CEE598 - Visual Sensing for Civil Infrastructure Eng. & Mgmt.

Session 15 – Perspective Structure from Motion

Mani Golparvar-Fard
Department of Civil and Environmental Engineering
3129D, Newmark Civil Engineering Lab
e-mail: mgolpar@illinois.edu

Structure from Motion

- http://www.youtube.com/watch?v=4cEQZreQ2zQ
- http://www.youtube.com/watch?v=mKHIn9JXIzQ&feature=related
- http://www.youtube.com/watch?v=iCvr47kcz5o&feature=related
- PhotoCity Gameplay

http://www.youtube.com/watch?v=iDbmpFHIrkg&feature=player_embedded

Outline

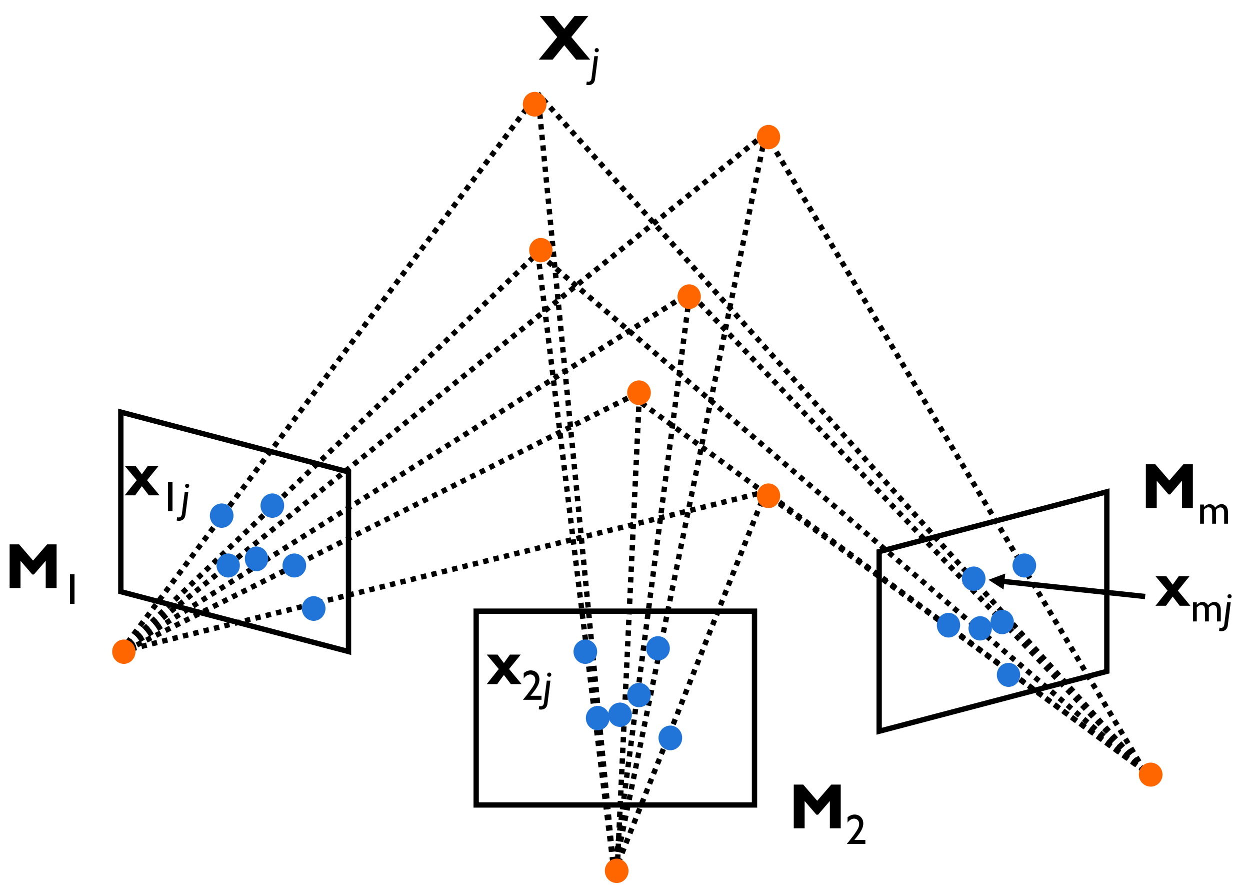
- Multiple view geometry
  Perspective Structure from Motion

- Perspective structure from motion problem
- Ambiguities
- Algebraic methods
- Factorization methods
- Bundle adjustment
- Self-calibration

- Reading: [HZ] Chapters: 10,18,19
  [FP] Chapter: 13

Some slides of this lectures are courtesy of prof Savarese, prof. J. Ponce, prof FF Li, prof S. Lazebnik & prof. M. Hebert

Structure from motion problem

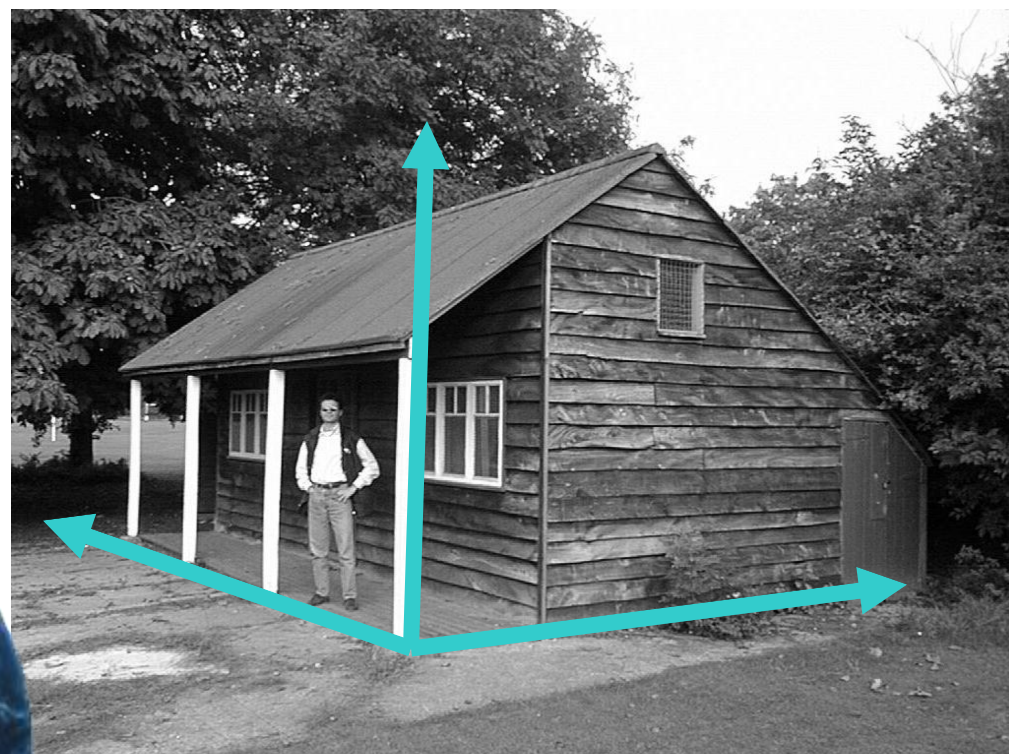


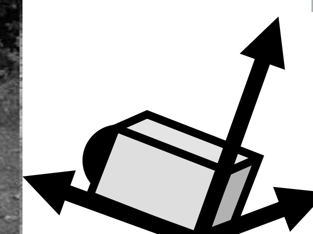
From the m×n correspondences $\mathbf{x}_{ij}$, estimate:

- *m* projection matrices $\mathbf{M}_i$ — motion
- *n* 3D points $\mathbf{X}_j$ — structure

Structure from motion ambiguity

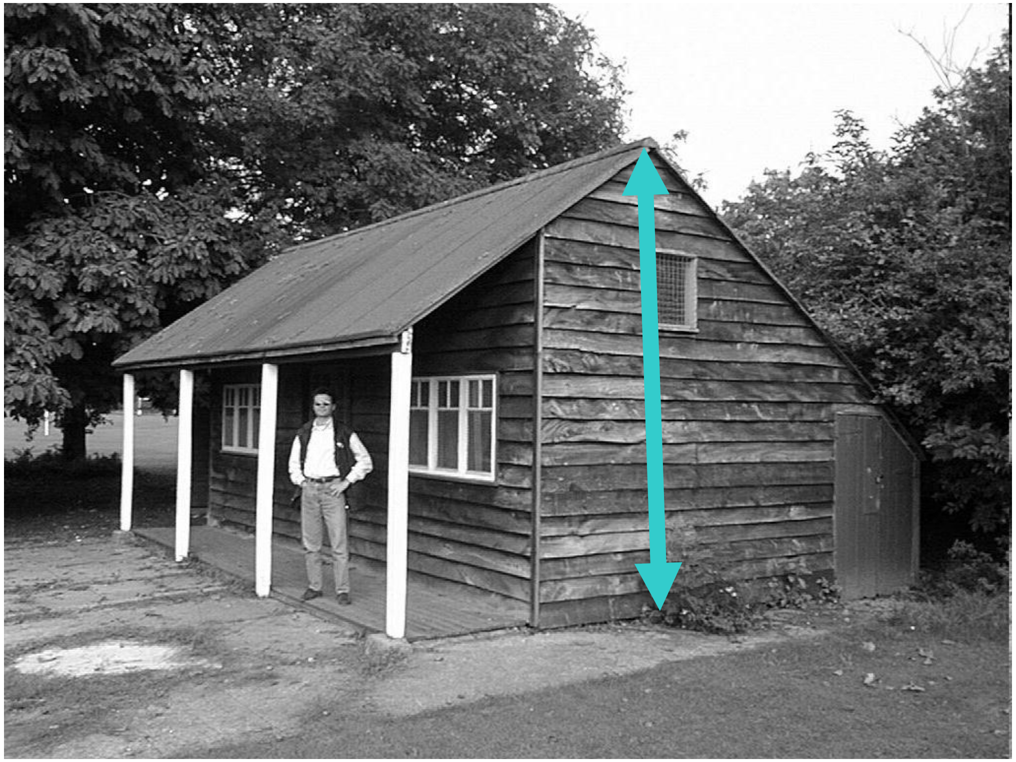
Position ambiguity: it is impossible based on the images alone to estimate the <u>absolute</u> location and pose of the scene w.r.t. a 3D world coordinate frame





