Modeling and Visualizing the World Through Internet Photo Collections

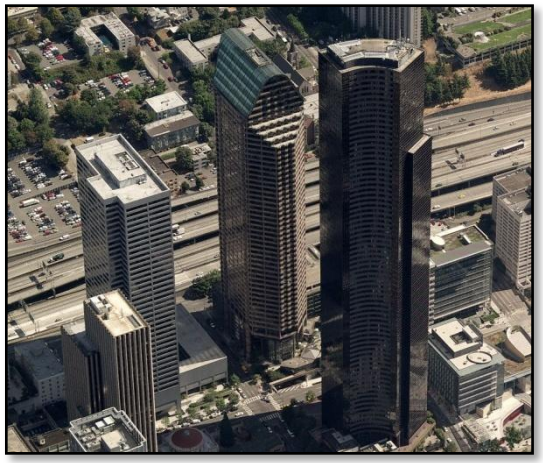
Noah Snavely

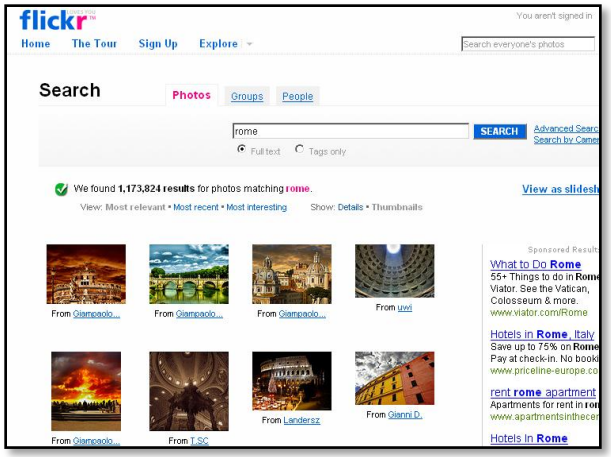
Cornell University

