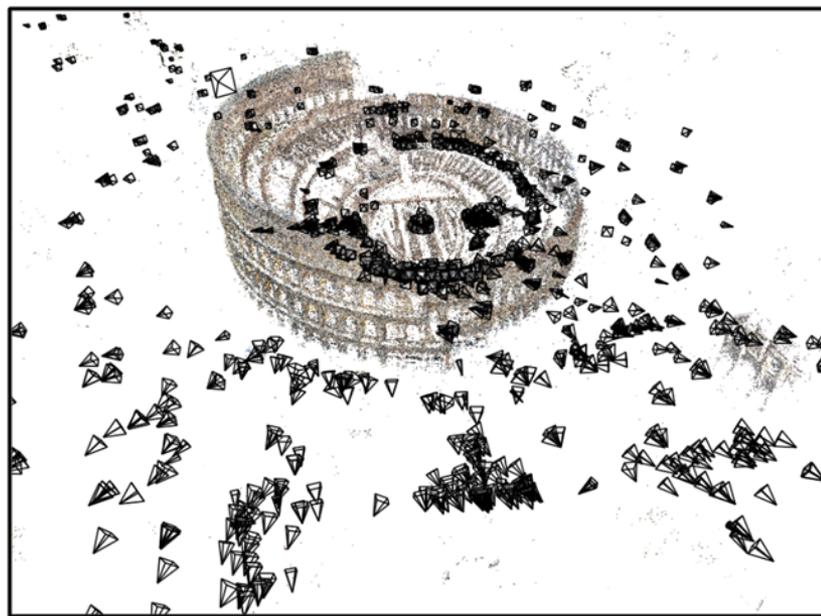
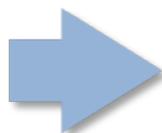


CS6670: Computer Vision

Noah Snavely

Lecture 11: Structure from motion



Announcements

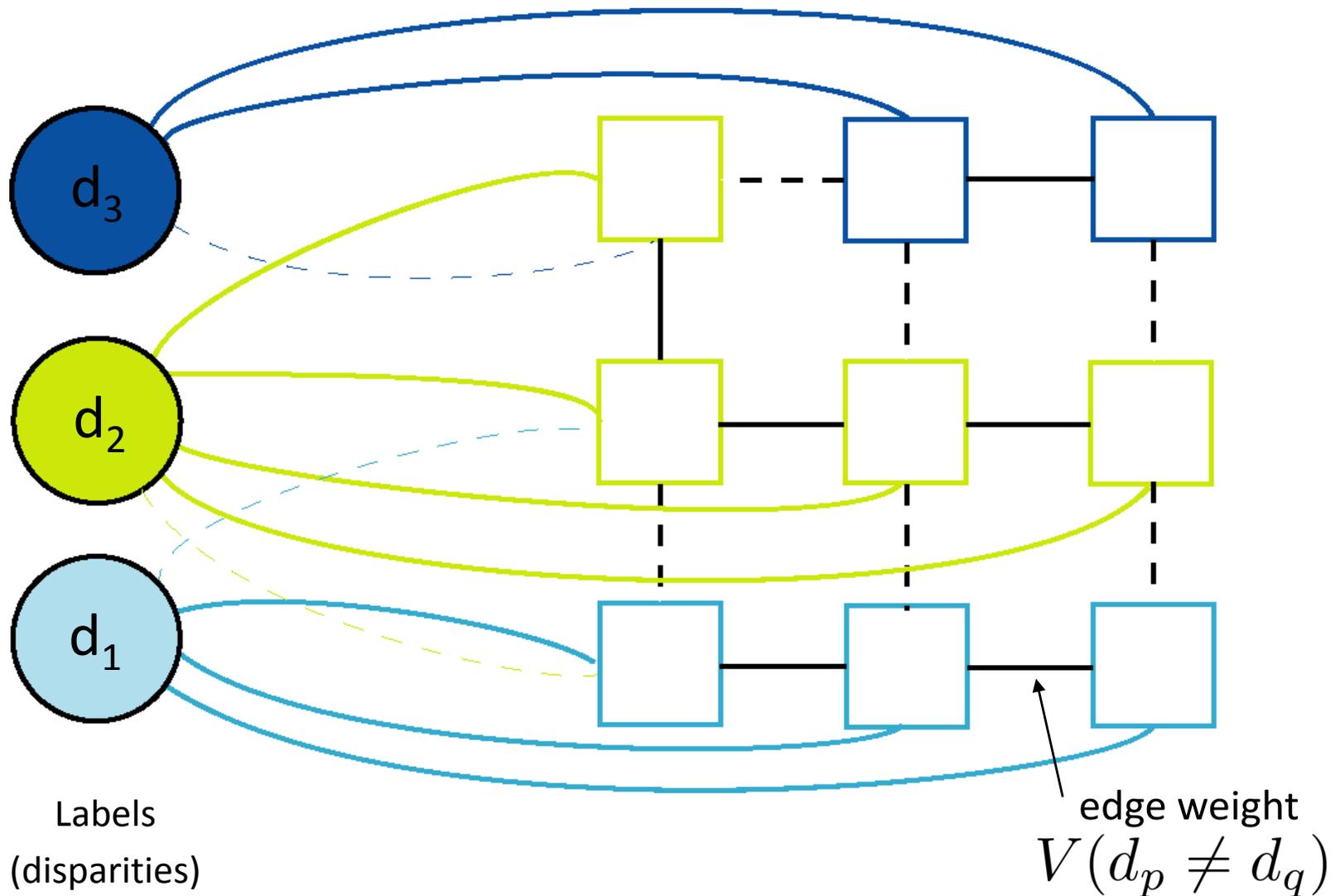
- Project 2 out, due next Wednesday, October 14
 - Artifact due Friday, October 16
- Questions?

Readings

- Szeliski, Chapter 7.2
- My thesis, Chapter 3

<http://www.cs.cornell.edu/~snaveley/publications/thesis/thesis.pdf>

Energy minimization via graph cuts



Other uses of graph cuts



Other uses of graph cuts



Agarwala et al., Interactive Digital Photo Montage, SIGGRAPH 2004

Panoramic stitching



- Problem: copy every pixel in the output image from *one* of the input images (without blending)
- Make transitions between two images seamless

Panoramic stitching

Input images:



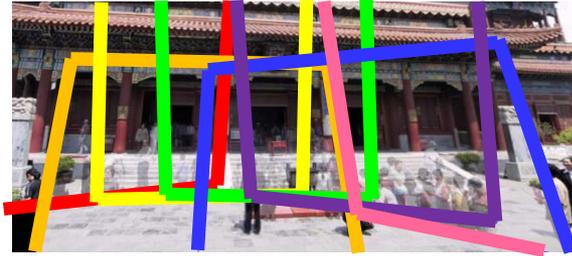
Output panorama:



- Can be posed as a labeling problem: label each pixel in the output image with one of the input images

Panoramic stitching

Input images:



Output panorama:



- Number of labels: $k =$ number of input images
- Objective function (in terms of labeling L)

$$E(L) = E_d(L) + \lambda E_s(L)$$

Panoramic stitching

$$E(L) = E_d(L) + \lambda E_s(L)$$

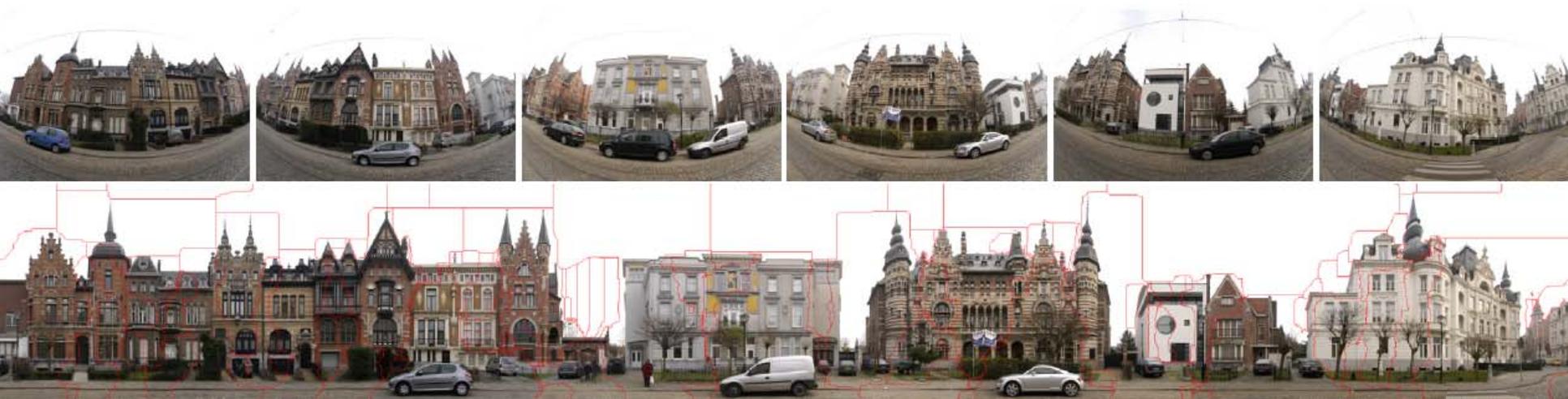
Infinite cost of labeling a pixel (x,y)
with image I if I doesn't cover (x,y)