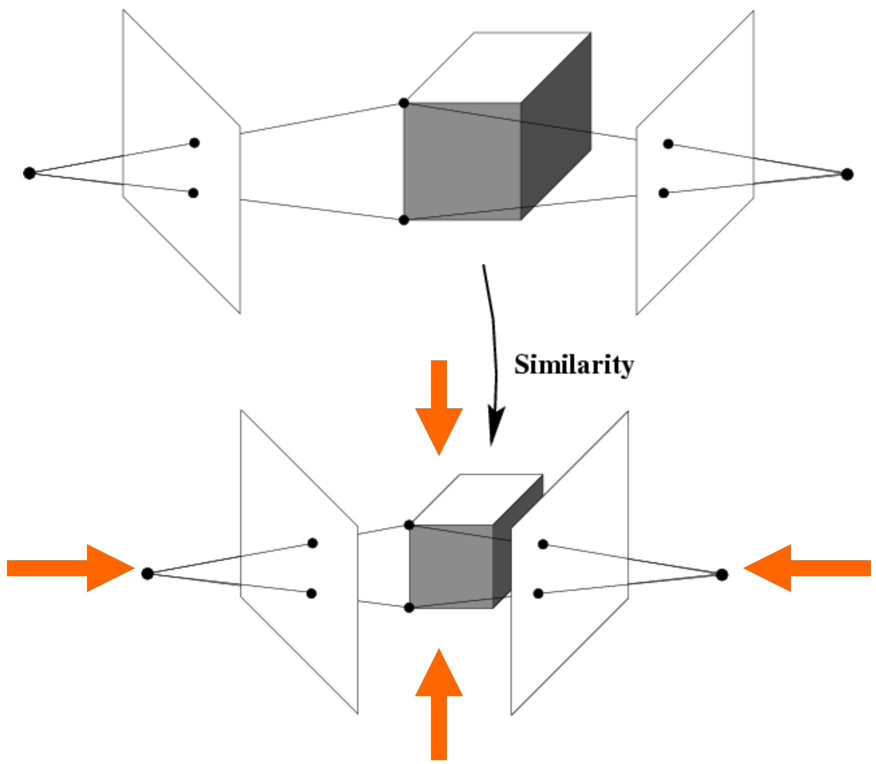
Structure from motion ambiguity

Scale ambiguity: It is impossible based on the images alone to estimate the absolute scale of the scene (i.e. house height)



Structure from motion ambiguity

-The scene is determined by the images only up a similarity transformation (rotation, translation and scaling)

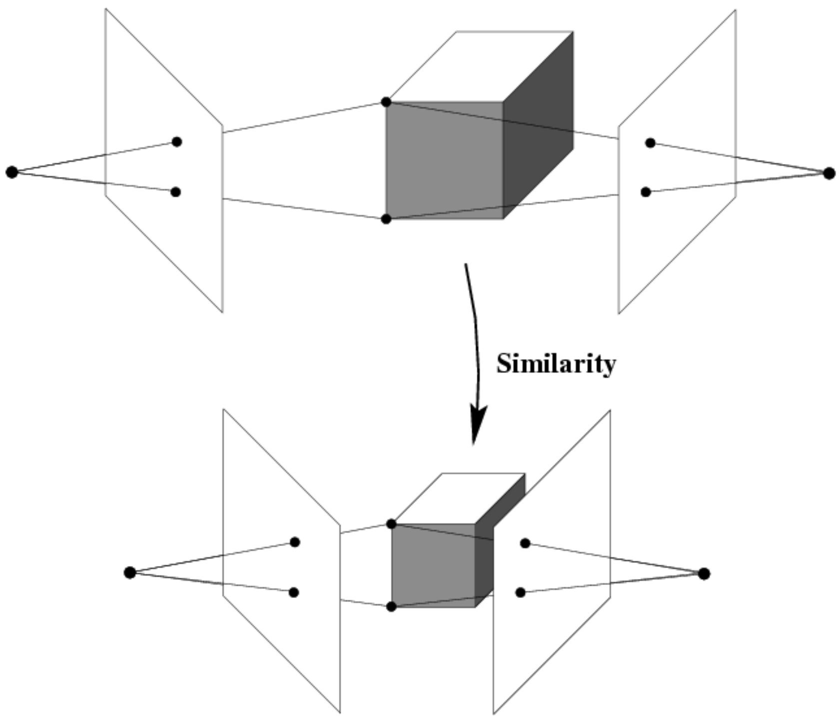


$$H_s = \begin{bmatrix} s & & \\ & s & \\ & & 1 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$$

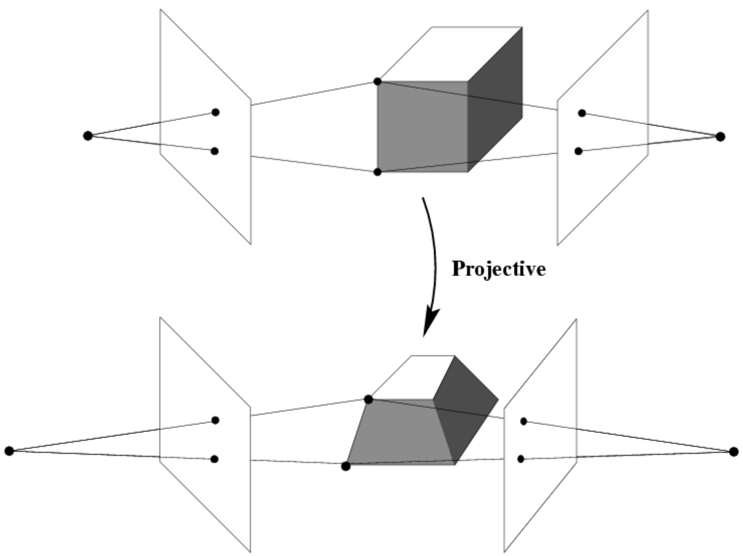
Structure from motion ambiguity

- The ambiguity exists even for calibrated cameras
- For calibrated cameras, the similarity ambiguity is the only ambiguity

[Longuet-Higgins '81]



Structure from motion ambiguity



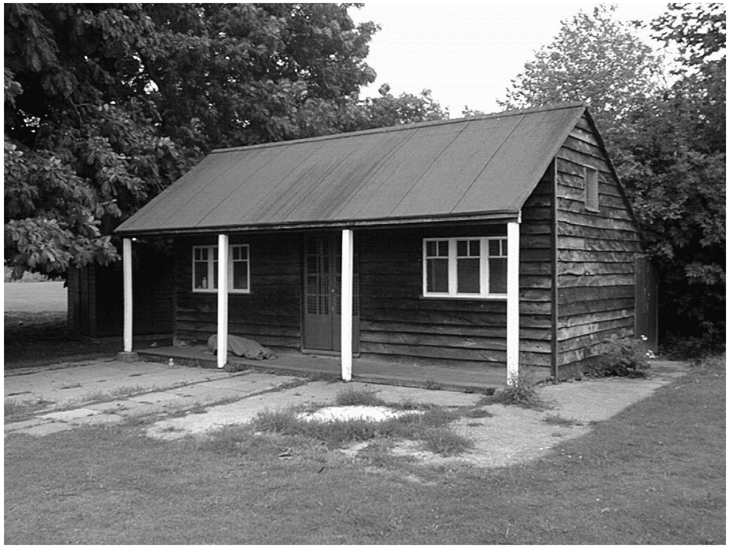
- In the general case (nothing is known) the ambiguity is expressed by an arbitrary affine or projective transformation