Goddard Space Flight Center, IS&T Colloquium April 29, 2009

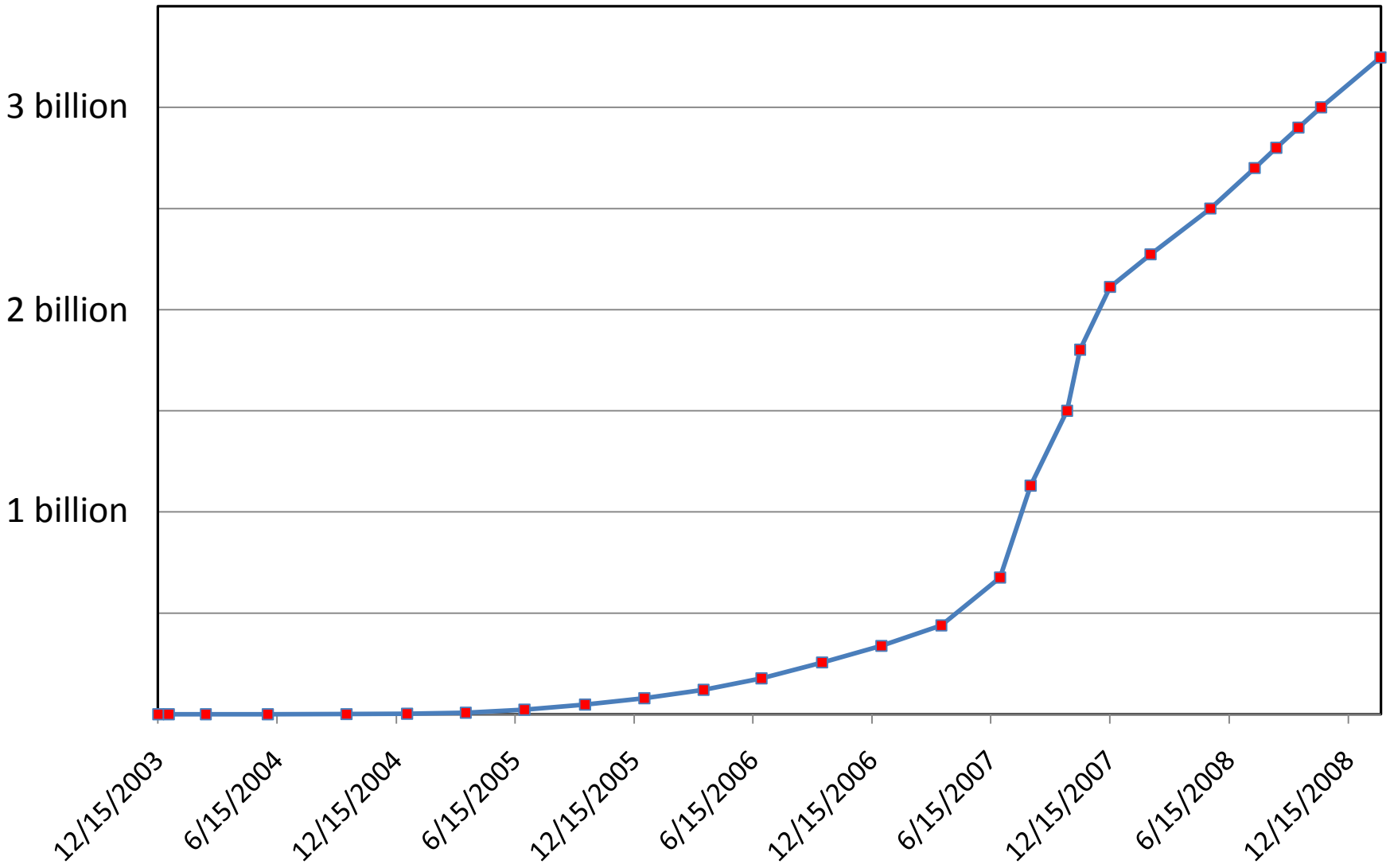
The world in photos

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

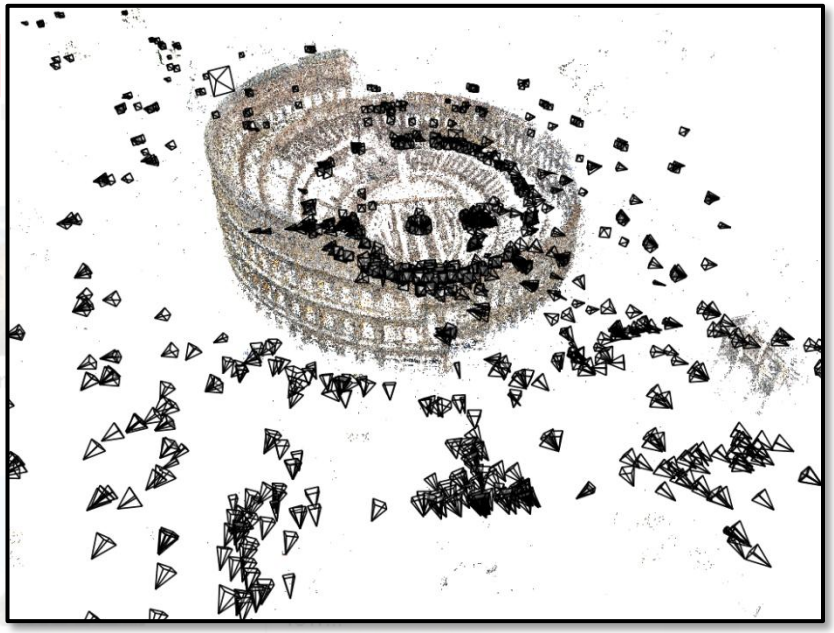




Flickr