Else, cost = 0

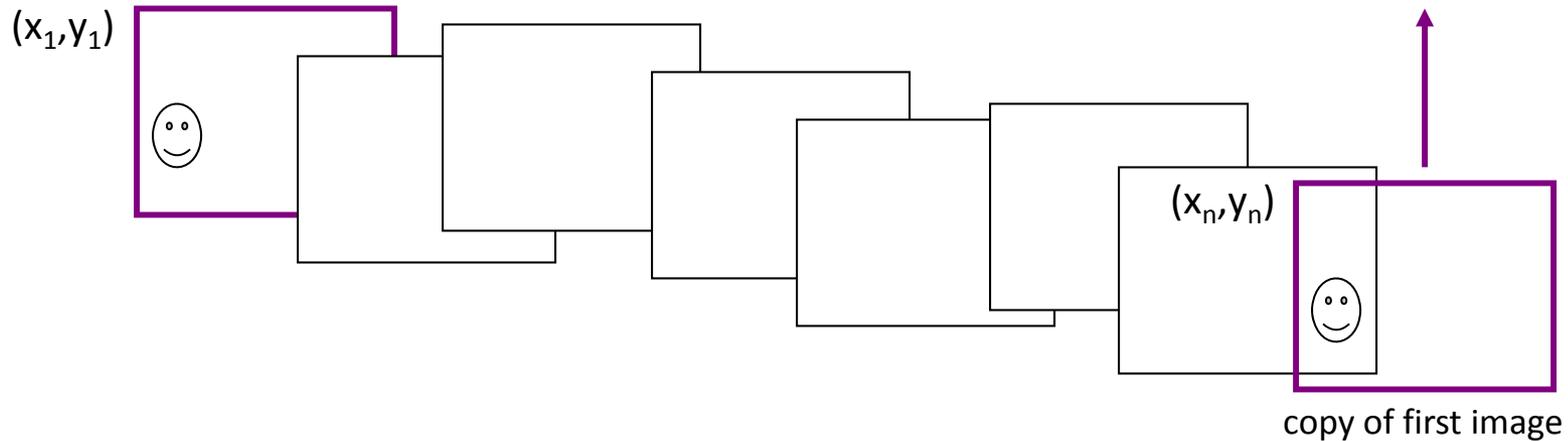
$$E_s(L) = \sum_{(p,q) \in \mathcal{E}} V(p, q, L(p), L(q))$$

$$V(p, q, L(p), L(q)) = \begin{cases} 0 & \text{if } L(p) = L(q) \\ \text{color similarity between } I_{L(p)} \text{ and } I_{L(q)} & \text{otherwise} \end{cases}$$

Photographing long scenes with multi-viewpoint panoramas

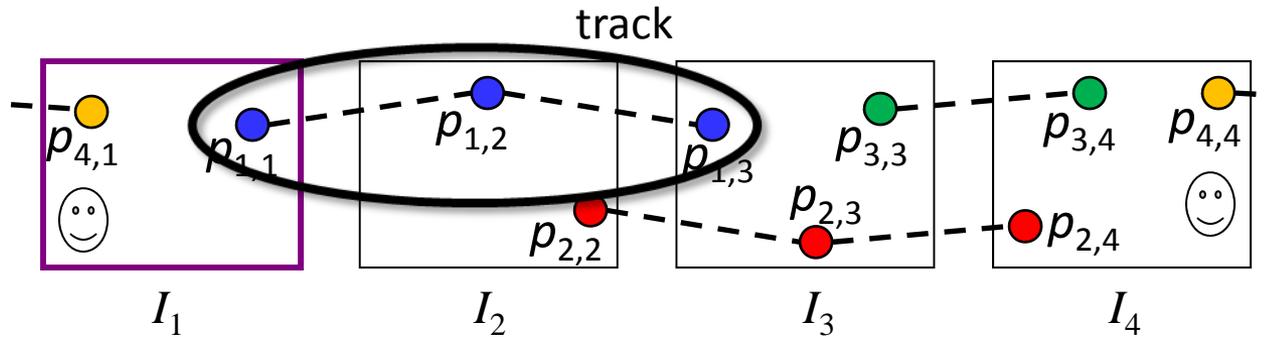


Today: Drift



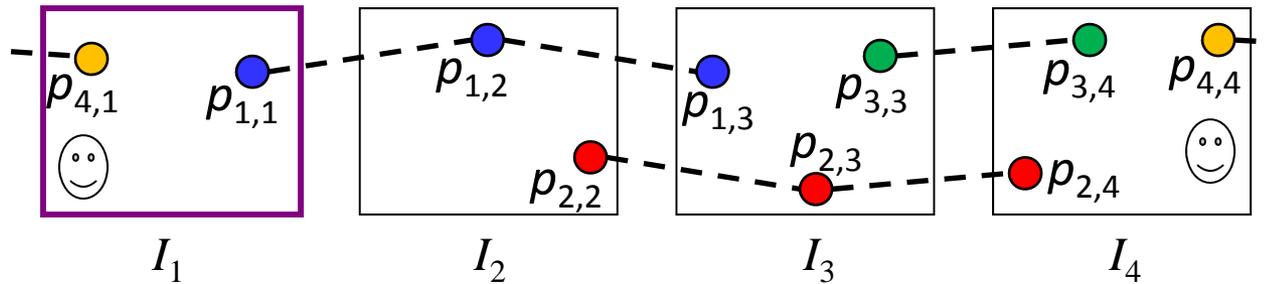
- add another copy of first image at the end
- this gives a constraint: $y_n = y_1$
- there are a bunch of ways to solve this problem
 - add displacement of $(y_1 - y_n)/(n - 1)$ to each image after the first
 - compute a global warp: $y' = y + ax$
 - run a big optimization problem, incorporating this constraint
 - best solution, but more complicated
 - known as “bundle adjustment”

Global optimization



- Minimize a global energy function:
 - What are the variables?
 - The translation $t_j = (x_j, y_j)$ for each image I_j
 - What is the objective function?
 - We have a set of matched features $p_{i,j} = (u_{i,j}, v_{i,j})$
 - We'll call these *tracks*
 - For each point match $(p_{i,j}, p_{i,j+1})$: $p_{i,j+1} - p_{i,j} = t_{j+1} - t_j$

Global optimization



$$p_{1,2} - p_{1,1} = t_2 - t_1$$

$$p_{1,3} - p_{1,2} = t_3 - t_2$$

$$p_{2,3} - p_{2,2} = t_3 - t_2$$

$$\dots$$

$$v_{4,1} - v_{4,4} = y_1 - y_4$$



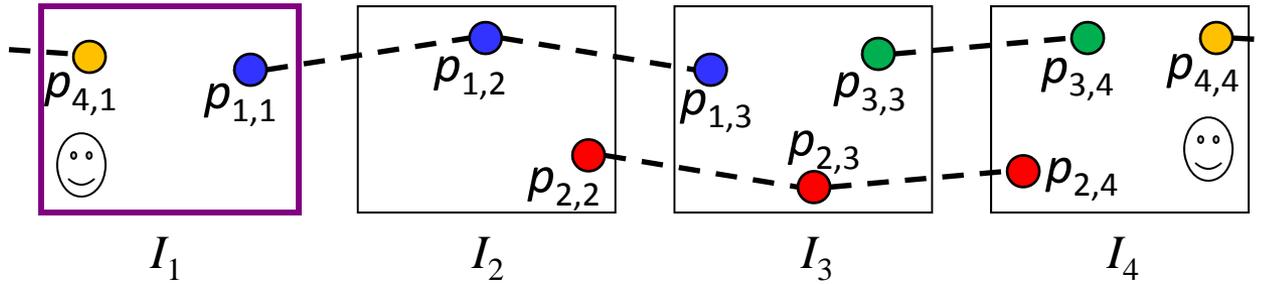
minimize

$$\sum_{i=1}^m \sum_{j=1}^{n-1} w_{ij} \cdot \left\| (p_{i,j+1} - p_{i,j}) - (t_{j+1} - t_j) \right\|^2$$

$$+ \sum_{i=1}^m w_{in} \cdot \left\| (v_{i,1} - v_{i,n}) - (y_1 - y_n) \right\|^2$$