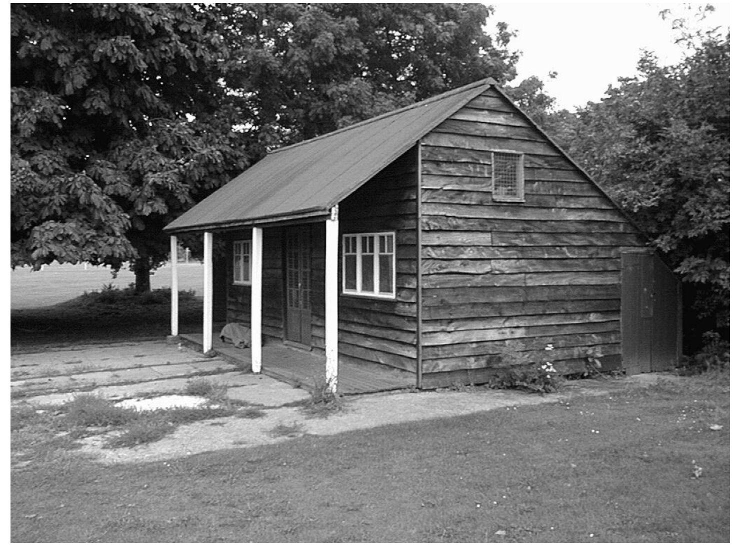
$$x_j = M_i X_j \quad \rightarrow \quad X_j \to H X_j$$

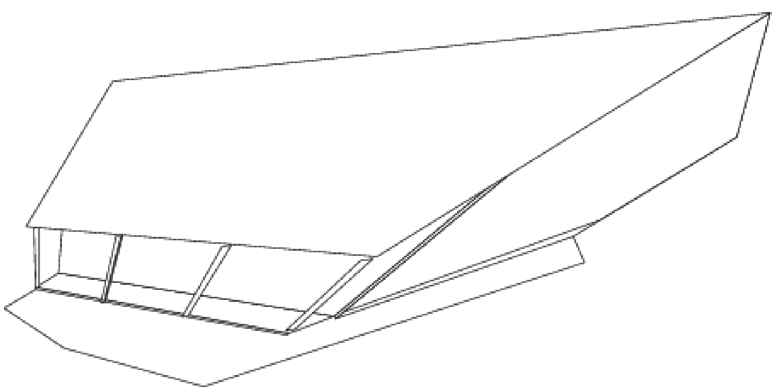
$$M_i = K_i \begin{bmatrix} R_i & T_i \end{bmatrix} \quad \rightarrow \quad M_i \to M_j H^{-1}$$

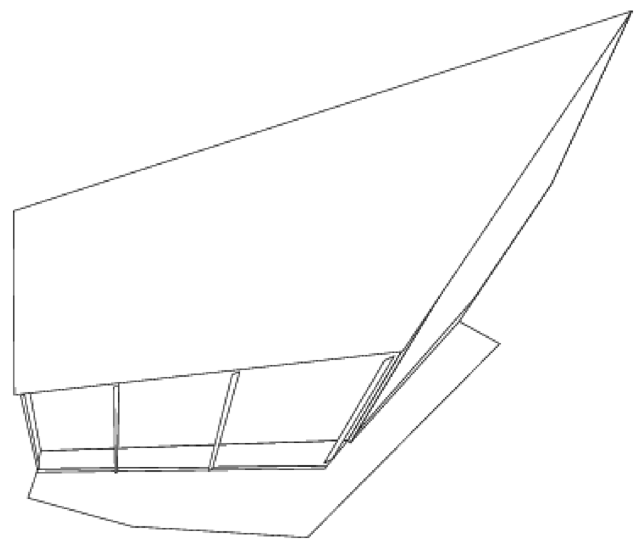
$$x_j = M_i X_j = \left(M_i H^{-1}\right)\left(H X_j\right)$$

Projective ambiguity

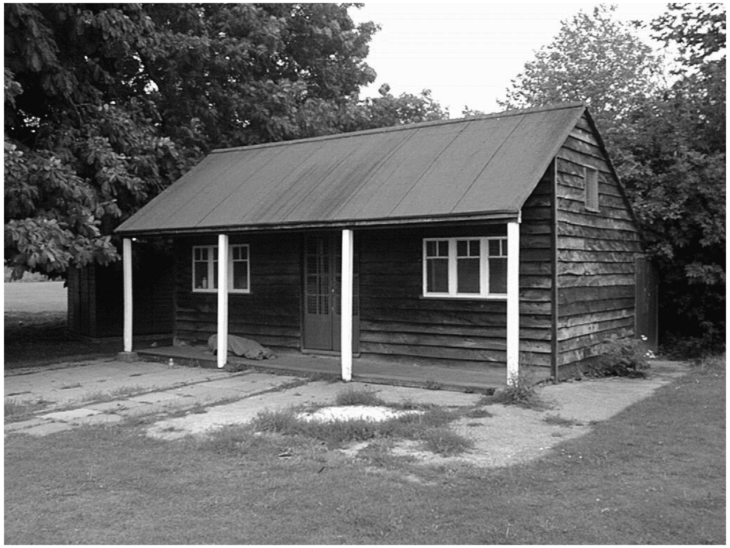


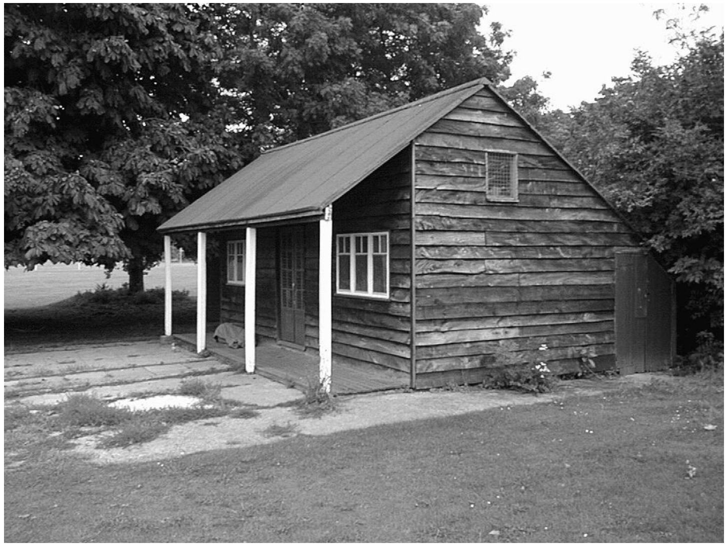


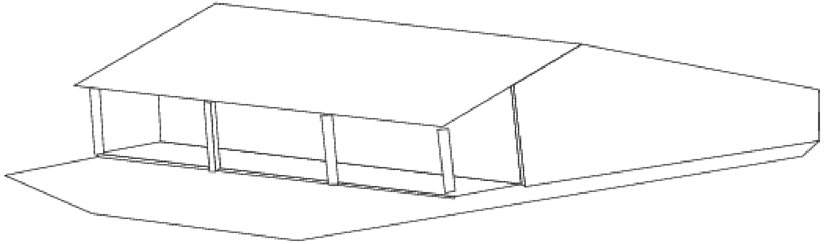


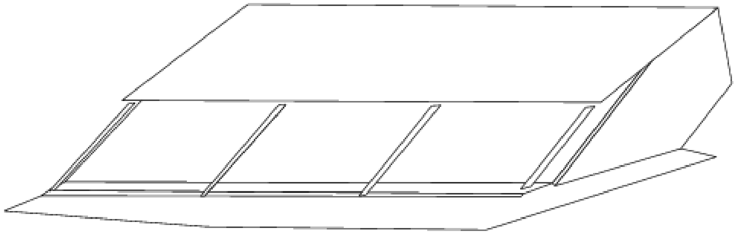


Affine ambiguity





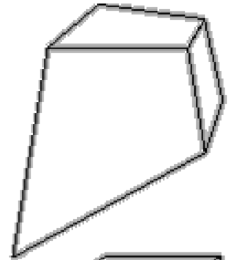




Hierarchy of 3D transformations

Projective
15dof

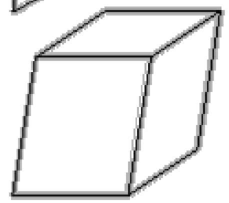
$$\begin{bmatrix} A & t \\ v^T & v \end{bmatrix}$$



Preserves intersection and tangency

Affine
12dof

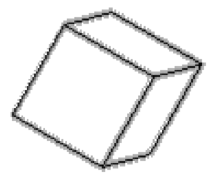
$$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$$



Preserves parallellism, volume ratios

Similarity
7dof

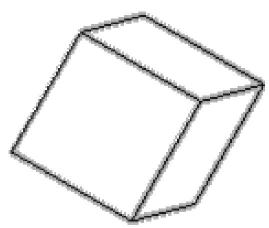
$$\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$$