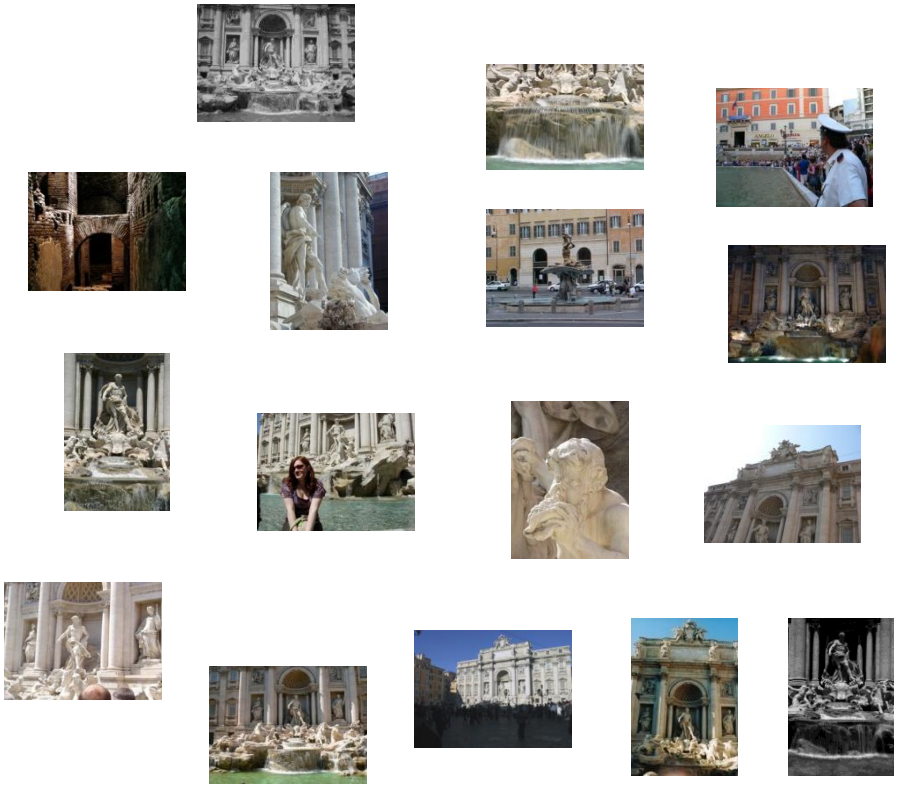


> 6.3 billion photos on Photobucket, > 10 billion on Facebook



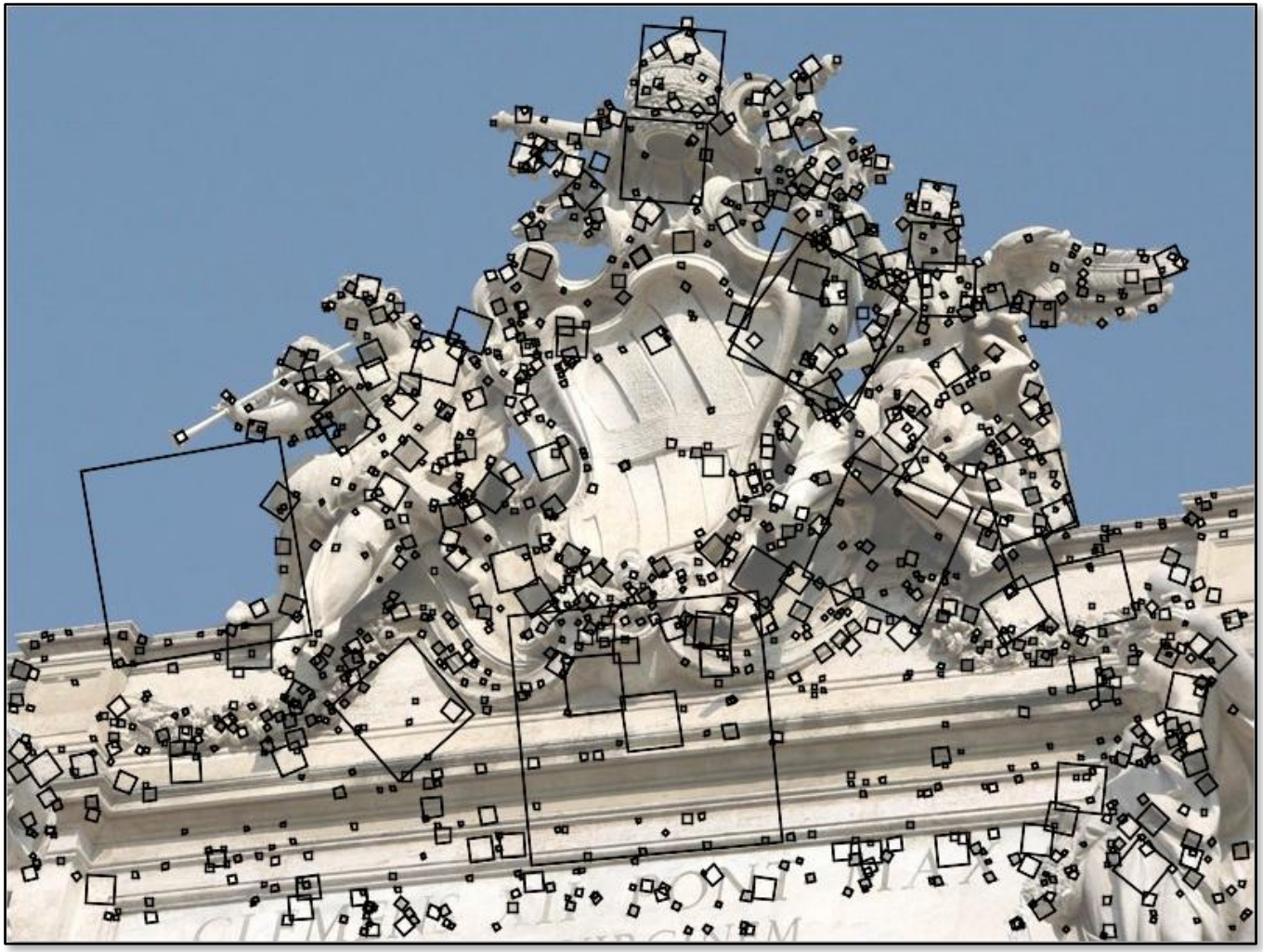
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