$w_{ij} = 1$ if track i is visible in images j and $j+1$
 0 otherwise

Global optimization



$$\begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ & & & \dots & & & & \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \\ x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} u_{1,2} - u_{1,1} \\ v_{1,2} - v_{1,1} \\ \vdots \\ v_{4,1} - v_{4,4} \end{bmatrix}$$

A
2m x 2n

x
2n x 1

b
2m x 1

Global optimization

$$\begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ & & & \dots & & & & \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \\ x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} u_{1,2} - u_{1,1} \\ v_{1,2} - v_{1,1} \\ \vdots \\ v_{4,1} - v_{4,4} \end{bmatrix}$$

A

2m x 2n

x

2n x 1

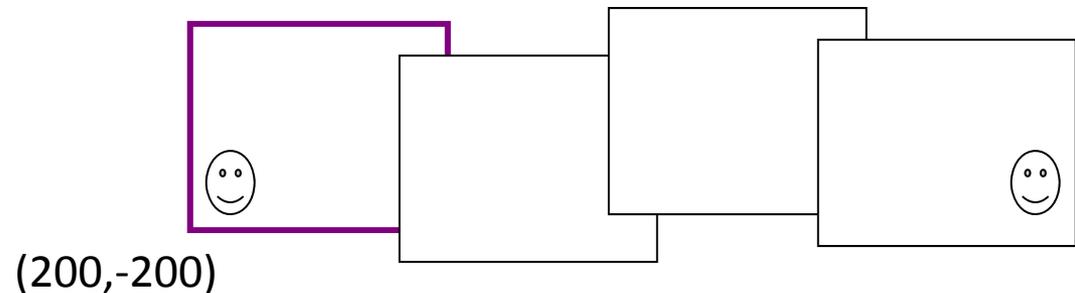
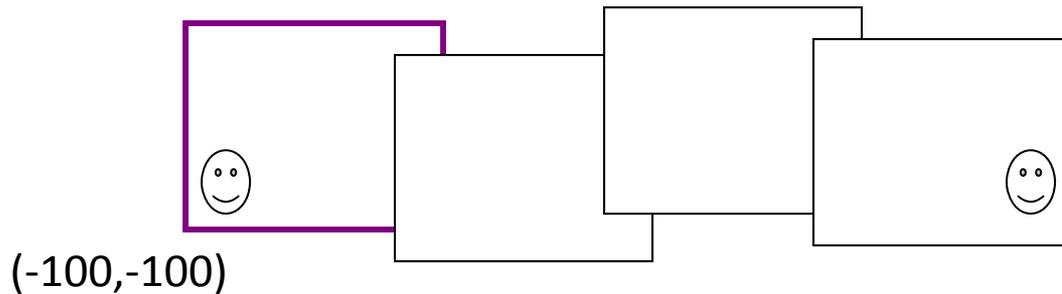
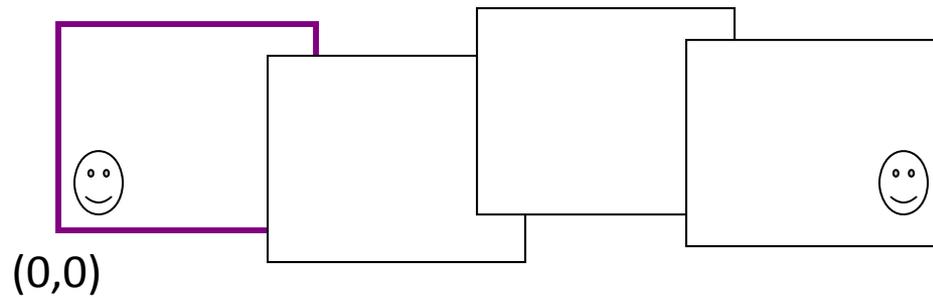
b

2m x 1

Defines a least squares problem: minimize $\|\mathbf{Ax} - \mathbf{b}\|$

- Solution: $\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$
- Problem: there is no unique solution for $\hat{\mathbf{x}}$! ($\det(\mathbf{A}^T \mathbf{A}) = 0$)
- We can add a global offset to a solution $\hat{\mathbf{x}}$ and get the same error

Ambiguity in global location

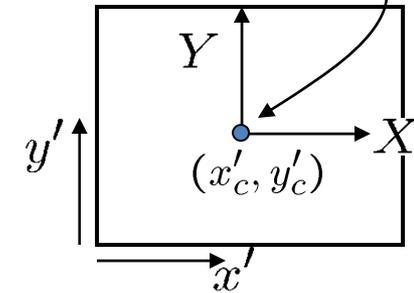


- Each of these solutions has the same error
- Called the *gauge ambiguity*
- Solution: fix the position of one image (e.g., make the origin of the 1st image $(0,0)$)

Solving for camera parameters

Recap: a camera is described by several parameters

- Translation \mathbf{t} of the optical center from the origin of world coords
- Rotation \mathbf{R} of the image plane
- focal length f , principle point (x'_c, y'_c) , pixel size (s_x, s_y)
- blue parameters are called “extrinsics,” red are “intrinsics”



Projection equation

$$p = \begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \underbrace{\begin{bmatrix} -fs_x & 0 & x'_c \\ 0 & -fs_y & y'_c \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}} (\mathbf{R}_{3 \times 3} \mathbf{x} + \mathbf{t}_{3 \times 1})$$

Solving for camera rotation

- Instead of spherically warping the images and solving for translation, we can directly solve for the rotation \mathbf{R}_j of each camera
- Can handle tilt / twist