Preserves angles, ratios of length

Euclidean
6dof

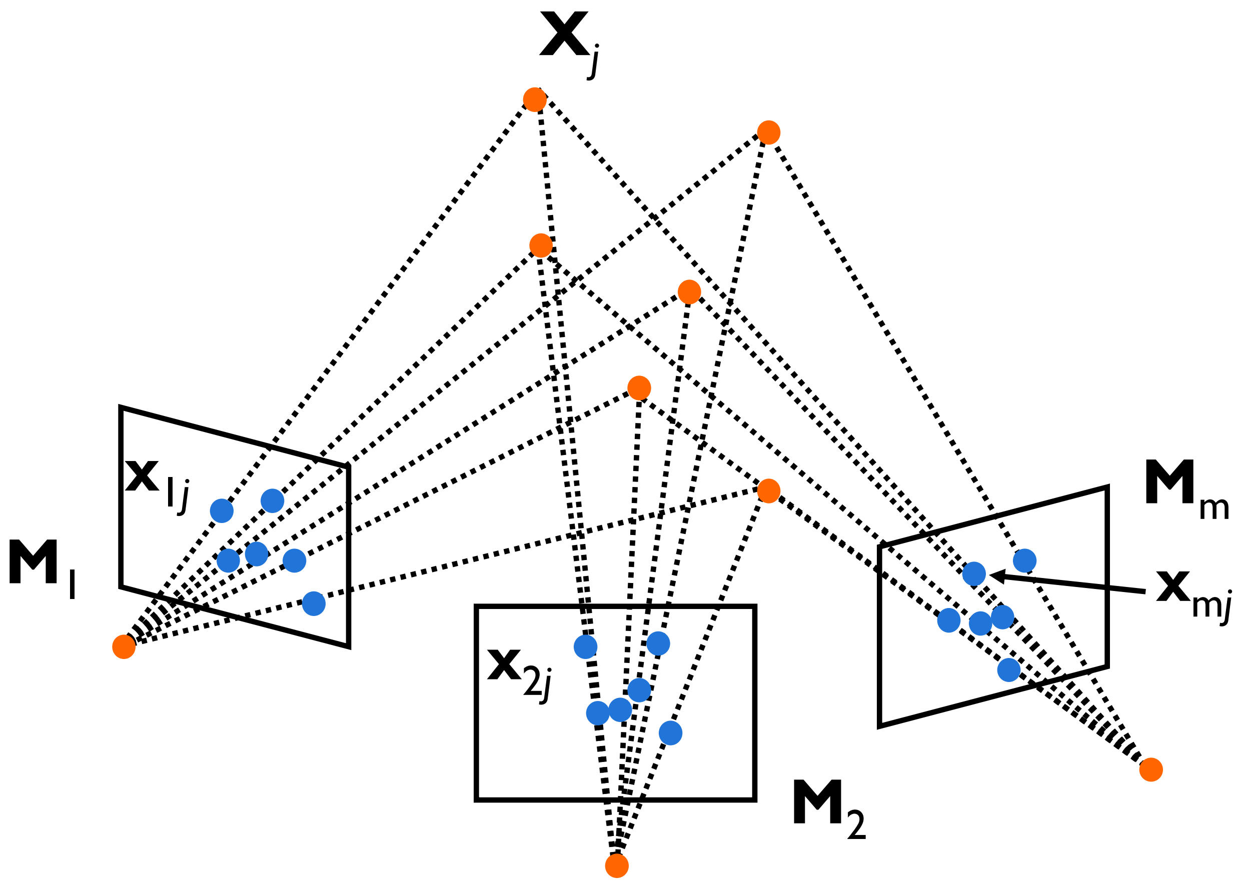
$$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$$



Preserves angles, lengths

- With no constraints on the camera calibration matrix or on the scene, we get a *projective* reconstruction
- Need additional information to *upgrade* the reconstruction to affine, similarity, or Euclidean

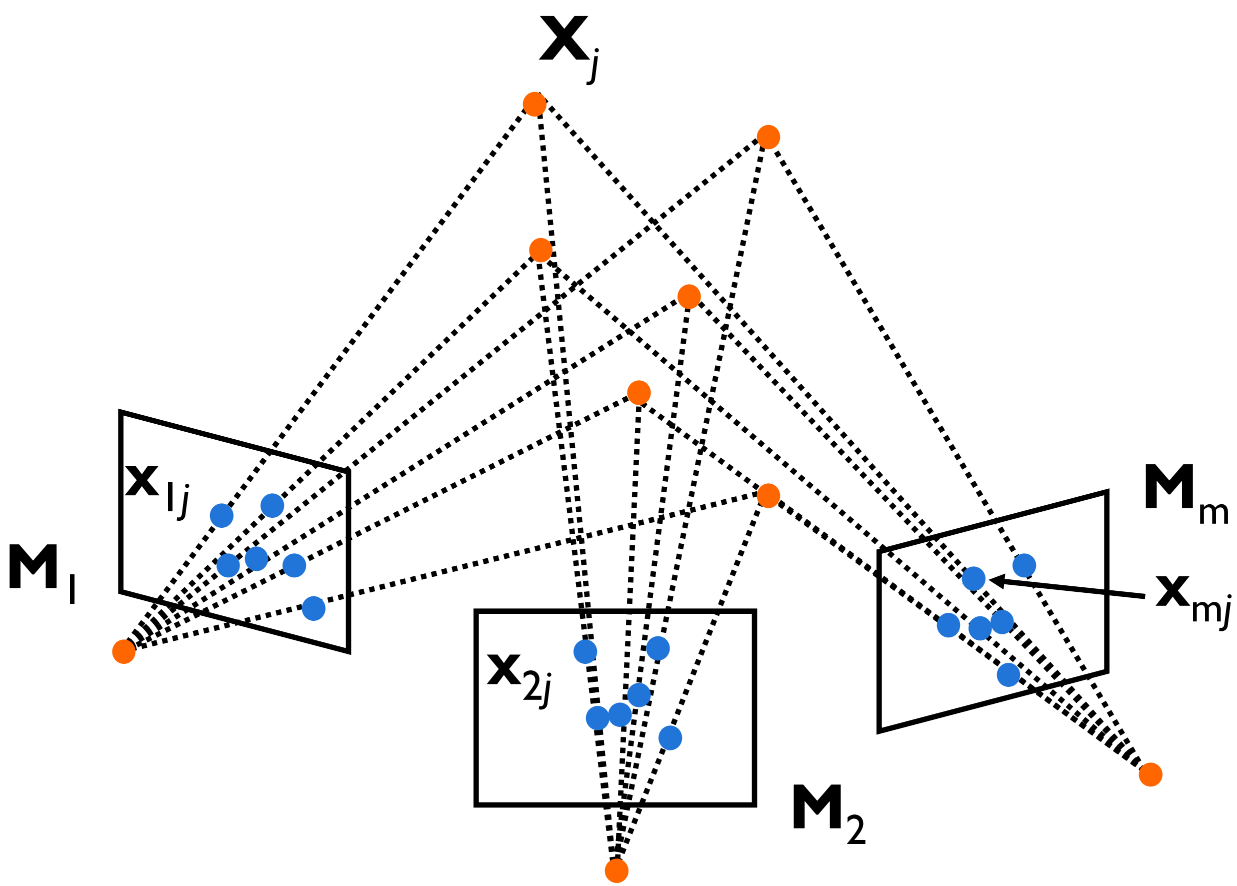
Structure from motion problem



Given *m* images of *n* fixed 3D points

- $\mathbf{x}_{ij} = \mathbf{M}_i \mathbf{X}_j, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$

Structure from motion problem



m cameras $M_1 \ldots M_m$

$$M_i = \begin{bmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & 1 \end{bmatrix}$$

The Structure-from-Motion Problem

Given *m* images of *n* fixed points X_j we can write

$$x_{ij} = M_i X_j \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n.$$

Problem: Estimate the m 3×4 matrices M_i and the n positions X_j from the m×n correspondences x_{ij} .

- With no calibration info, cameras and points can only be recovered up to a 4x4 projective
- Given two cameras, how many points are needed?
- How many equations and how many unknown?

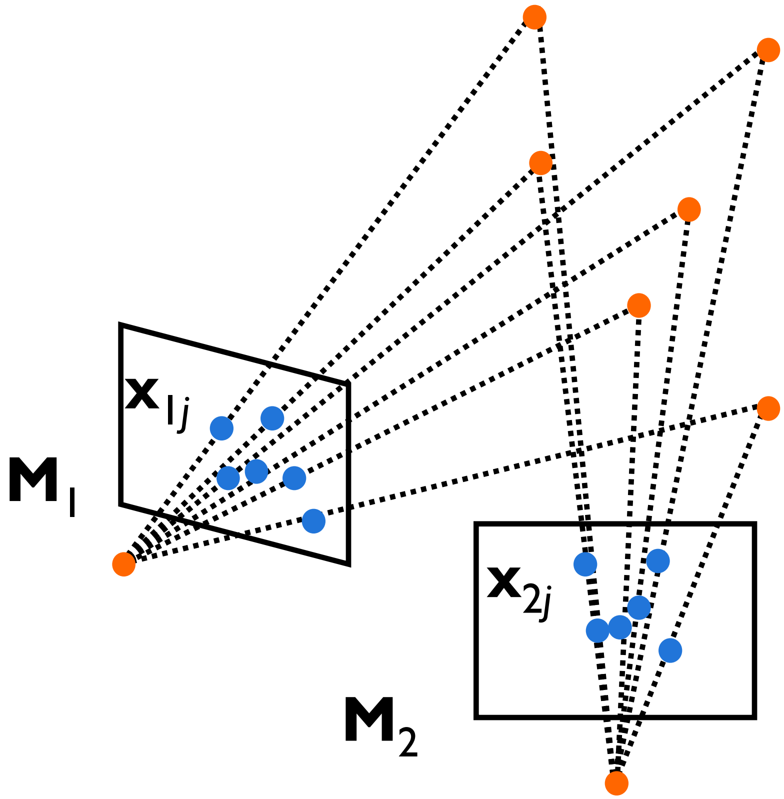
2m × n equations in 11m+3n – 15 unknowns

