



Photo Tourism ++

Computer Vision
 CSE P 576, Spring 2011
 Richard Szeliski
 Microsoft Research



Photo Tourism ++

Richard Szeliski
 Microsoft Research
 Interactive Visual Media Group

Noah Snavely **Steve Seitz, Ian Simon, Sameer Agarwal**
 Cornell University University of Washington &
 Microsoft & Google

The world in photos

- There are **billions** of photos online
- Photographic record of the surface of the earth
- Photo sharing on a massive scale



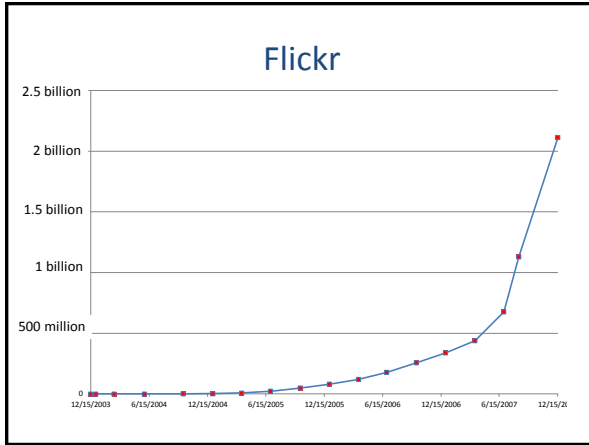


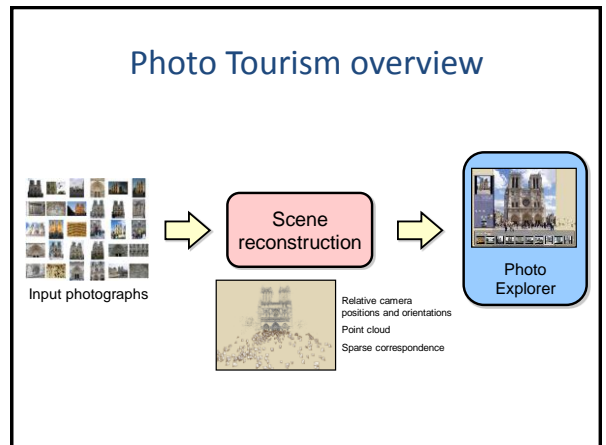
Photo Sharing

How can we manipulate (visualize, navigate, find) such immense numbers of images?

Photo Tourism

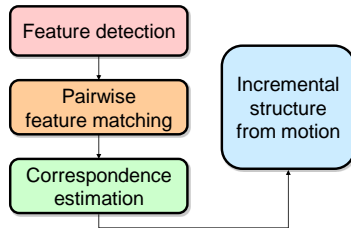
[Snavely, Seitz, Szeliski, SIGGRAPH 2006]

Images on the Internet
Computed 3D structure

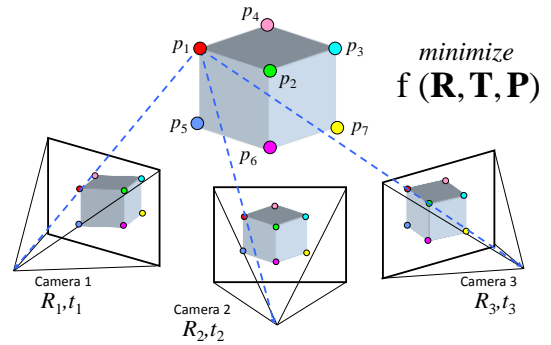


Scene reconstruction

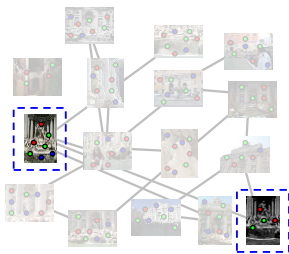
- Automatically estimate
 - position, orientation, and focal length of cameras
 - 3D positions of feature points



Structure from motion



Incremental structure from motion



- Automatically select an initial pair of images

Incremental structure from motion



Navigation: Prague Old Town Square



Annotations: Notre Dame



Hierarchical annotations



Locking the camera (stabilization)