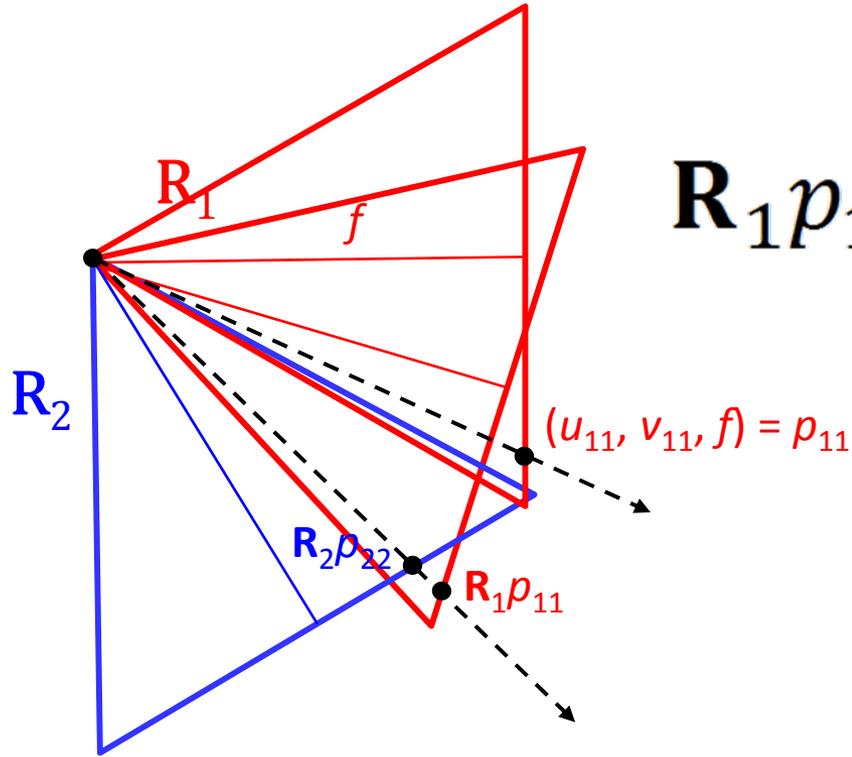
Solving for rotations

$$p_{11} = (u_{11}, v_{11}) \bullet$$

I_1

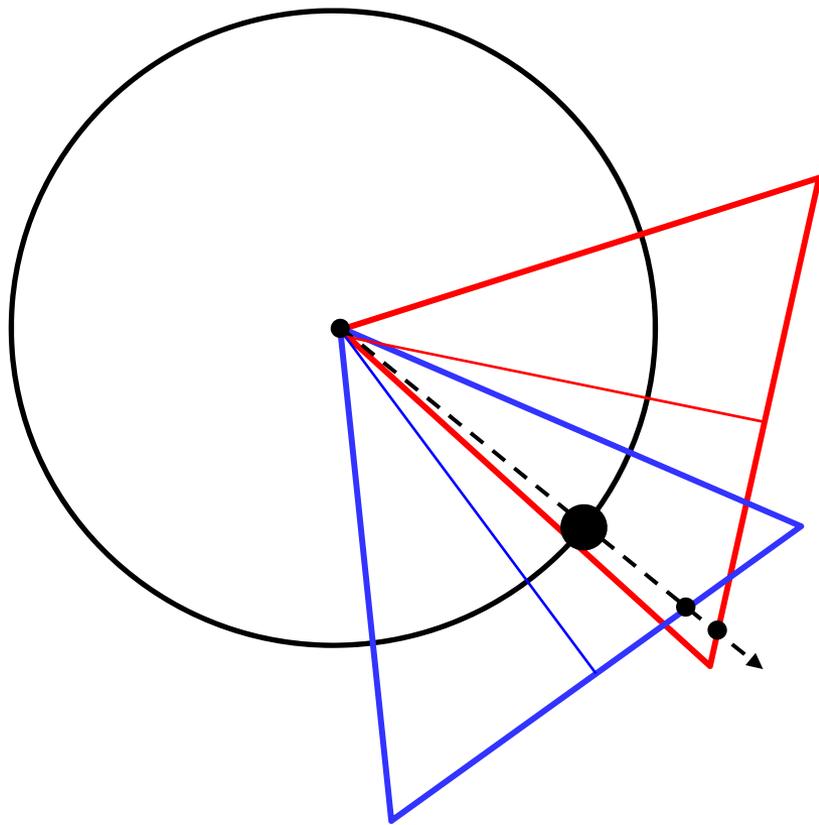
$$\bullet p_{12} = (u_{12}, v_{12})$$

I_2



$$R_1 p_{11} \cong R_2 p_{12}$$

Solving for rotations



$$\mathbf{R}_1 \mathbf{p}_{11} \cong \mathbf{R}_2 \mathbf{p}_{12}$$

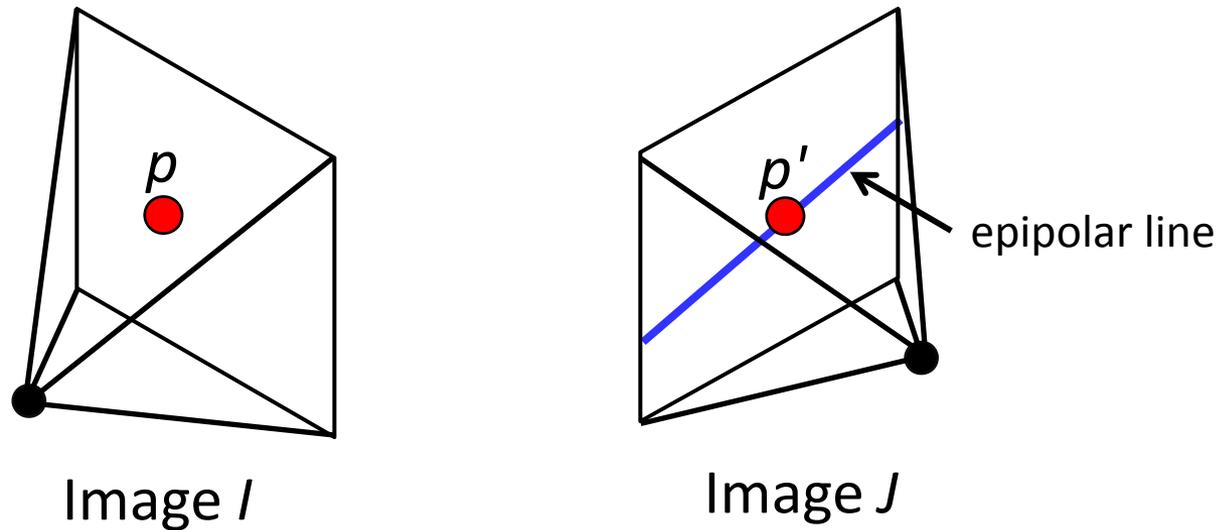
$$\mathbf{R}_1 \hat{\mathbf{p}}_{11} = \mathbf{R}_2 \hat{\mathbf{p}}_{12}$$

$$\text{minimize } \sum_{i=1}^m \sum_{j=1}^n w_{ij} \cdot \left\| \mathbf{R}_{j+1} \hat{\mathbf{p}}_{i,j+1} - \mathbf{R}_j \hat{\mathbf{p}}_{i,j} \right\|^2$$

3D rotations

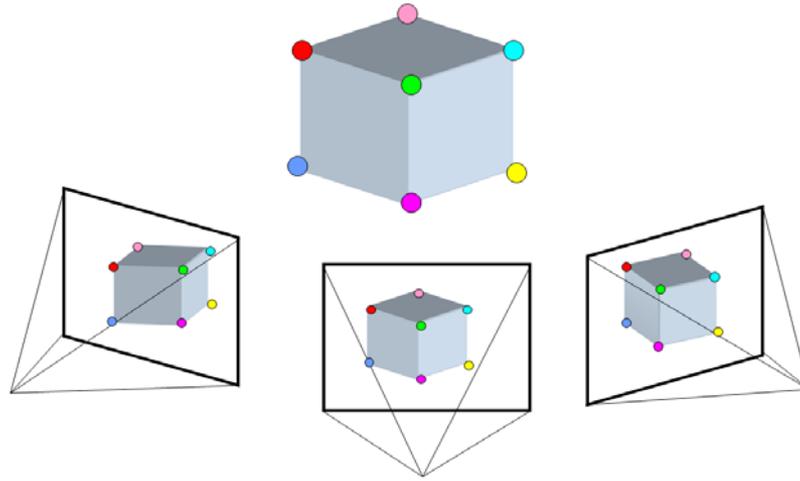
- How many degrees of freedom are there?
- How do we represent a rotation?
 - Rotation matrix (too many degrees of freedom)
 - Euler angles (e.g. yaw, pitch, and roll) – bad idea
 - Quaternions (4-vector on unit sphere)
- Usually involves non-linear optimization

Revisiting stereo



- How do we find the epipolar lines?
- Need to *calibrate* the cameras (estimate relative position, orientation)

Solving for rotations and translations



- *Structure from motion (SfM)*
- Unlike with panoramas, we often need to solve for *structure* (3D point positions) as well as *motion* (camera parameters)

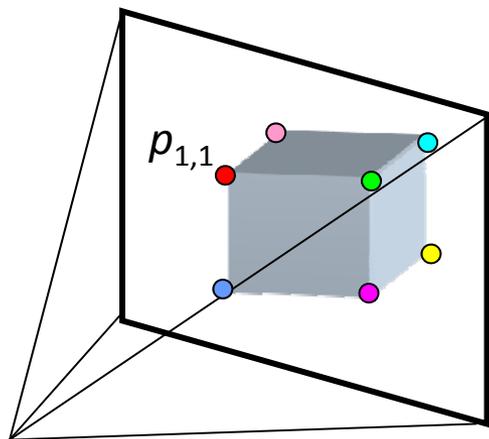
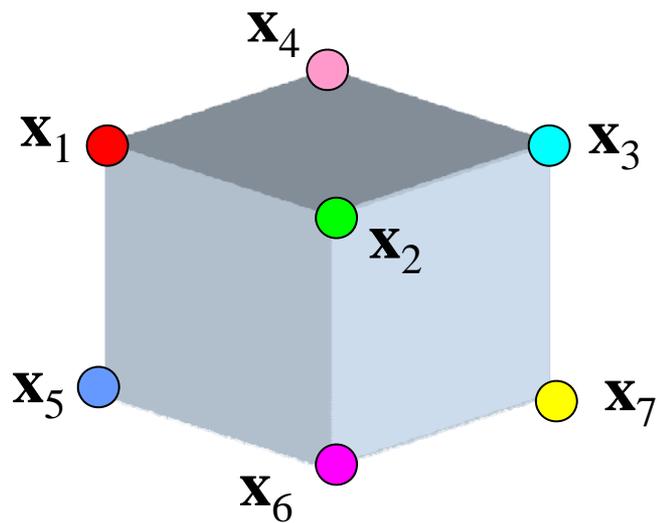