So 7 points! [2x2x7 = 28; 11x2 + 3x7 – 15 = 28]

Structure-from-Motion Algorithms

- Algebraic approach (by fundamental matrix)
- Factorization method (by SVD)
- Bundle adjustment

Algebraic approach (2-view case)



$$x_{ij} = M_i X_j$$

Apply a projective transformation H such that:

$$M_1 H^{-1} = \begin{bmatrix} I & 0 \end{bmatrix} \qquad M_2 H^{-1} = \begin{bmatrix} A & b \end{bmatrix}$$

Canonical perspective cameras

Algebraic approach (Fundamental matrix)

$$\mathbf{x} = \mathrm{M}_1\,\mathrm{H}^{-1}\,\tilde{\mathbf{X}} = [\mathbf{I}\,|\,\mathbf{0}]\tilde{\mathbf{X}} \qquad \mathbf{x}' = \mathrm{M}_2\,\mathrm{H}^{-1}\,\tilde{\mathbf{X}} = [\mathbf{A}\,|\,\mathbf{b}]\tilde{\mathbf{X}}$$

$$\tilde{\mathbf{X}} = \mathrm{H}\,\mathbf{X}$$

$$\mathbf{x}' = [\mathbf{A}\,|\,\mathbf{b}]\tilde{\mathbf{X}} = [\mathbf{A}\,|\,\mathbf{b}]\begin{bmatrix}\tilde{X}_1\\ \tilde{X}_2\\ \tilde{X}_3\\ 1\end{bmatrix} = \mathbf{A}[\mathrm{I}\,|\,0]\begin{bmatrix}\tilde{X}_1\\ \tilde{X}_2\\ \tilde{X}_3\\ 1\end{bmatrix} + \mathbf{b} = \mathbf{A}\underbrace{[\mathrm{I}\,|\,0]\,\tilde{\mathbf{X}}}_{?} + \mathrm{b} \quad = \mathrm{A}\mathbf{x} + \mathrm{b}$$

$$\mathbf{x}' \times \mathbf{b} = (\mathbf{Ax} + \mathbf{b}) \times \mathbf{b} = \mathbf{Ax} \times \mathbf{b}$$

$$(\mathbf{Ax} \times \mathbf{b}) \cdot \mathbf{x}' = (\mathbf{x}' \times \mathbf{b}) \cdot \mathbf{x}' = 0$$

Cross product as matrix multiplication

$$\mathbf{a} \times \mathbf{b} = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = [\mathbf{a}_\times]\mathbf{b}$$

Algebraic approach (Fundamental matrix)

$$\mathbf{x} = M_1 H^{-1} \tilde{\mathbf{X}} = [\mathbf{I} \mid \mathbf{0}]\tilde{\mathbf{X}} \qquad \mathbf{x}' = M_2 H^{-1} = [\mathbf{A} \mid \mathbf{b}]\tilde{\mathbf{X}}$$

$$\mathbf{x}' = [\mathbf{A} \mid \mathbf{b}]\tilde{\mathbf{X}} = [\mathbf{A} \mid \mathbf{b}]\begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \tilde{X}_3 \\ 1 \end{bmatrix} = \mathbf{A}[I \mid 0]\begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \tilde{X}_3 \\ 1 \end{bmatrix} + \mathbf{b} = \mathbf{A}[I \mid 0]\,\tilde{\mathbf{X}} + b \overset{?}{=} A\mathbf{x} + b$$

$\tilde{X} = H\,X$

$$\mathbf{x}' \times \mathbf{b} = (\mathbf{Ax} + \mathbf{b}) \times \mathbf{b} = \mathbf{Ax} \times \mathbf{b}$$

$$(\mathbf{Ax} \times \mathbf{b}) \cdot \mathbf{x}' = (\mathbf{x}' \times \mathbf{b}) \cdot \mathbf{x}' = 0$$

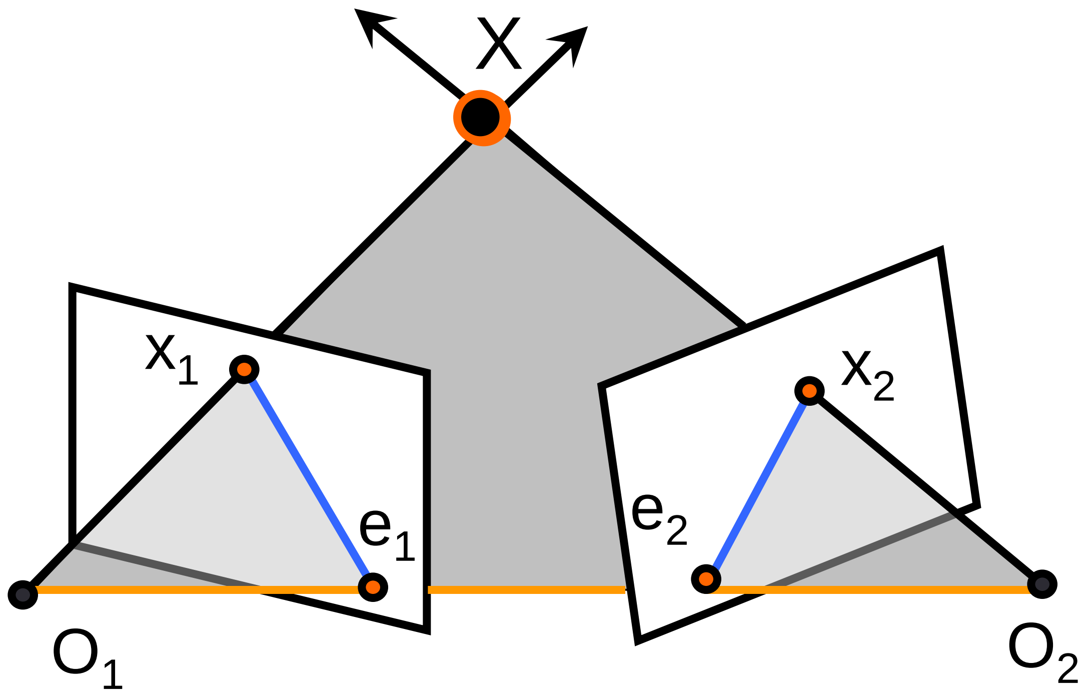
$$\mathbf{x}'^{T}[\mathbf{b}_{\times}]\mathbf{A}\mathbf{x} = 0$$

is this familiar?

$$\mathbf{F} = [\mathbf{b}_{\times}]\mathbf{A}$$

$$\mathbf{x}'^{T} F\,\mathbf{x} = 0$$

Epipolar Constraint [from previous lectures]



- $F\,x_2$ is the epipolar line associated with x_2 ($l_1 = F\,x_2$)
- $F^T x_1$ is the epipolar line associated with x_1 ($l_2 = F^T x_1$)
- F is singular (rank two)
- $F\,e_2 = 0$ and $F^T e_1 = 0$
- F is 3x3 matrix; 7 DOF

Algebraic approach (Fundamental matrix)

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

$$\mathbf{F} = [\mathbf{b}_{\times}]\mathbf{A}$$

$$\mathbf{F} \cdot b = [\mathbf{b}_{\times}]\mathbf{A} \cdot b = 0$$

Since **F b** = 0, **b=e** → Compute **b** as least sq. solution of **F b** = 0
det(F)=0; |b|=1 →

$$\mathbf{A} = [\mathbf{b}_{\times}]^{-1} \mathbf{F} = -[\mathbf{b}_{\times}]\mathbf{F}$$

$$M^{p}_{1} = \begin{bmatrix} I & 0 \end{bmatrix} \qquad M^{p}_{2} = \begin{bmatrix} -[\mathbf{e}_{x}]\mathbf{F} & \mathbf{e} \end{bmatrix}$$

Perspective cameras are known

Structure-from-Motion Algorithms

- Algebraic approach (by fundamental matrix)
- Factorization method (by SVD)
- Bundle adjustment

Projective factorization

$$\mathbf{D} = \begin{bmatrix} z_{11}\mathbf{x}_{11} & z_{12}\mathbf{x}_{12} & \cdots & z_{1n}\mathbf{x}_{1n} \\ z_{21}\mathbf{x}_{21} & z_{22}\mathbf{x}_{22} & \cdots & z_{2n}\mathbf{x}_{2n} \\ & & \ddots & \\ z_{m1}\mathbf{x}_{m1} & z_{m2}\mathbf{x}_{m2} & \cdots & z_{mn}\mathbf{x}_{mn} \end{bmatrix} = \begin{bmatrix} \mathrm{M}_1 \\ \mathrm{M}_2 \\ \vdots \\ \mathrm{M}_m \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n \end{bmatrix}$$

points ($4 \times n$)

cameras ($3m \times 4$)