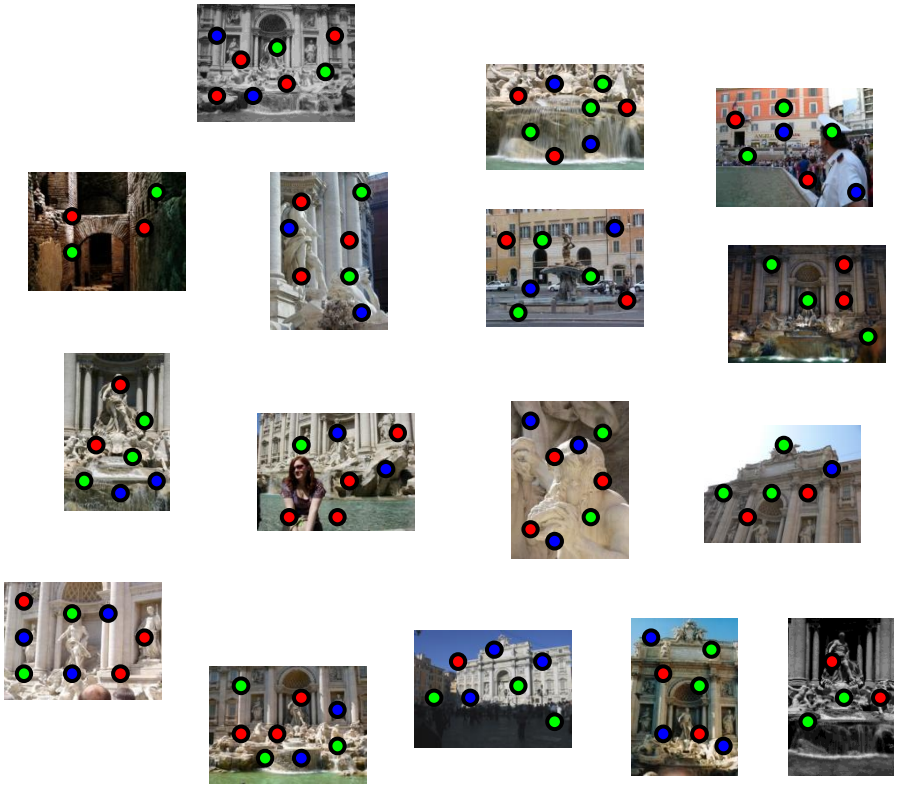
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]



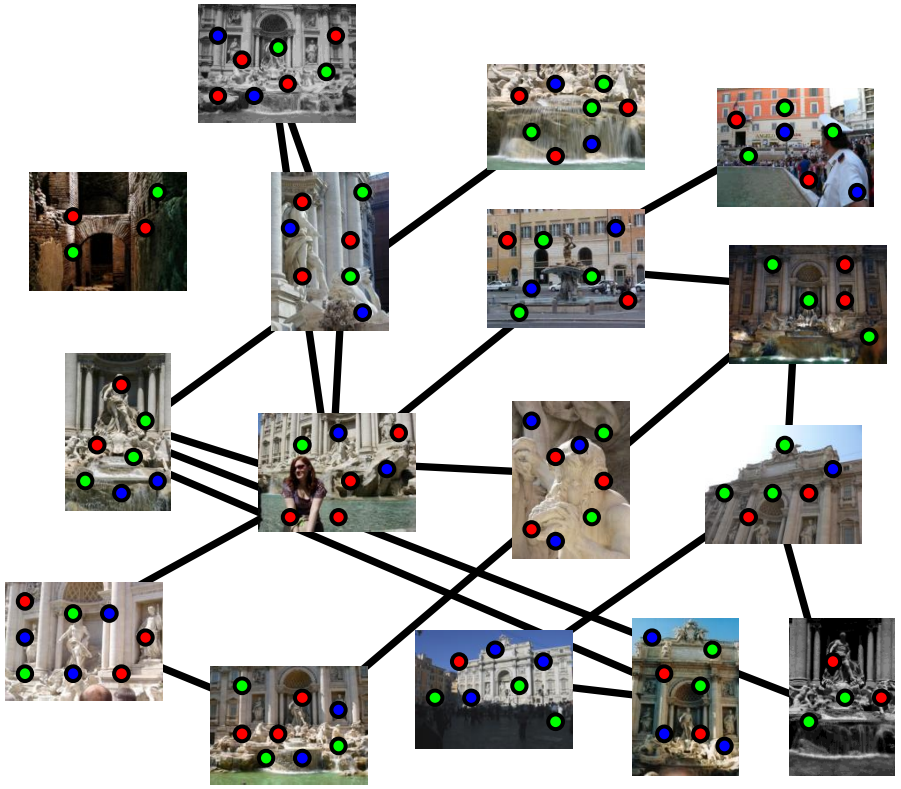
Feature detection

Detect features using SIFT [Lowe, IJCV 2004]