Microsoft Photosynth

<http://photosynth.net/>

Microsoft Photosynth

- 3D reconstruction
- Multi-resolution streaming & zooming
- Quad-based exploration
- Community photo sharing

<http://photosynth.net/>

Skeletal graphs for efficient structure from motion

Noah Snavely, Steve Seitz,
Richard Szeliski
CVPR'2008

Large-scale reconstruction

- Most of the models shown so far have had ~500 images

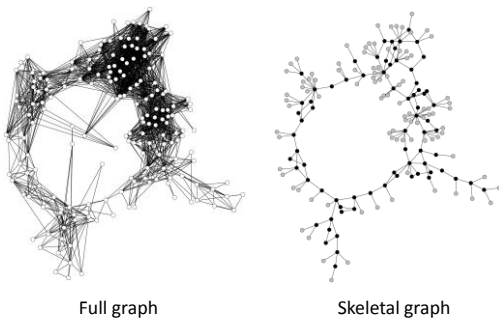
✓ We found 39,609 results for photos matching **colosseum** and **rome**.

- How do we scale from 100s to 10,000s of images?
- Observation: Internet collections represent very non-uniform samplings of viewpoint

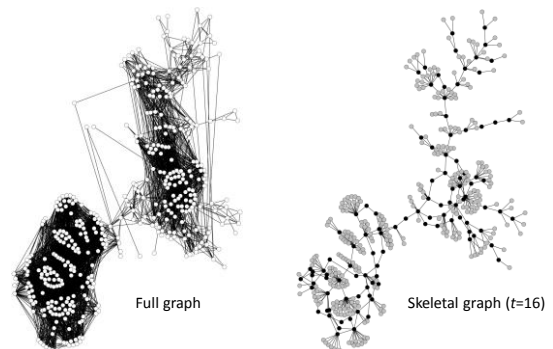
Skeletal set

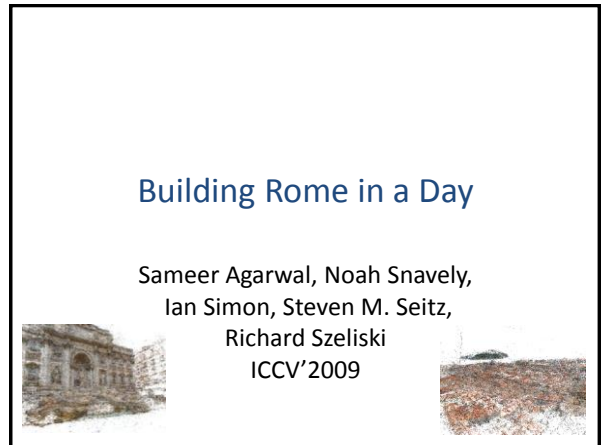
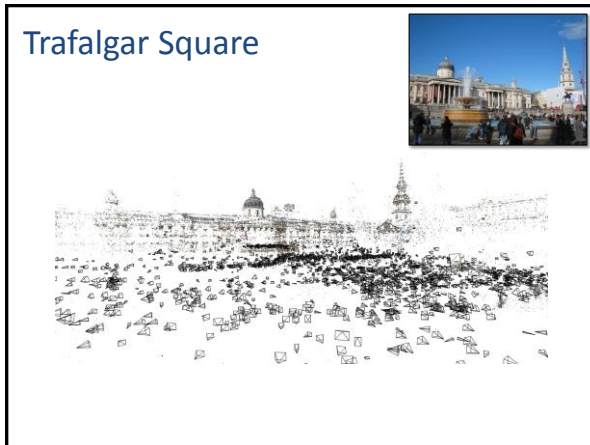
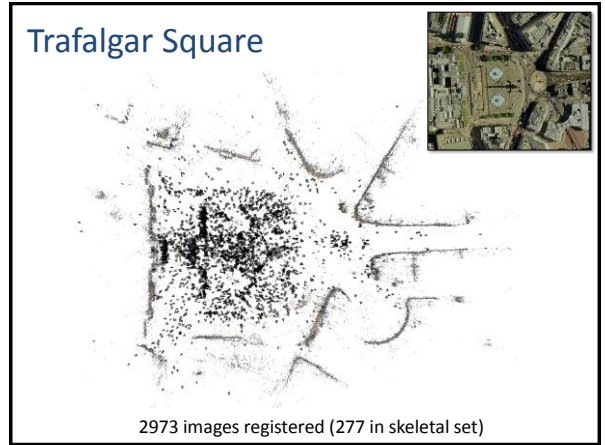
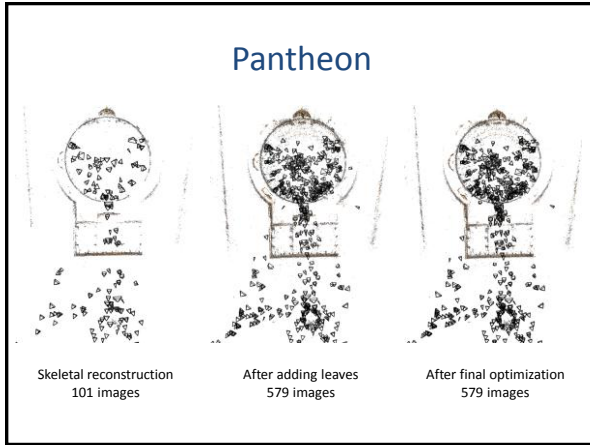
- Goal: given an image graph G_I , select a small set S of **important** images to reconstruct, bounding the loss in **quality** of the reconstruction
- Reconstruct the skeletal set S
- Estimate the remaining images with much faster pose estimation steps

