for ratio $\lambda : \mu$

A note on minimizing residuals

We have seen two examples of needing to minimize residuals of the form || A $\mathbf{x}$ || over $\mathbf{x}$

1. In computing the fundamental matrix from point correspondences over two views

   For $n > 8$ point correspondences, $\mathtt{A}$ is a $n \times 9$ matrix,

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

2. In triangulating the 3D position of a point from its image in two or more views

$$\mathbf{x} = \mathtt{P}\mathbf{X} \qquad \mathbf{x}' = \mathtt{P}'\mathbf{X} \qquad \begin{bmatrix} x\mathbf{p}^{3\top} - \mathbf{p}^{1\top} \\ y\mathbf{p}^{3\top} - \mathbf{p}^{2\top} \\ x'\mathbf{p}'^{3\top} - \mathbf{p}'^{1\top} \\ y'\mathbf{p}'^{3\top} - \mathbf{p}'^{2\top} \end{bmatrix} \mathbf{X} = \mathbf{0}$$

   For m views $\mathtt{A}$ is a $2m \times 4$ matrix

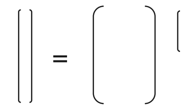
We want to avoid the trivial solution $\mathbf{x} = \mathbf{0}$, so add the constraint that || $\mathbf{x}$ || = 1

$$\min_{\mathbf{x}} ||\mathtt{A}\mathbf{x}|| \quad \text{subject to} \quad ||\mathbf{x}|| = 1$$

For a m x n matrix (with m > n) the vector $\mathbf{x}$ that minimizes || A $\mathbf{x}$ || subject to || $\mathbf{x}$ || = 1 is given by the eigenvector of $A^{\top}$ A corresponding to the least eigenvalue

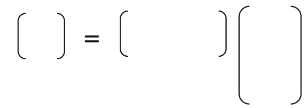
Proof

Write the residuals as a m-vector $\mathbf{r} = \mathtt{A}\mathbf{x}$



Then $||\mathbf{r}||^2 = \mathbf{r}^\top \mathbf{r} = \mathbf{x}^\top \mathtt{A}^\top \mathtt{A}\mathbf{x}$

Write $\mathtt{M} = \mathtt{A}^\top \mathtt{A}$. This is a $n \times n$ (i.e. square) positive semi-definite symmetric matrix:



- The eigenvalues λ_i are real, and the eigenvectors $\mathbf{e}_i$ are orthonormal, $\mathbf{e}_i.\mathbf{e}_j = \delta_{ij}$
- Let $\mathbf{x} = \mathbf{e}_i$ then $\mathbf{e}_i{}^\top \mathtt{M}\mathbf{e}_i = \lambda_i$, and since $\mathbf{x}^\top \mathtt{M}\mathbf{x} \geq 0 \; \forall \; \mathbf{x}$ it follows that $\lambda_i \geq 0$

Eigenvector decomposition

$$\mathtt{M} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 & \ldots & \mathbf{e}_n \\ \vdots & \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{e}_1{}^\top \ldots \\ \mathbf{e}_2{}^\top \ldots \\ \mathbf{e}_3{}^\top \ldots \\ \vdots \\ \mathbf{e}_n{}^\top \ldots \end{bmatrix} = \sum_i \lambda_i \mathbf{e}_i \mathbf{e}_i{}^\top$$

where $0 \leq \lambda_1 < \lambda_2 < \ldots < \lambda_n$

Then

$$\mathbf{x}^\top \mathtt{M}\mathbf{x} = \lambda_1(\mathbf{x}.\mathbf{e}_i)^2 + \lambda_2(\mathbf{x}.\mathbf{e}_2)^2 \ldots + \lambda_n(\mathbf{x}.\mathbf{e}_n)^2$$

This is minimized if $\mathbf{x} = \mathbf{e}_1$.

Automatic Computation of the fundamental matrix

Given Image pair

Find The fundamental matrix F and correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$.

- Compute image points
- Compute correspondences
- Compute epipolar geometry

Step 1: interest points

Harris corner detector

100's of points per image

Step 2a: match points – proximity

• proximity - search within disparity window

Step 2b: match points – cross-correlate

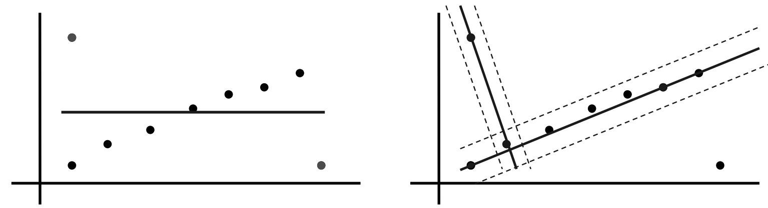
• cross-correlate on intensity neighbourhoods

Correlation matching results

• Many wrong matches (10-50%), but enough to compute F

Robust line estimation - RANSAC

Fit a line to 2D data containing outliers



There are two problems

1. a line fit which minimizes perpendicular distance
2. a classification into inliers (valid points) and outliers

Solution: use robust statistical estimation algorithm RANSAC

(RANdom Sample Consensus) [Fishler & Bolles, 1981]

RANSAC robust line estimation

Repeat

1. Select random sample of 2 points
2. Compute the line through these points
3. Measure support (number of points within threshold distance of the line)

Choose the line with the largest number of inliers

- Compute least squares fit of line to inliers (regression)

Algorithm summary – RANSAC robust F estimation

Repeat

1. Select random sample of 7 correspondences
2. Compute F (1 or 3 solutions)
3. Measure support (number of inliers within threshold distance of epipolar line)

Choose the F with the largest number of inliers

Correspondences consistent with epipolar geometry

- Use RANSAC robust estimation algorithm
- Obtain correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ and $\mathtt{F}$

Computed epipolar geometry

Determining cameras from the fundamental matrix

Decomposing the fundamental matrix

$F = K'^{-\top}[\mathbf{t}]_\times R K^{-1}$

Form the Essential matrix $E = [\mathbf{t}]_\times R = K'^\top F K$

1. Compute $\mathbf{t}$ as left null-vector of E, i.e. $E^\top \mathbf{t} = \mathbf{0}$
   This determines $\mathbf{t}$ up to scale.