Image 1
 $\mathbf{R}_1, \mathbf{t}_1$

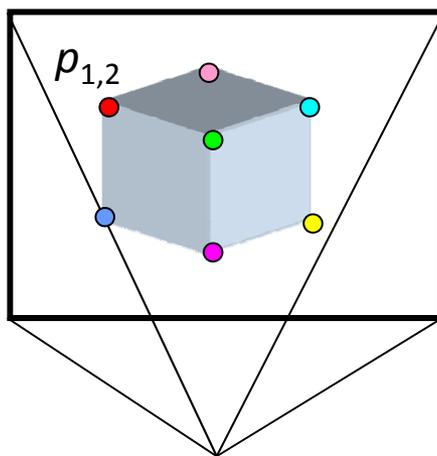


Image 2
 $\mathbf{R}_2, \mathbf{t}_2$

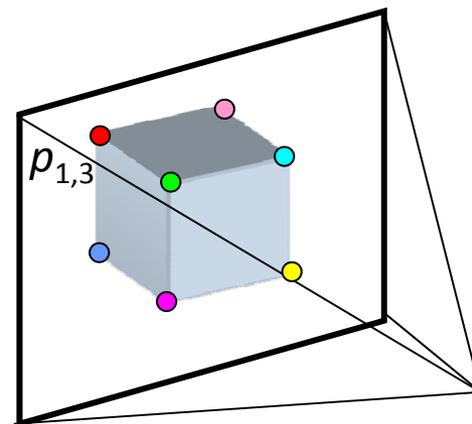
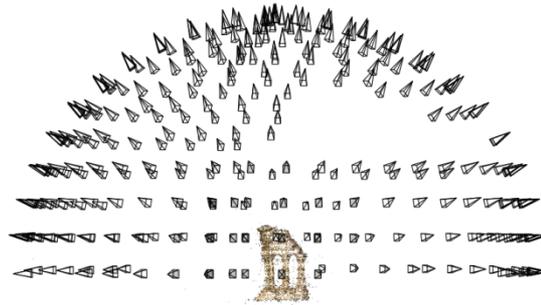
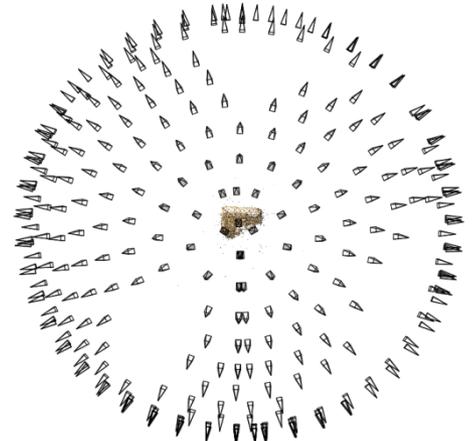


Image 3
 $\mathbf{R}_3, \mathbf{t}_3$

Structure from motion



Reconstruction (side)



(top)

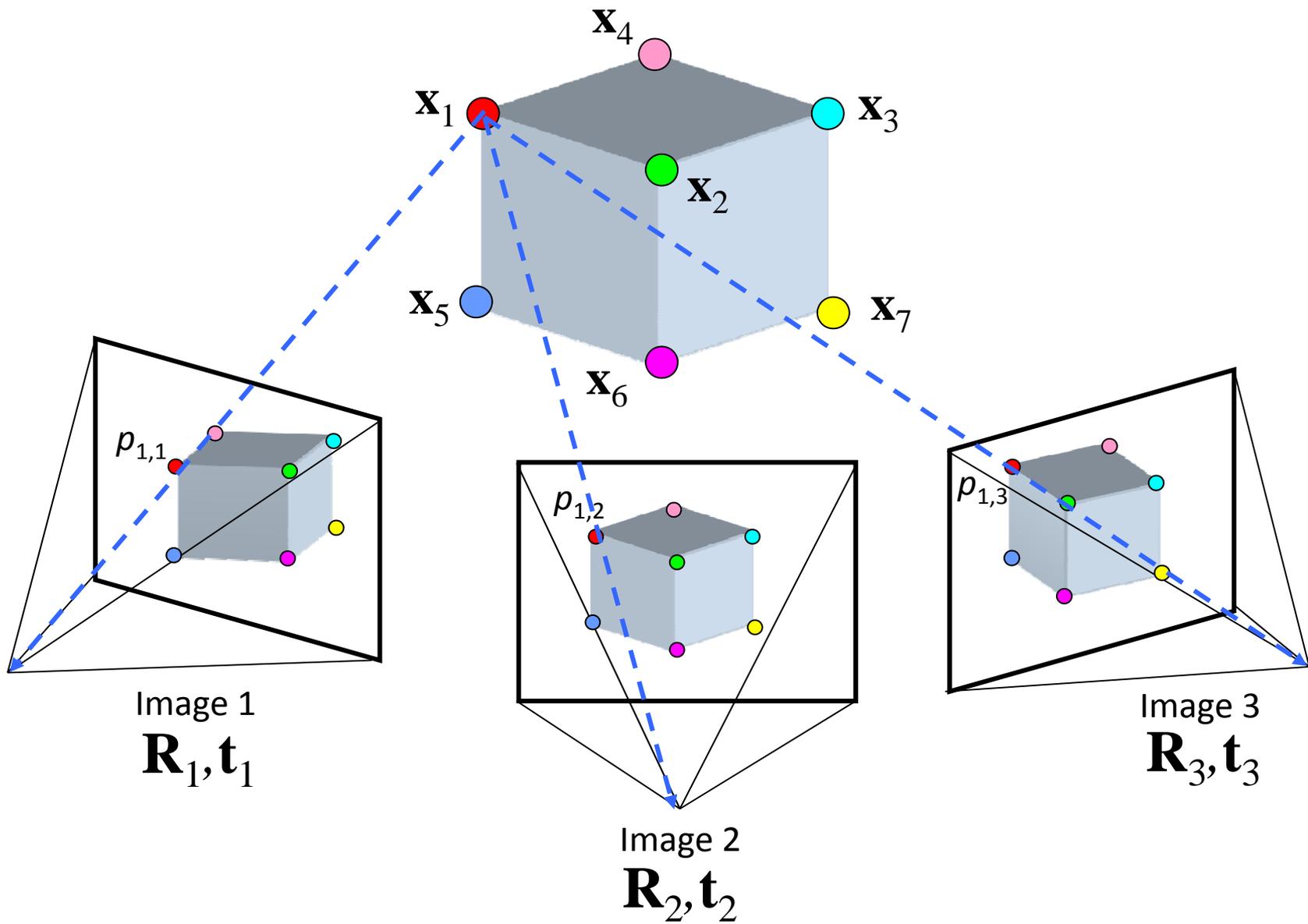
- Input: images with points in correspondence

$$p_{i,j} = (u_{i,j}, v_{i,j})$$

- Output

- structure: 3D location \mathbf{x}_i for each point p_i
- motion: camera parameters $\mathbf{R}_j, \mathbf{t}_j$

- Objective function: minimize *reprojection error*



SfM objective function

- Given point \mathbf{x} and rotation and translation \mathbf{R}, \mathbf{t}

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathbf{R}\mathbf{x} + \mathbf{t} \quad \begin{matrix} u' = \frac{fx'}{z'} \\ v' = \frac{fy'}{z'} \end{matrix} \quad \begin{bmatrix} u' \\ v' \end{bmatrix} = \mathbf{P}(\mathbf{x}, \mathbf{R}, \mathbf{t})$$

- Minimize sum of squared reprojection errors:

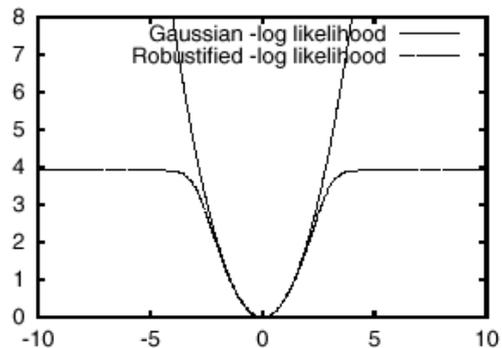
$$g(\mathbf{X}, \mathbf{R}, \mathbf{T}) = \sum_{i=1}^m \sum_{j=1}^n w_{ij} \cdot \left\| \underbrace{\mathbf{P}(\mathbf{x}_i, \mathbf{R}_j, \mathbf{t}_j)}_{\substack{\text{predicted} \\ \text{image location}}} - \underbrace{\begin{bmatrix} u_{i,j} \\ v_{i,j} \end{bmatrix}}_{\substack{\text{observed} \\ \text{image location}}} \right\|^2$$

Solving structure from motion

- Minimizing g is difficult
 - g is non-linear due to rotations, perspective division
 - lots of parameters: 3 for each 3D point, 6 for each camera
 - difficult to initialize
 - gauge ambiguity: error is invariant to a similarity transform (translation, rotation, uniform scale)
- Many techniques use non-linear least-squares (NLLS) optimization (*bundle adjustment*)
 - Levenberg-Marquardt is one common algorithm for NLLS
 - Lourakis, **The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the Levenberg-Marquardt Algorithm**,
<http://www.ics.forth.gr/~lourakis/sba/>
 - http://en.wikipedia.org/wiki/Levenberg-Marquardt_algorithm

Extensions to SfM

- Can also solve for intrinsic parameters (focal length, radial distortion, etc.)
- Can use a more robust function than squared error, to avoid fitting to outliers



- For more information, see: Triggs, *et al*, “Bundle Adjustment – A Modern Synthesis”, *Vision Algorithms* 2000.

