Announcements

Midterm: out by the end of the week
Project 1 artifact winners

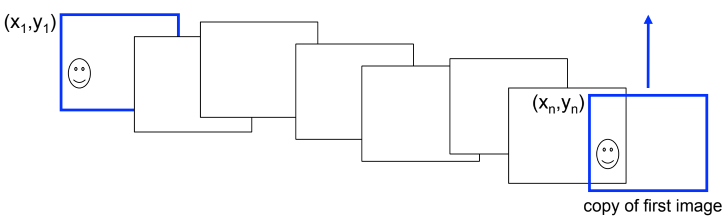
Global Alignment and Structure from Motion

(Adapted from slides by Noah Snavely)

Today's Readings

- Photo Tourism (Snavely et al., SIGGRAPH 2006)
  - http://phototour.cs.washington.edu/Photo_Tourism.pdf

Problem: Drift

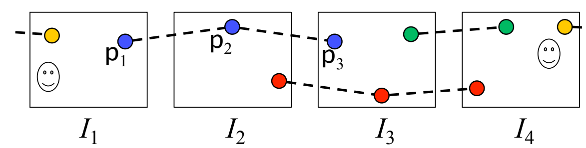


Solution

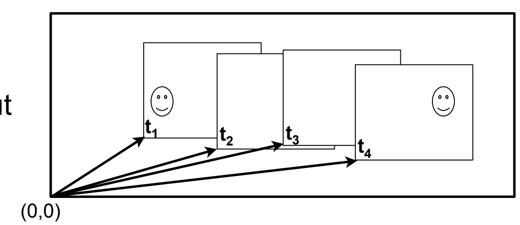
- add another copy of first image at the end
- this gives a constraint: $y_n = y_1$
- there are a bunch of ways to solve this problem
  - add displacement of $(y_1 - y_n)/(n - 1)$ to each image after the first
  - compute a global warp: $y' = y + ax$
  - run a big optimization problem, incorporating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

Global optimization

Input



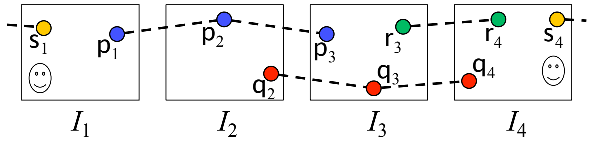
Output



We want to estimate $\mathbf{t}_i$. We know $\mathbf{p}_i$

- how do $\mathbf{t}_i$ relate to $\mathbf{p}_i$?

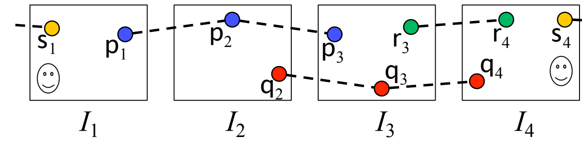
Global optimization



Recipe

1. Identify the **variables** you want to estimate
   - in our case: $\mathbf{t_i}$
2. Identify a set of **objectives** you want to satisfy
   - in our case: $\mathbf{p_i} - \mathbf{p_j} = \mathbf{t_i} - \mathbf{t_j}$ and similar for **q**, **r**, **s**
3. Define an **objective function** F over these variables, whose minimum occurs at the "answer" for these variables
4. Find the **minimum** of F

Objective function

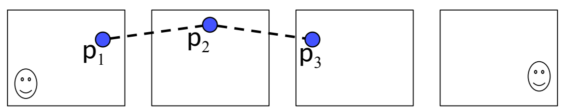


Objective function

$$\sum_{i=2}^{3} \|(\mathbf{p_i} - \mathbf{p_{i-1}}) - (\mathbf{t_i} - \mathbf{t_{i-1}})\|^2$$

+ similar terms for **q**, **r**, **s**

Objective function



Objective function $\sum_{i=2}^{3} \|(\mathbf{p_i} - \mathbf{p_{i-1}}) - (\mathbf{t_i} - \mathbf{t_{i-1}})\|^2$

$\mathbf{t_i} = (x_i, y_i)$

$\mathbf{p_i} = (u_i, v_i)$

Matrix form

$$\begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \\ x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} u_2 - u_1 \\ v_2 - v_1 \\ u_3 - u_2 \\ v_3 - v_2 \end{bmatrix}$$