$\mathbf{D} = \mathbf{MS}$ has rank 4

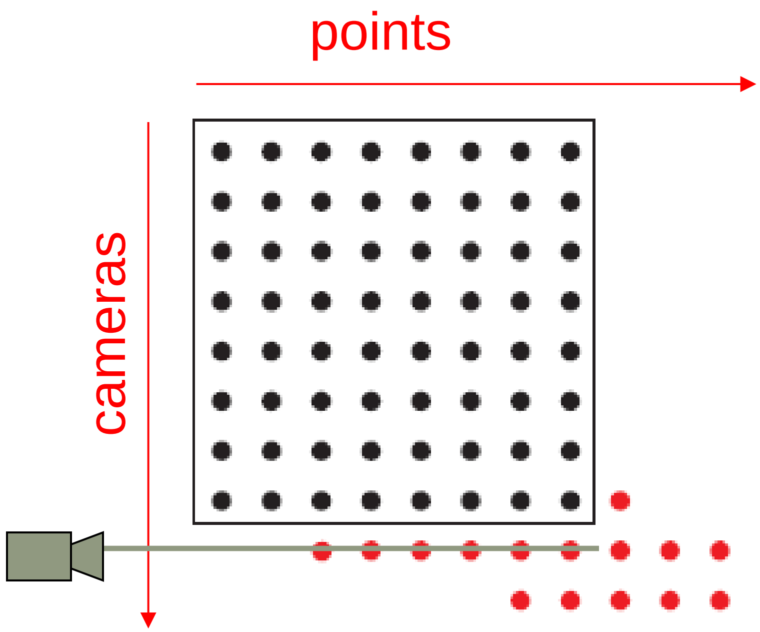
- If we knew the depths *z*, we could factorize **D** to estimate **M** and **S**
- If we knew **M** and **S**, we could solve for *z*
- Solution: iterative approach (alternate between above two steps)

Structure-from-Motion Algorithms

- Algebraic approach (by fundamental matrix)
- Factorization method (by SVD)
- Bundle adjustment

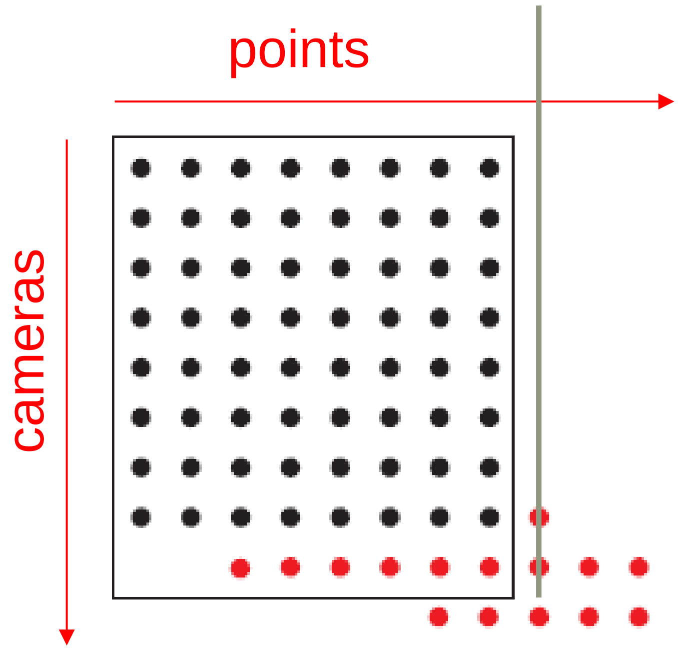
Sequential structure from motion

- Initialize motion from two images using fundamental matrix
- Initialize structure
- For each additional view:
  - Determine projection matrix of new camera using all the known 3D points that are visible in its image – *calibration*



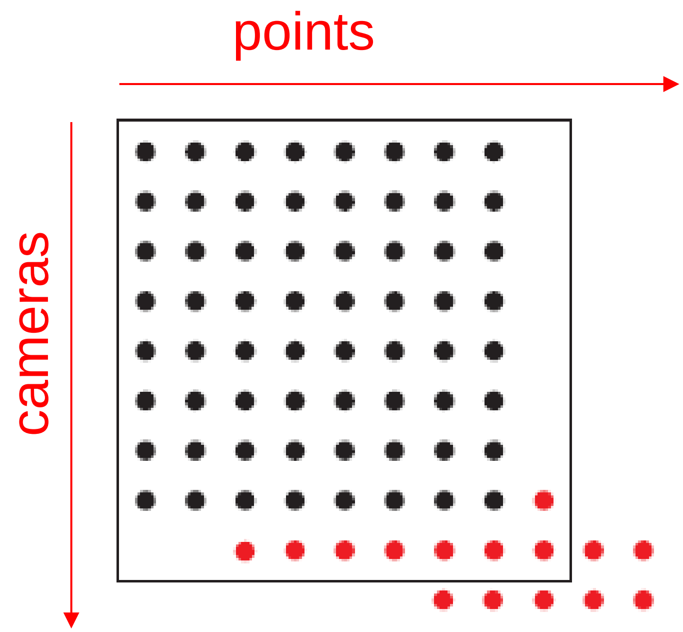
Sequential structure from motion

- Initialize motion from two images using fundamental matrix
- Initialize structure
- For each additional view:
  - Determine projection matrix of new camera using all the known 3D points that are visible in its image – *calibration*
  - Refine and extend structure: compute new 3D points, re-optimize existing points that are also seen by this camera – *triangulation*



Sequential structure from motion

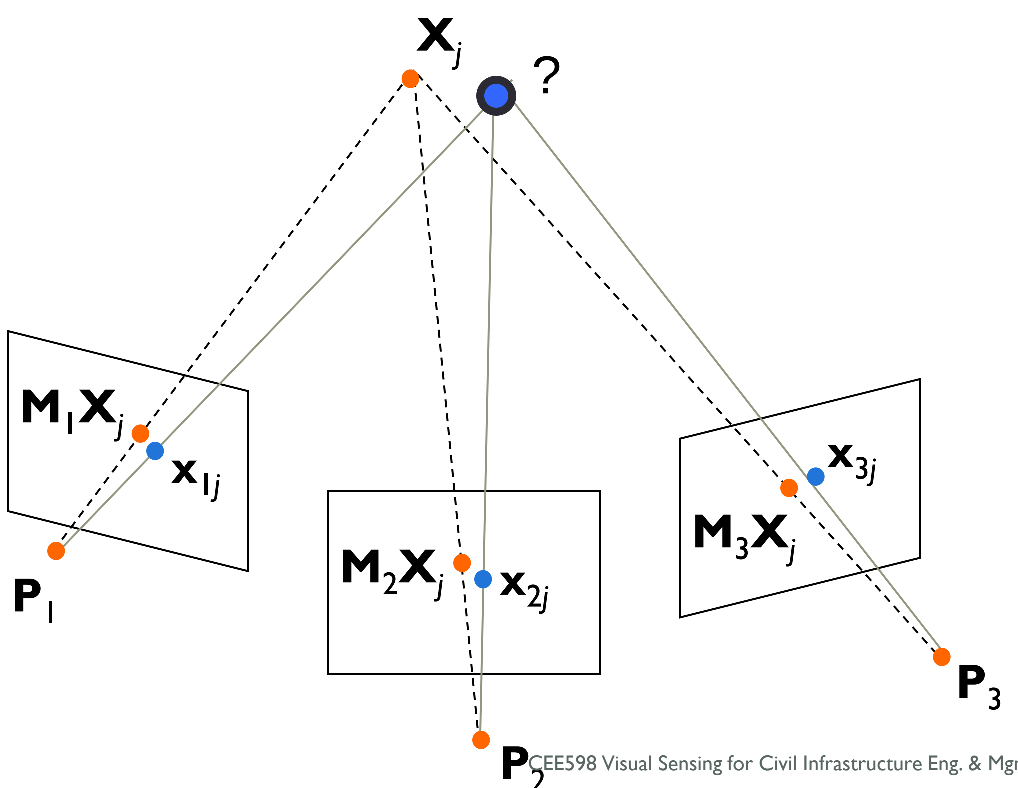
- Initialize motion from two images using fundamental matrix
- Initialize structure
- For each additional view:
  - Determine projection matrix of new camera using all the known 3D points that are visible in its image – *calibration*
  - Refine and extend structure: compute new 3D points, re-optimize existing points that are also seen by this camera – *triangulation*
- Refine structure and motion: bundle adjustment



Bundle adjustment

- Non-linear method for refining structure and motion
- Minimizing re-projection error

$$E(M,\mathbf{X}) = \sum_{i=1}^{m}\sum_{j=1}^{n} D\left(\mathbf{x}_{ij}, M_i\mathbf{X}_j\right)^2$$



Bundle adjustment

- Non-linear method for refining structure and motion
- Minimizing re-projection error

$$E(M, \mathbf{X}) = \sum_{i=1}^{m} \sum_{j=1}^{n} D\left(\mathbf{x}_{ij}, M_i \mathbf{X}_j\right)^2$$