Photo Tourism

- Structure from motion on Internet photo collections

The screenshot shows the Flickr search interface. At the top, the Flickr logo is on the left, and navigation links for Home, The Tour, Sign Up, and Explore are in the center. On the right, there's a search bar with the text 'Search everyone's photos' and a 'Search' button. Below the navigation, the search results are displayed. The search term 'trevi fountain' is entered in the search bar. The results section shows '27,761 results for photos matching trevi and fountain.' Below this, there are several photo thumbnails arranged in a grid. Each thumbnail has a caption indicating the user who uploaded it. On the right side of the page, there are sponsored results for 'Cruise The Mediteranian', 'Trevi Fountain Hotel', 'Trevi Fountains', and 'Hotels Trevi Fountain - Italy'.

flickr LOVES YOU™

You aren't signed in [Sign In](#) [Help](#)

[Home](#) [The Tour](#) [Sign Up](#) [Explore](#)

Search everyone's photos [Search](#)

Search [Photos](#) [Groups](#) [People](#)

trevi fountain [SEARCH](#) [Advanced Search](#)
[Search by Camera](#)

Full text Tags only

✓ We found **27,761 results** for photos matching **trevi** and **fountain**. [View as slideshow](#) (85)

View: [Most relevant](#) • [Most recent](#) • [Most interesting](#) Show: [Details](#) • [Thumbnails](#)

 From [NupurB](#)

 From [BRUNO MÉNDEZ...](#)

 From [Toshio](#)

 From [Maqik &...](#)

 From [Christopher...](#)

 From [Nastrina1981](#)

 From [jaki good](#)

 From [Fabrizio...](#)

 From [Skazi](#)

 From [Alida's...](#)

 From [hbomb1947](#)

 From [David d'O](#)

Sponsored Results

[Cruise The Mediteranian](#)
Enjoy 11 days in Spain, France and Italy on this Mediteranian tour.
[www.mayflowertours.com/medtour](#)

[Trevi Fountain Hotel](#)
Compare hotel prices from over 120 top websites and save up to 70%.
[Hotels.SideStep.com](#)

[Trevi Fountains](#)
Compare prices, brands, and more at Smarter.com.
[www.smarter.com](#)

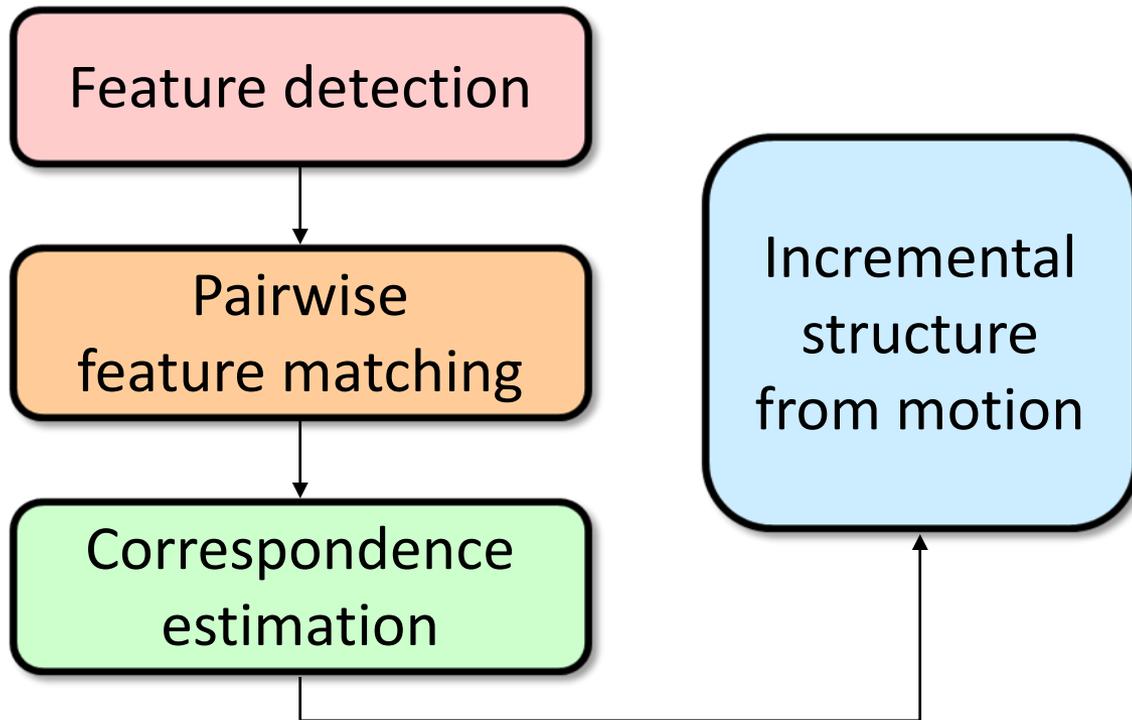
[Hotels Trevi Fountain - Italy](#)
Book now and save up to 75%. No reservation fee and pay at your hotel.
[booking.com/hotels-trevi-Fountain](#)

[Fountain Trevi](#)
Enjoy Savings & Selection On **Fountain Trevi**.
[Shopzilla.com](#)