Objective Function

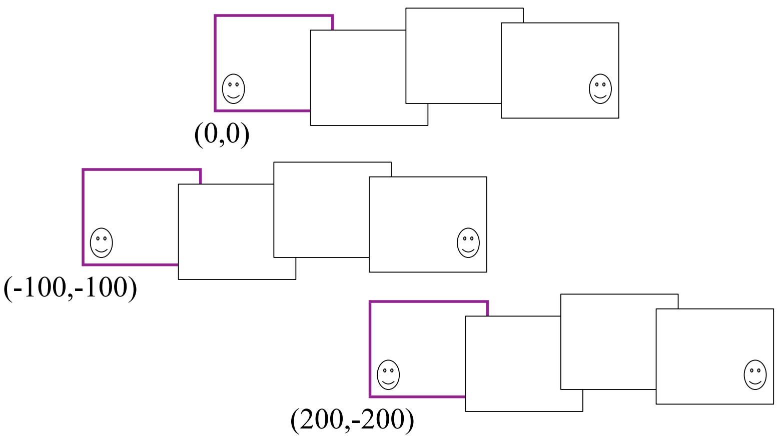
Adding in **q**, **r**, **s** give a larger matrix equation

$$\underbrace{\begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ & & & \vdots & & & & \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \\ x_4 \\ y_4 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} u_2 - u_1 \\ v_2 - v_1 \\ u_3 - u_2 \\ v_3 - v_2 \\ \vdots \end{bmatrix}}_{\mathbf{b}}$$

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

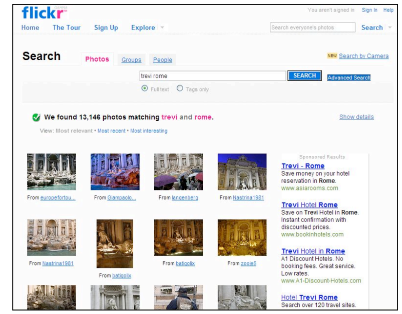
- Solution: $\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$
- Problem: there are multiple solutions for $\hat{\mathbf{x}}$! $(\det(\mathbf{A}^T\mathbf{A}) = 0)$
- We can add a global offset to a solution $\hat{\mathbf{x}}$ and get the same error

Ambiguity in the solution



- Each of these solutions has the same error
- Called the gauge ambiguity
- Solution: fix the translation of one image ($\mathbf{t}_1$ = (0,0))

Structure from motion

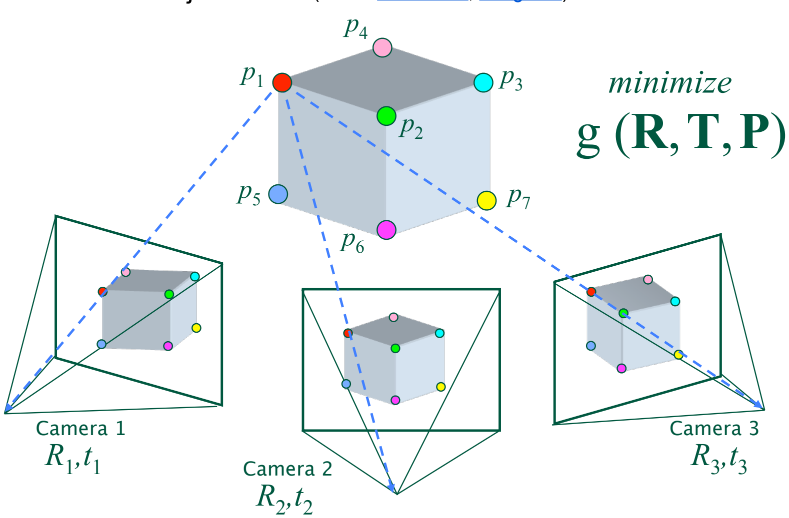


Images on the Internet

Computed 3D structure

Structure from motion

aka "bundle adjustment" (texts: Zisserman; Faugeras)



SfM objective function

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

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

Minimize sum of squared reprojection errors:

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

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 optimization (*bundle adjustment*)

- Levenberg-Marquardt is a popular algorithm
- http://en.wikipedia.org/wiki/Levenberg-Marquardt_algorithm

Good code online

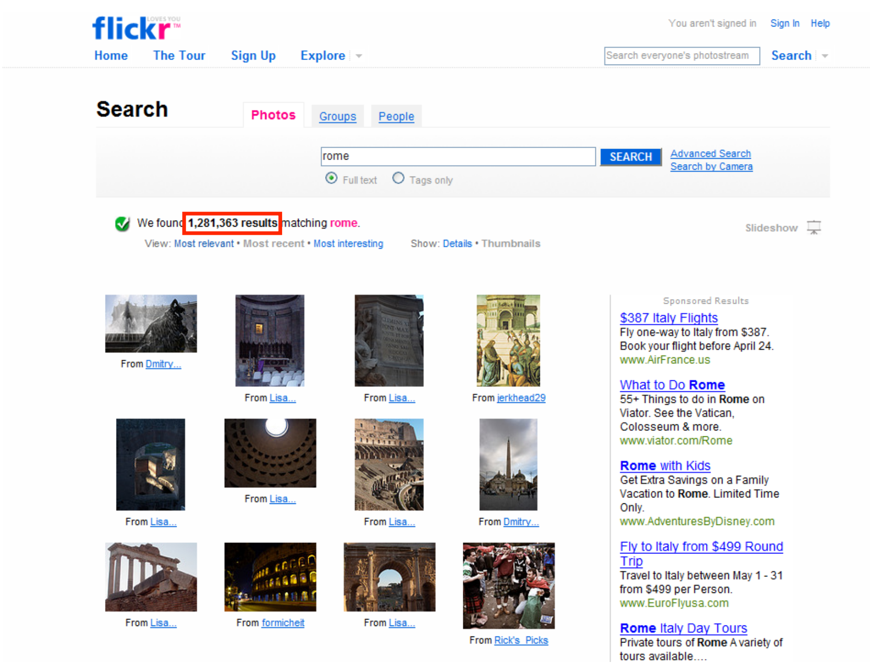
- Bundler: http://phototour.cs.washington.edu/bundler/
- Multicore: http://grail.cs.washington.edu/projects/mcba/