- Advantages
  - Handle large number of views
  - Handle missing data

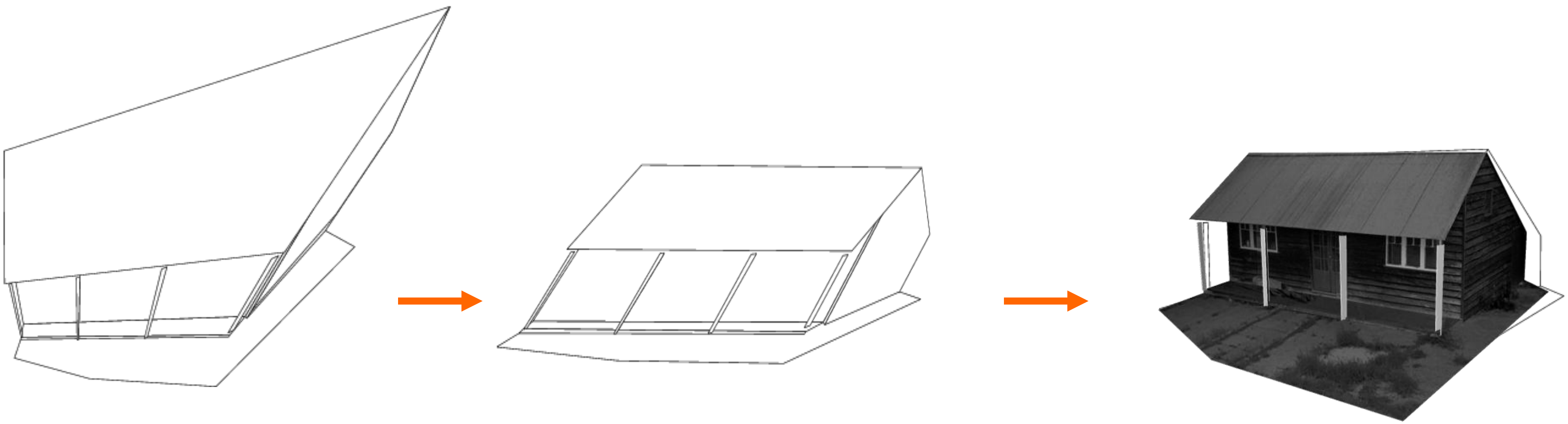
- Limitations
  - Large minimization problem (parameters grow with number of views)
  - requires good initial condition

Used as the final step of SFM

Removing the ambiguities: the Stratified reconstruction

- Up grade reconstruction from perspective to affine
  [by measuring the plane at infinity]

- up grade reconstruction from affine to metric
  [by measuring the absolute conic]



Recovering the metric reconstruction
from the perspective one is called **self-calibration**

Self-calibration

Process of determining intrinsic camera parameters directly from un-calibrated images

Suppose we have a projective reconstruction $\{M_i, X_j\}$

GOAL: find a rectifying (non-singular) homography H such that

$\{M_i\, H, H^{-1} X_j\}$ is a metric reconstruction

$\overline{M}_i \qquad \overline{X}_j$

$$\overline{M}_i = M_i\, H \qquad i = 1 \cdots m \qquad \overline{M}_i = K_i [R_i \quad T_i]$$

If world ref. system = camera 1 ref. system: $\overline{M}_1 = K_1 [I \quad 0]$

If the perspective camera is canonical: $M_1 = [I \quad 0]$

Self-calibration

$$\overline{M}_i = M_i\ H$$

$$[K_1 \quad 0] = [I \quad 0]\ H$$

$$A = K_1$$

$$t = 0$$

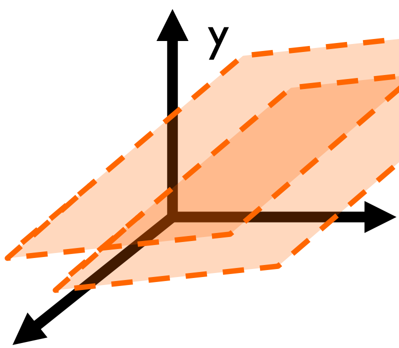
$$H = \begin{bmatrix} K_1 & 0 \\ v & 1 \end{bmatrix}$$

$$\begin{cases} \overline{M}_1 = K_1[I \quad 0] \\ M_1 = [I \quad 0] \\ H = \begin{bmatrix} A & t \\ v & k \end{bmatrix} \end{cases}$$

We can set k=1
(this fixes the scale of the reconstruction)

Planes at infinity (Repeat from Prev. Lectures)

$$\Pi_\infty = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$



In the metric (Euclidean) world coordinates

2 planes are parallel iff their intersections is a line that belongs to Π_∞

The projective transformation of a plane at infinity can be expressed as

$$\boldsymbol{\pi}_\infty = \mathrm{H}^{-1}\,\Pi_\infty = \begin{bmatrix} \mathrm{p} \\ 1 \end{bmatrix}$$

Self-calibration

$$\pi_\infty = \begin{bmatrix} p \\ 1 \end{bmatrix} = H^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad H = \begin{bmatrix} K_1 & 0 \\ v & 1 \end{bmatrix}$$

$$= \begin{bmatrix} K_1^{-T} & -K_1^{-T} v \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -K_1^{-T} v \\ 1 \end{bmatrix} \rightarrow v = -p^T K_1$$

Self-calibration

GOAL: find a rectifying homography H such that

$$\{M_i, X_j\} \rightarrow \{M_i\ H, H^{-1}X_j\}$$ is a metric reconstruction

$$H = \begin{bmatrix} K_1 & 0 \\ -p^T K_1 & 1 \end{bmatrix}$$

$$K = \begin{bmatrix} \alpha & -\alpha \cot\theta & u_o \\ 0 & \frac{\beta}{\sin\theta} & v_o \\ 0 & 0 & 1 \end{bmatrix}$$

K_1 = calibration matrix of first camera

5 unknowns

$\pi_\infty = [p\ \ 1]^T$ = plane at infinity in the projective reconstruction

3 unknowns

Self-calibration basic equation

$$\overline{M}_i = M_i \; H \qquad i = 2 \cdots m$$

$$\begin{cases} M_i = [A_i \quad a_i] & \text{= perspective reconstruction of the camera (known)} \\ \overline{M}_i = K_i [R_i \quad T_i] & \text{= metric reconstruction of the camera} \\ H = \begin{bmatrix} K_1 & 0 \\ -p^T K_1 & 1 \end{bmatrix} & \text{= rectifying homography (unknown)} \end{cases}$$

$$[K_i R_i \quad T_i'] = [A_i \quad a_i] \begin{bmatrix} K_1 & 0 \\ -p^T K_1 & 1 \end{bmatrix} = [A_i K_1 - a_i p^T K_1 \quad a_i]$$

$$K_i R_i = \left(A_i - a_i p^T\right) K_1 \rightarrow R_i = K_i^{-1}\left(A_i - a_i p^T\right) K_1$$

Self-calibration basic equation

$$\begin{cases} R_i = K_i^{-1}\left(A_i - a_i p^T\right) K_1 \\ R_i^T = K_1^T \left(A_i - a_i p^T\right)^T K_i^{-T} \end{cases}$$

$$R_i \; R_i^T = I$$

$$K_i^{-1}\left(A_i - a_i p^T\right) K_1 \; K_1^T \left(A_i - a_i p^T\right)^T K_i^{-T} = I$$

$$\left(A_i - a_i p^T\right) K_1 K_1^T \left(A_i - a_i p^T\right)^T = \boxed{K_i \; K_i^T} \leftarrow ?$$

Absolute conic Ω_∞ is a $C \in \Pi_\infty$

Any $X \in \Omega_\infty$ satisfies:

$$x^T \Omega_\infty \, x = 0 \qquad \Omega_\infty = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 0 \end{bmatrix} \qquad \begin{cases} x_1^2 + x_2^2 + x_3^2 = 0 \\ x_4 = 0 \end{cases}$$

Projective transformation of Ω_∞

$$\boldsymbol{\omega} = (K^T K)^{-1}$$

$$\boldsymbol{\omega}^* = K\,K^T$$

Dual image of the absolute conic

[From previous lectures]

Properties of ω

$$\omega = (K^{T}K)^{-1}$$

- It is not function of R, T

- symmetric (5 unknowns)

$$\omega = \begin{bmatrix} \omega_1 & \omega_2 & \omega_4 \\ \omega_2 & \omega_3 & \omega_5 \\ \omega_4 & \omega_5 & \omega_6 \end{bmatrix}$$

Self-calibration basic equation

$$\left(A_i - a_i p^T\right) K_1 K_1^T \left(A_i - a_i p^T\right)^T = K_i \; K_i^T$$

$$\boxed{\left(A_i - a_i p^T\right) \boldsymbol{\omega}_1^* \left(A_i - a_i p^T\right)^T = \boldsymbol{\omega}_i^*}$$ i=2…m

[A_i and a_i are known]

How many unknowns?