Photo Tourism

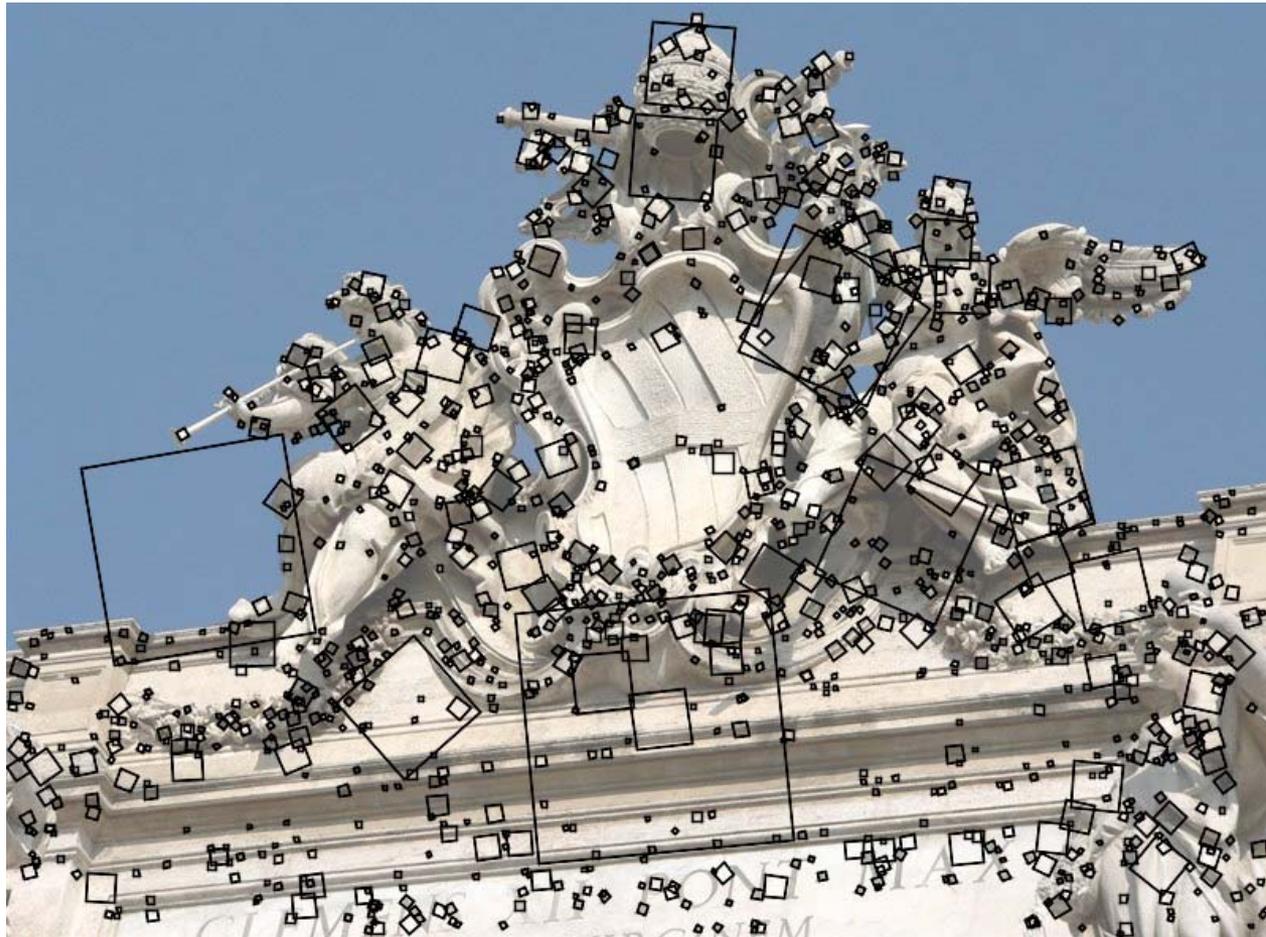


Scene reconstruction



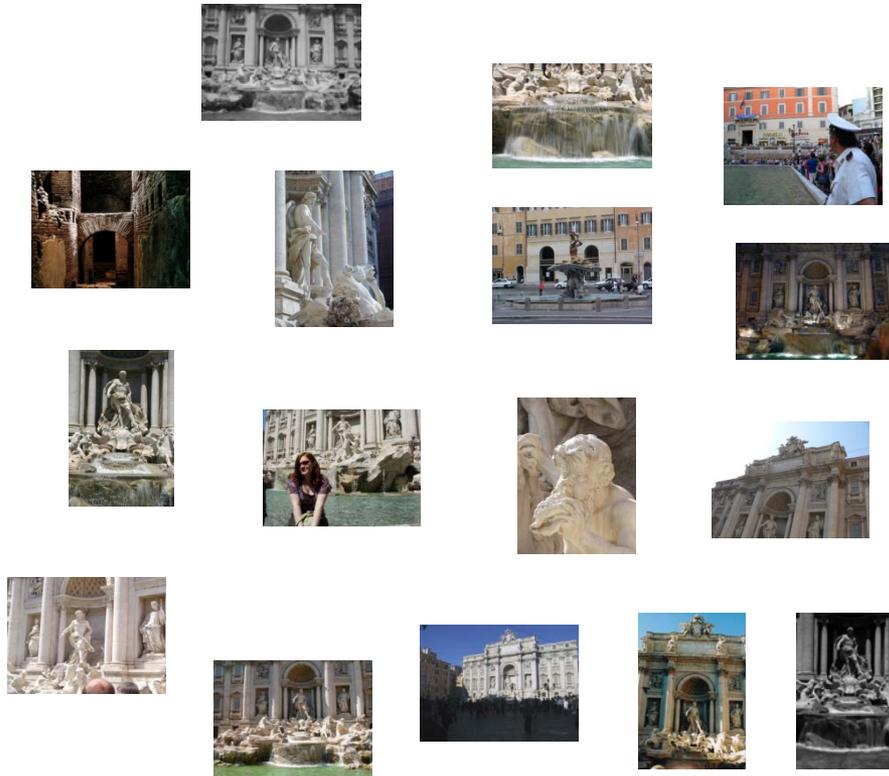
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



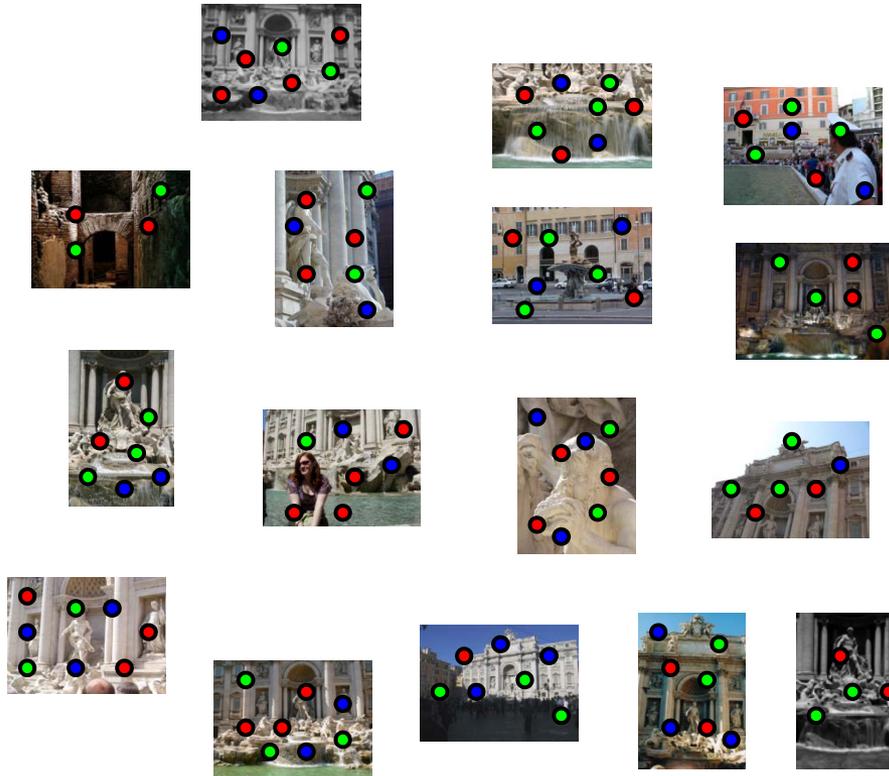
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



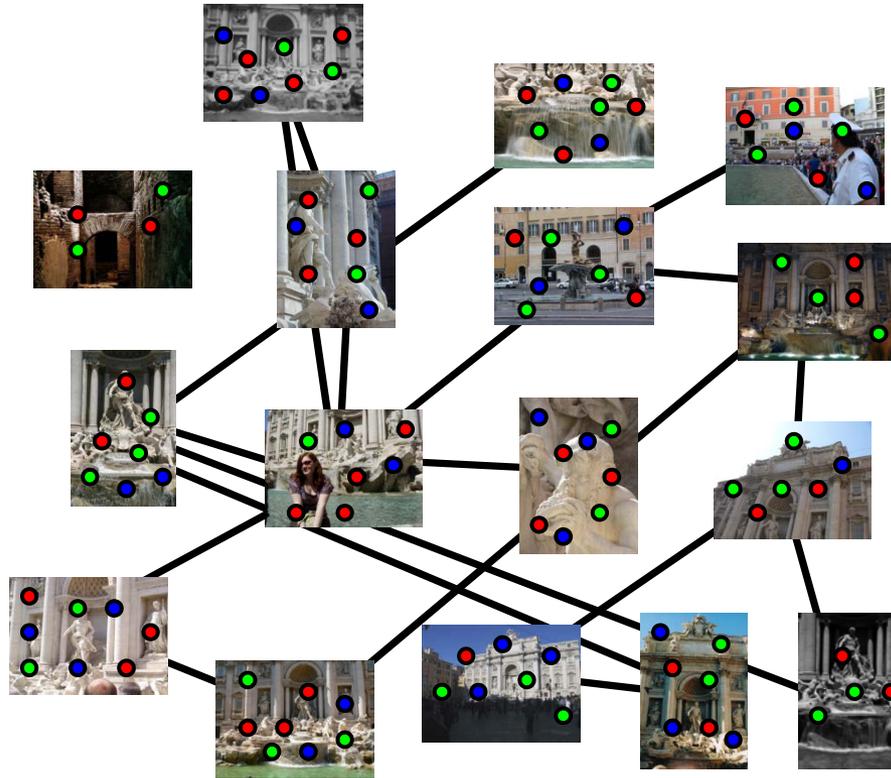
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



Feature matching

Match features between each pair of images



Feature matching

Refine matching using RANSAC [Fischler & Bolles 1987] to estimate fundamental matrices between pairs