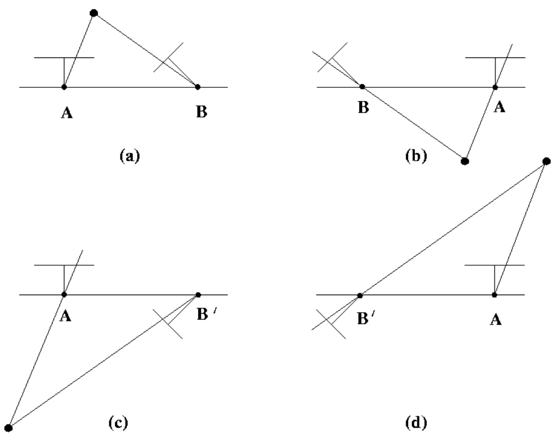
2. Compute R from E (see below)
   There are two solutions R_1 and R_2.

3. Set $P = K[I \mid \mathbf{0}]$ for the first camera

   The four solutions for the second camera are

$$P' = K'[R_1 \mid \mu\mathbf{t}] \qquad P' = K'[R_1 \mid -\mu\mathbf{t}] \qquad \mu > 0$$
$$P' = K'[R_2 \mid \mu\mathbf{t}] \qquad P' = K'[R_2 \mid -\mu\mathbf{t}]$$

The four camera solutions



The 3D point is only in front of both cameras in one case

Computing the rotation matrix from the Essential matrix (non-examinable)

- Compute the SVD of $E = U \operatorname{diag}(1,1,0) V^\top$

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

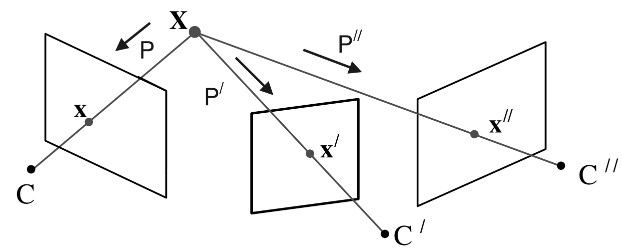
- Solutions are $R_1 = UWV^\top \quad R_2 = UW^\top V^\top$

Structure and Motion for more than 2 views

Notation for three or more views

For 3 views the cameras are P, P/ and P//, and a 3D point is imaged as

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



For m views, a point $\mathbf{X}_j$ is imaged in the "i" th view as

$\mathbf{x}^i_j = P^i \mathbf{X}_j$

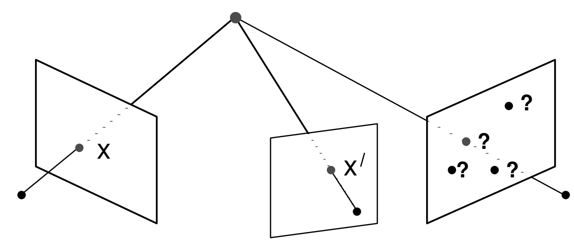
What is gained by having more than 2 views?

1. The two view ambiguities do not get worse
   - there is an overall scale ambiguity
2. Matching:
   - matches can be verified
3. Estimation:
   - accuracy increased by using more measurements

Point correspondence over 3 views

Given: the cameras P, P/ and P//, and matching points $\mathbf{x}$ and $\mathbf{x}'$

Find: the matching point in the third view



Algorithm:

- compute the 3D point from $\mathbf{x}$ and $\mathbf{x}'$ and project it into the third view
- the matching point coincides with the projected point

Problem statement: structure and motion

Given: n matching image points $\mathbf{x}^i_j$ over m views

Find: the cameras P^i and the 3D points $\mathbf{X}_j$ such that $\mathbf{x}^i_j = P^i \mathbf{X}_j$

$$\min_{P^i \, \mathbf{X}_j} \sum_{j \in \text{points}} \sum_{i \in \text{views}} d\left(\mathbf{x}^i_j, \; P^i \mathbf{X}_j\right)^2$$

number of parameters

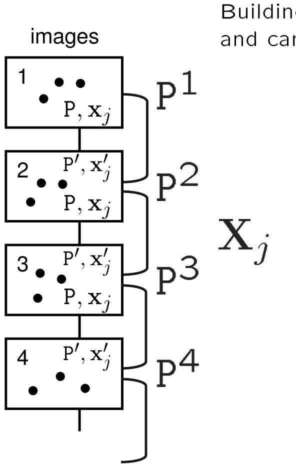
- for each camera there are 6 parameters
- for each 3D point there are 3 parameters

a total of 6 m + 3 n parameters must be estimated

Application: Augmented reality

original sequence

Algorithm for structure and motion



Building block is computing correspondences $\{\mathbf{x}_j \leftrightarrow \mathbf{x}'_j\}$ and cameras P, P' for an image pair via F

Algorithm

- Compute interest points in each image
- Compute point correspondences between consecutive image pairs
- Extend and verify correspondences and cameras over image triplets
- Extend correspondences and cameras over all images
- Optimize over $\{P^i, \mathbf{X}_j\}$

Augmentation