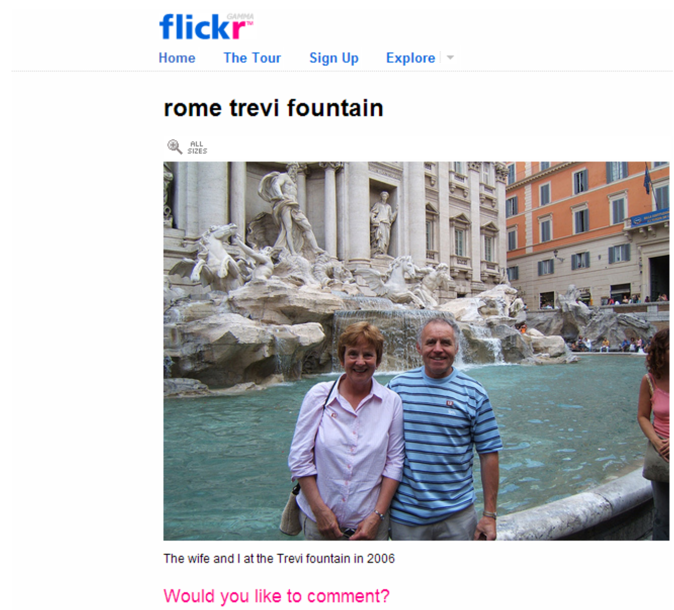
Photo Tourism

Photo tourism video: http://www.youtube.com/watch?v=5Ji84zb2r8s

Microsoft Photosynth: http://photosynth.net/

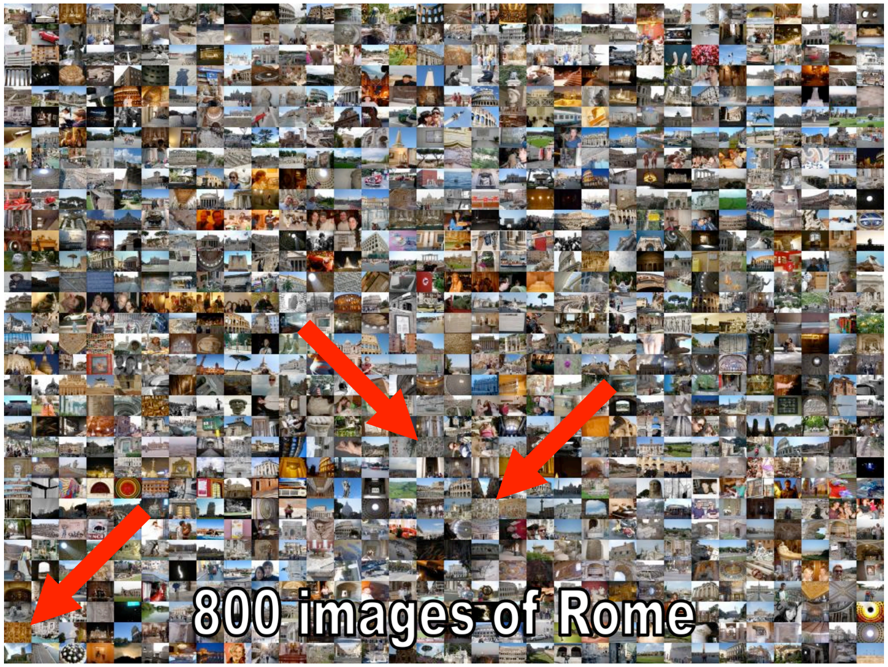
Google Photo Tours: http://maps.google.com/phototours