- 3 from **p**
- 5 from ω_i [per view]

How many equations? 5 independent equations [per view]

Art of self-calibration:

use constraints on ω (K) to generate enough equations on the unknowns

Self-calibration – identical Ks

$$\left(A_i - a_i p^T\right)\omega_1^*\left(A_i - a_i p^T\right)^T = \omega_i^*$$

$$\downarrow$$

$$\left(A_i - a_i p^T\right)\omega^*\left(A_i - a_i p^T\right)^T = \omega^*$$

- For m views, 5(m-1) constraints
- Number of unknowns: 8

→ m>=3 provides enough constraints

To solve the self-calibration problem with identical cameras we need at least 3 views

Properties of ω

$$\omega = (K^T K)^{-1}$$

1. $$\omega = \begin{bmatrix} \omega_1 & \omega_2 & \omega_4 \\ \omega_2 & \omega_3 & \omega_5 \\ \omega_4 & \omega_5 & \omega_6 \end{bmatrix}$$

2. $\omega_2 = 0$ zero-skew

3. $\omega_2 = 0$
   $\omega_1 = \omega_3$
   square pixel

4. $\omega_4 = \omega_5 = 0$
   zero-offset

Self-calibration – other constraints

$$\left(A_i - a_i p^T\right) \omega_1^* \left(A_i - a_i p^T\right)^T = \omega_i^*$$

- zero-offset $\omega_4 = \omega_5 = 0$ → 2 m linear constraints
- zero-skew $\omega_2 = 0$ → m linear constraints

etc…

Self-calibration - summary

Condition	N. Views
•Constant internal parameters	3
•Aspect ratio and skew known •Focal length and offset vary	4*
•Aspect ratio and skew known •Focal length and offset vary	5*
•skew =0, all other parameters vary	8*

Issue: the larger is the number of view,
the harder is the correspondence problem

Bundle adjustment helps!

SFM problem - summary

1. Estimate structure and motion up perspective transformation
    1. Algebraic
    2. factorization method
    3. bundle adjustment

2. Convert from perspective to metric (self-calibration)

3. Bundle adjustment

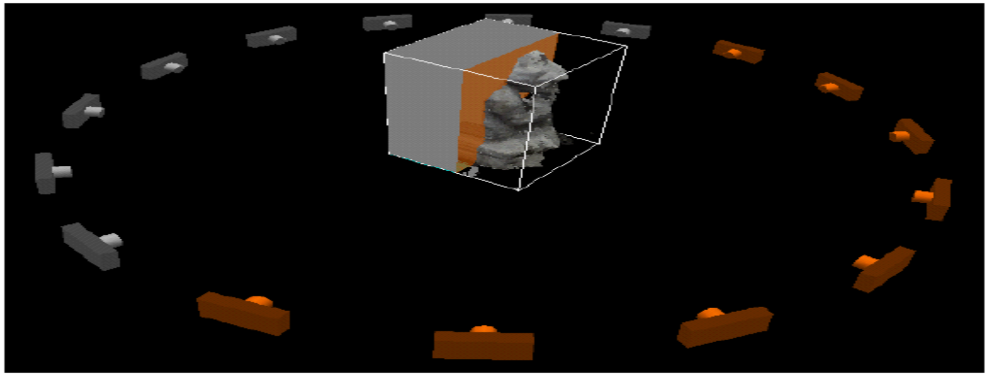
** or **

1. Bundle adjustment with self-calibration constraints

Self-calibration - Summary

Constraints on camera motion can be incorporated

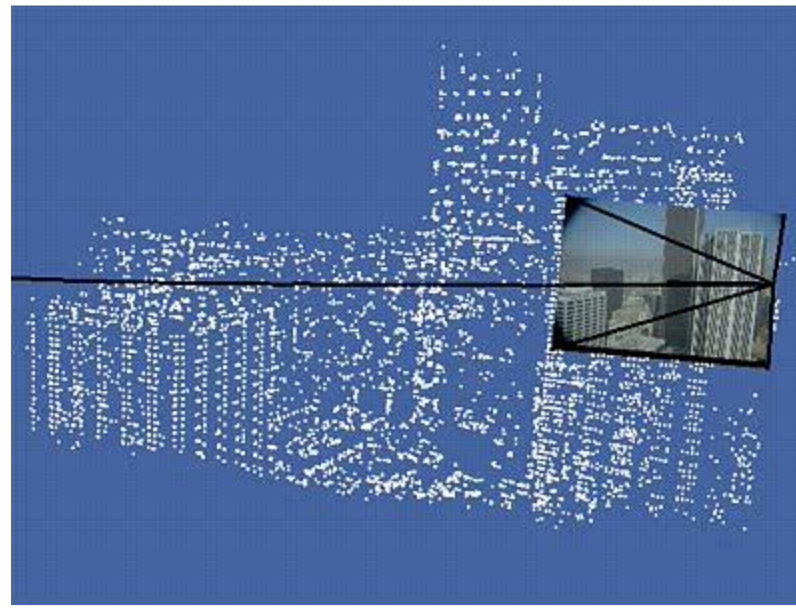
- Linearly translating camera



- Single axis of rotation: turntable motion

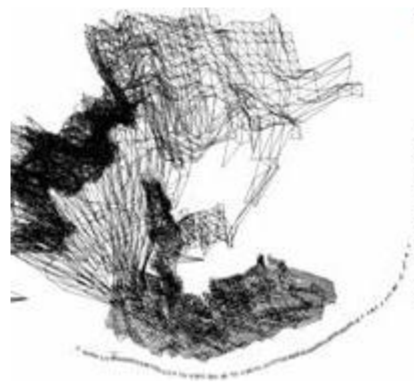
Applications

Courtesy of Oxford **Visual Geometry Group**



Applications

D. Nistér, PhD thesis ‘01

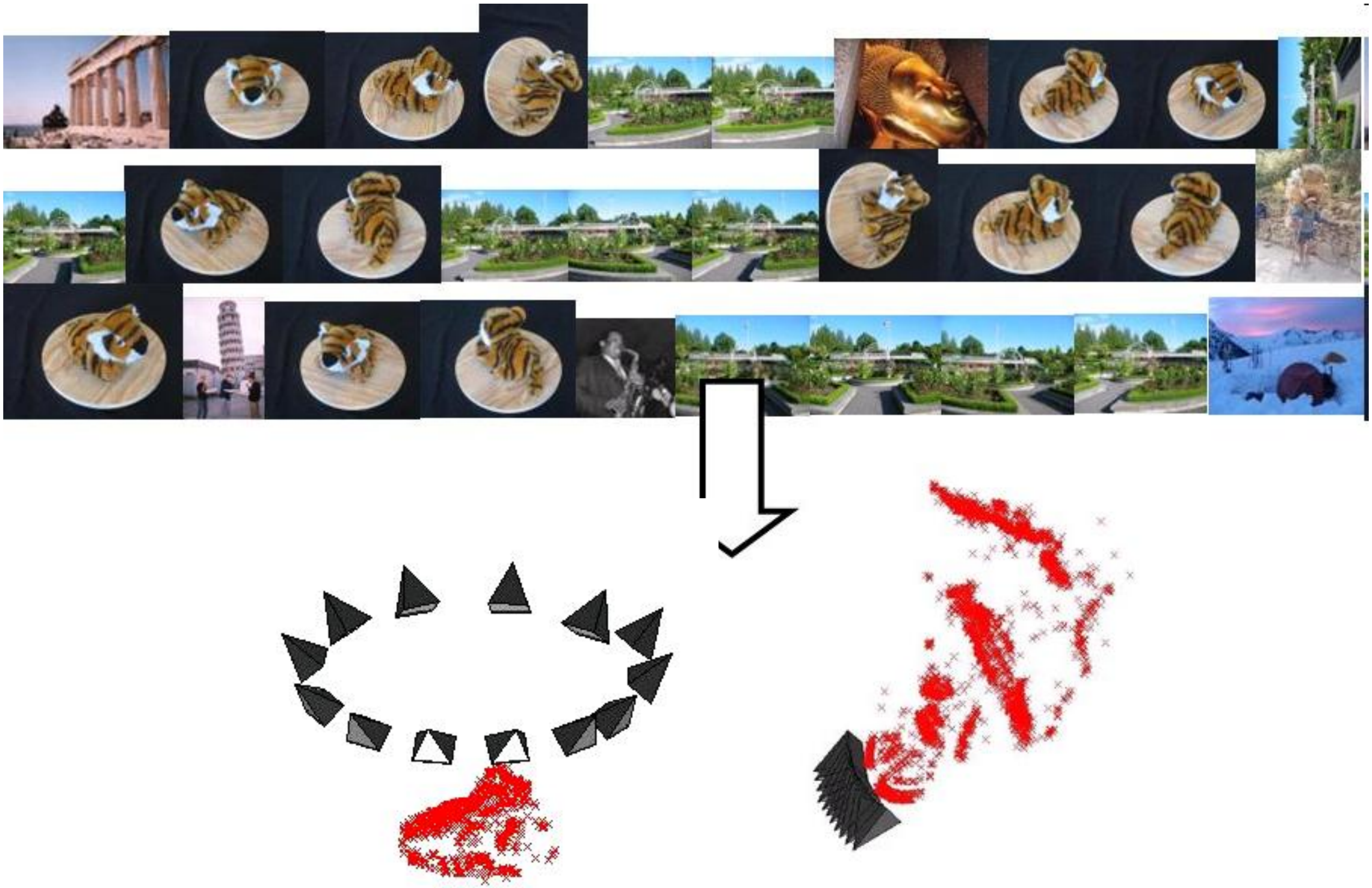


Applications

M. Pollefeys et al 98---

Applications

M. Brown and D. G. Lowe. Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets. *(3DIM2005)*



Photosynth

Noah Snavely, Steven M. Seitz, Richard Szeliski, "Photo tourism: Exploring photo collections in 3D," ACM Transactions on Graphics (SIGGRAPH Proceedings),2006,