Example



Pantheon





Results: Dubrovnik



(a) Dubrovnik: Four different views and associated images from the largest connected component. Note that the component captures the entire old city, with both street-level and roof-top detail. The reconstruction consists of 4,585 images and 2,662,981 3D points with 11,839,682 observed features.



Results: Rome



Colosseum: 2,097 images, 819,242 points

Trevi Fountain: 1,935 images, 1,055,153 points



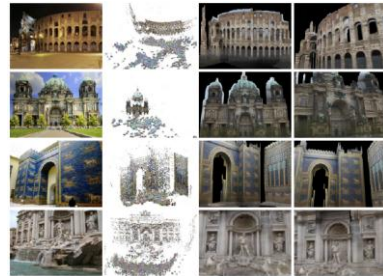
Pantheon: 1,032 images, 530,076 points

Hall of Maps: 275 images, 230,182 points

(b) Rome: Four of the largest connected components visualized at canonical viewpoints [14].

Reconstructing Rome on a Cloudless Day

[J.M. Frahm *et al.*, ECCV 2010]



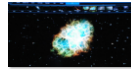
Applications

Community photo collections



- “Wikipedia for photos” – visual record of world through community of photographers

- *Geograph British Isles*
<http://www.geograph.org.uk/>



- Users can tag and comment on photos, link to other content

- *World-wide telescope*



- “Where should I take a photo?”

- <http://photocitygame.com/>

Community photo collections

- Leveraging large databases of photos, large number of users
 - Annotations / augmented reality

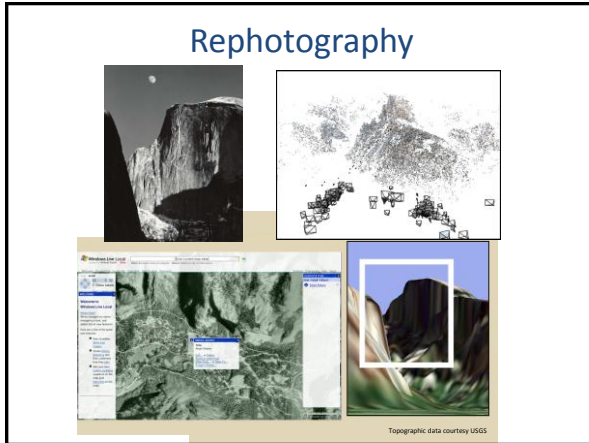


Virtual tour guide scenario






St. Peter's Basilica
 • Built: 1506-1626
 • http://en.wikipedia.org/wiki/St_Peter's_Basilica





Summary

- 
• Large (Internet) photo collections: the next frontier in computer vision
- 
• Efficient algorithms for large-scale 3D matching and reconstruction
- 
• New challenges for 3D recognition