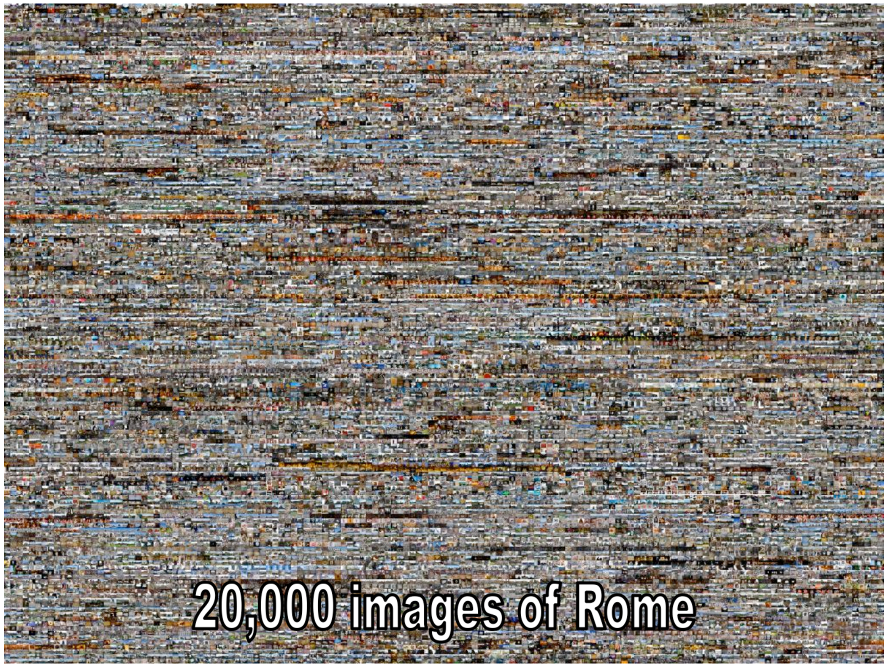





50 images of Rome


200 images of Rome


800 images of Rome

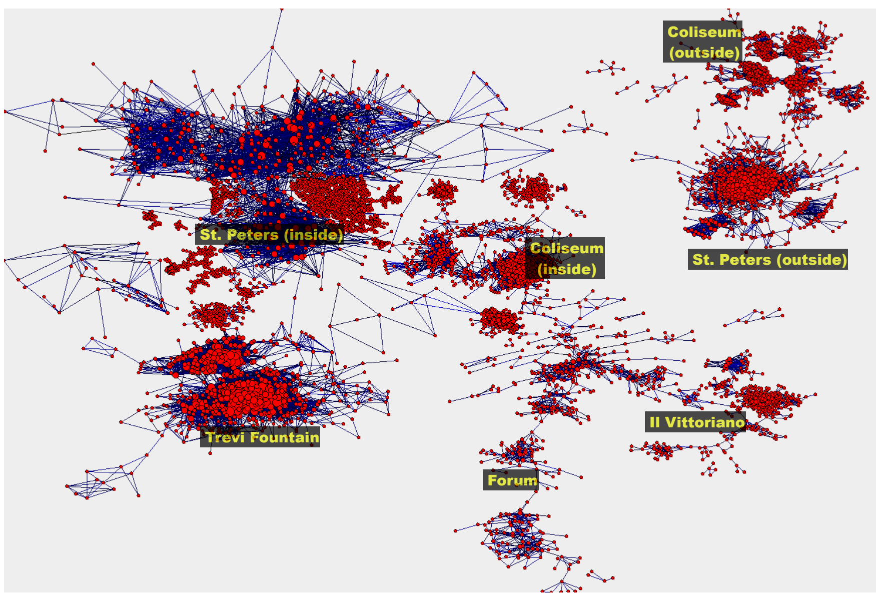

20,000 images of Rome

Reconstructing Rome

In a day...

From ~1M images
Using ~1000 cores

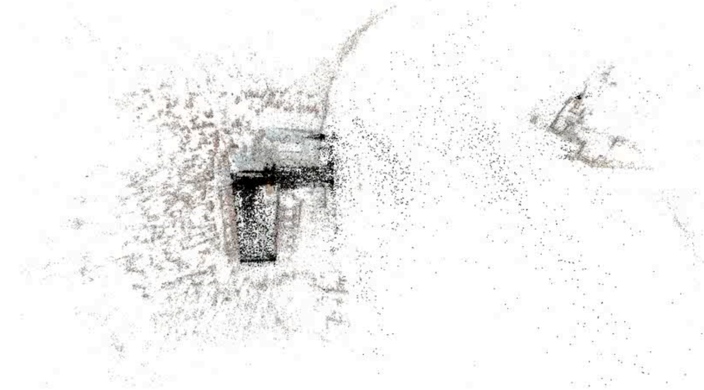
Sameer Agarwal, Noah Snavely, Rick Szeliski, Steve Seitz
http://grail.cs.washington.edu/rome



Rome 150K: Colosseum

Rome: St. Peters

Venice (250K images)



Venice: Canal

Dubrovnik

More info

- Rome-in-a-day page
  - http://grail.cs.washington.edu/rome