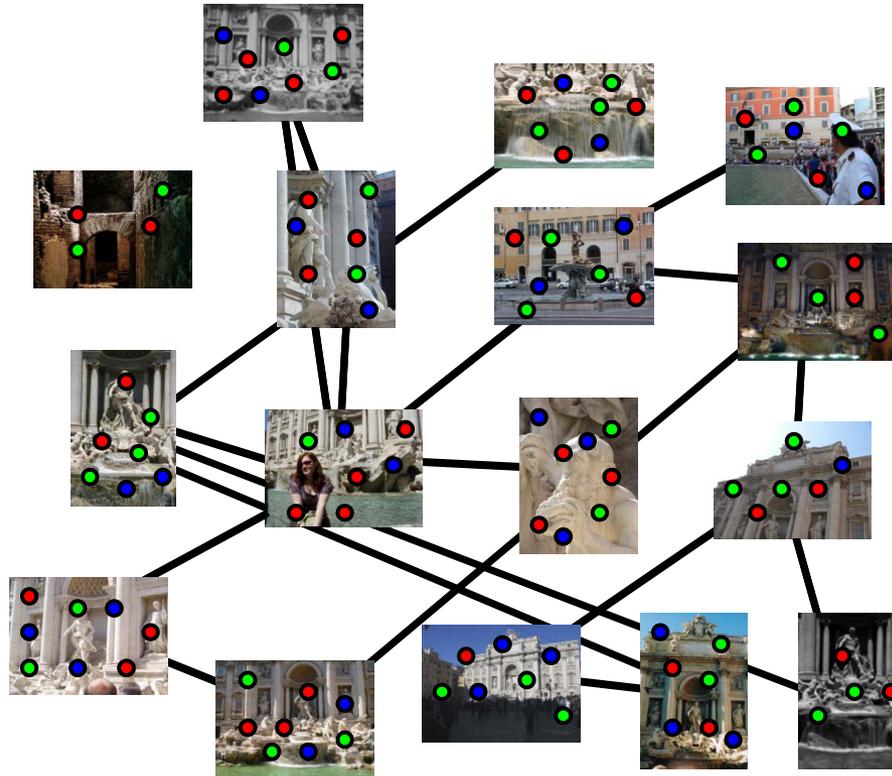
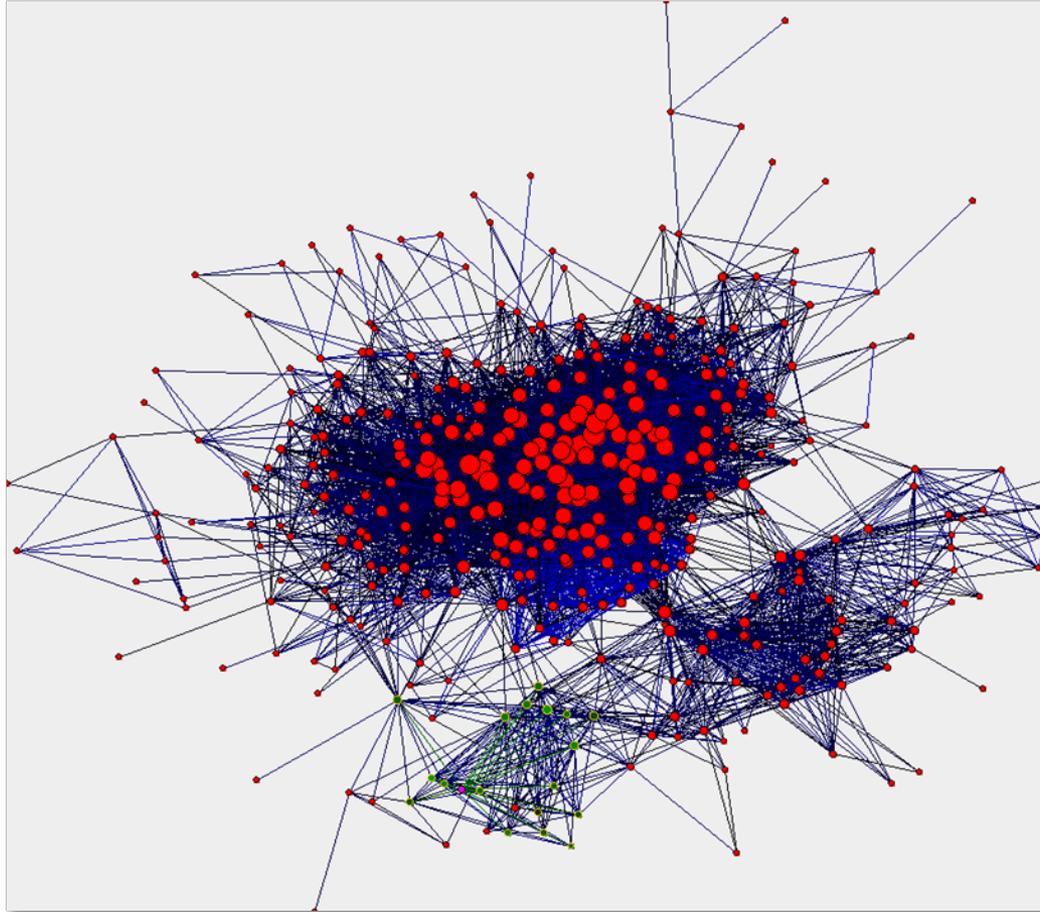
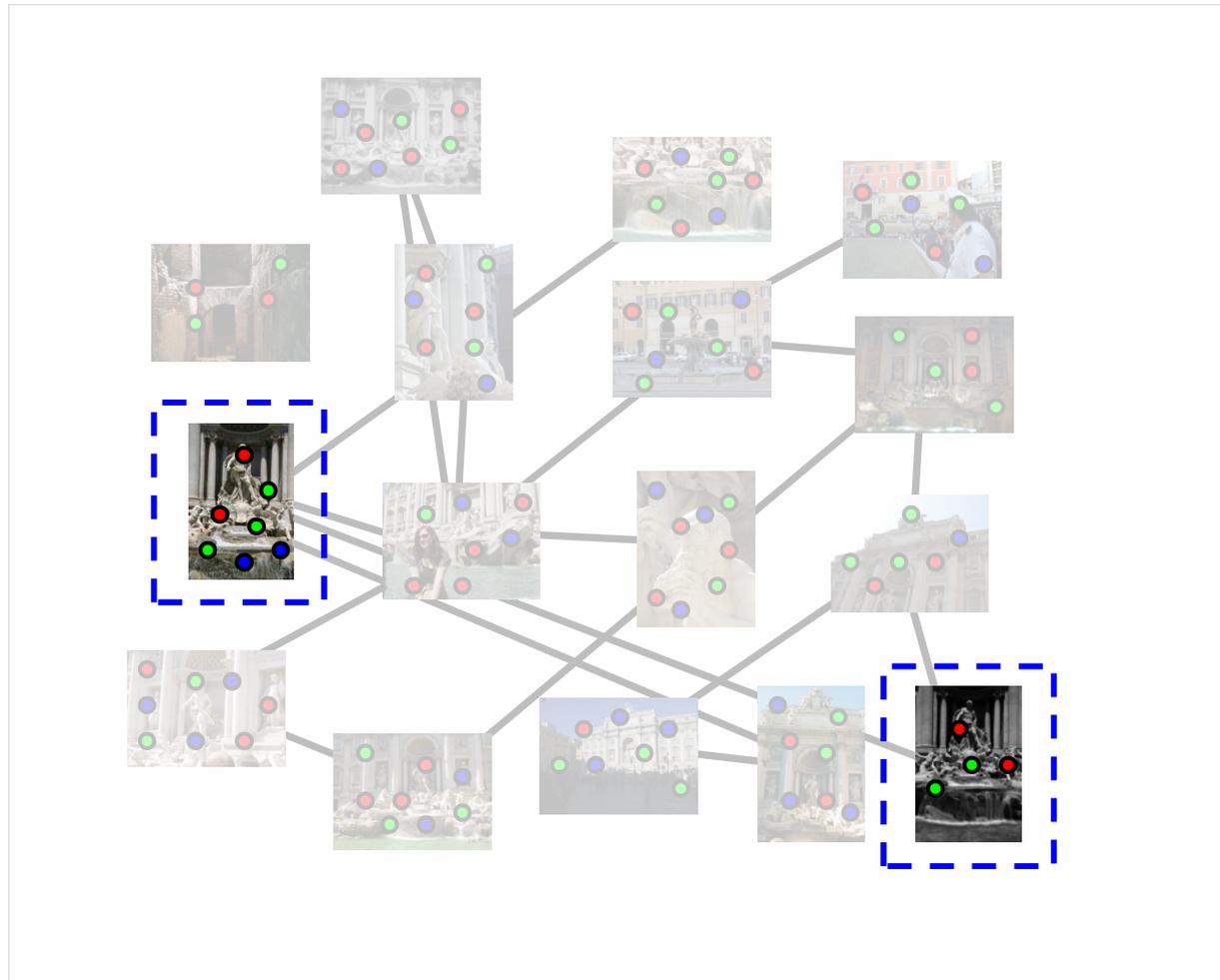


Image connectivity graph



(graph layout produced using the Graphviz toolkit: <http://www.graphviz.org/>)

Incremental structure from motion



Incremental structure from motion



Incremental structure from motion