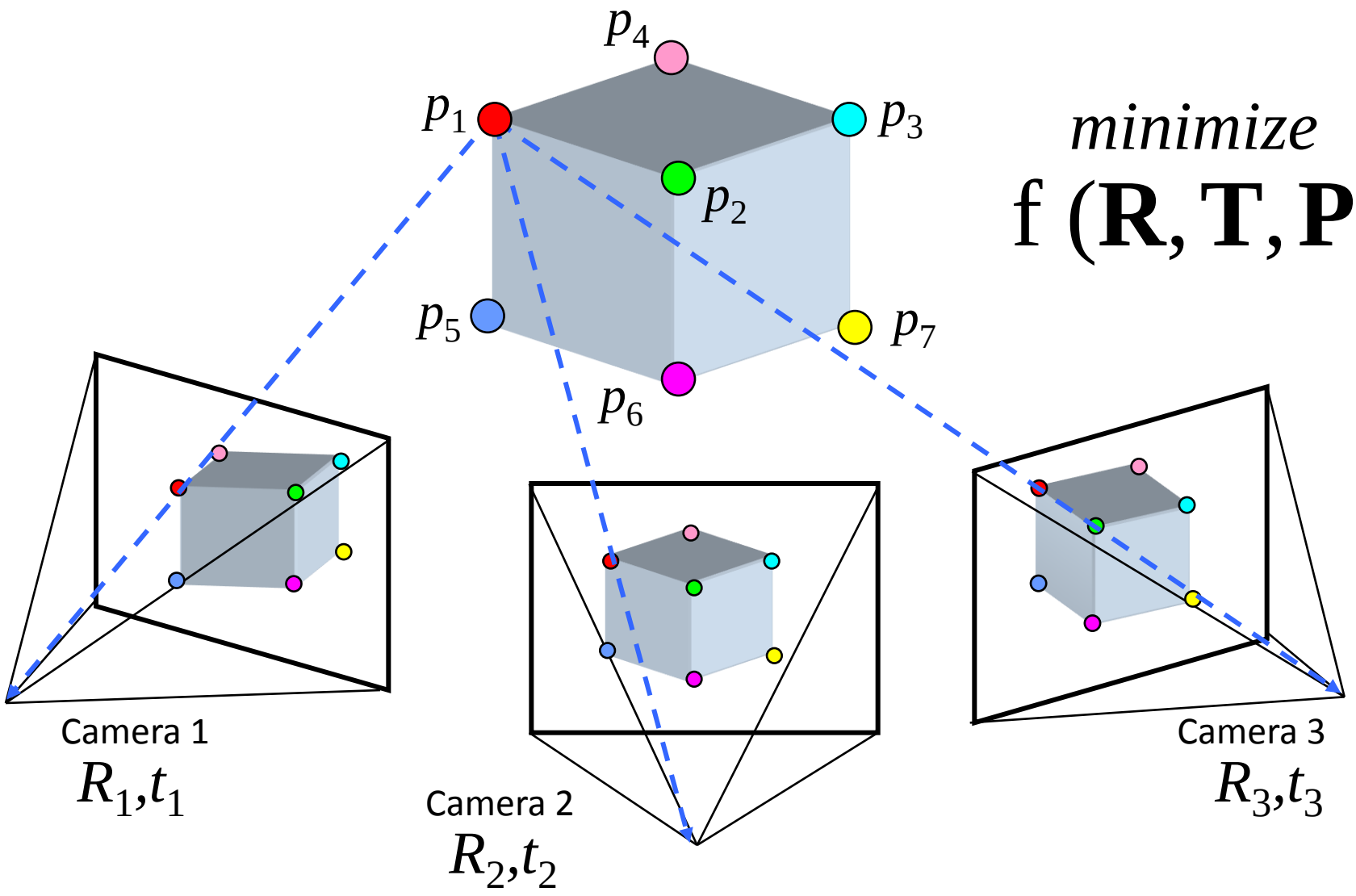


Feature matching

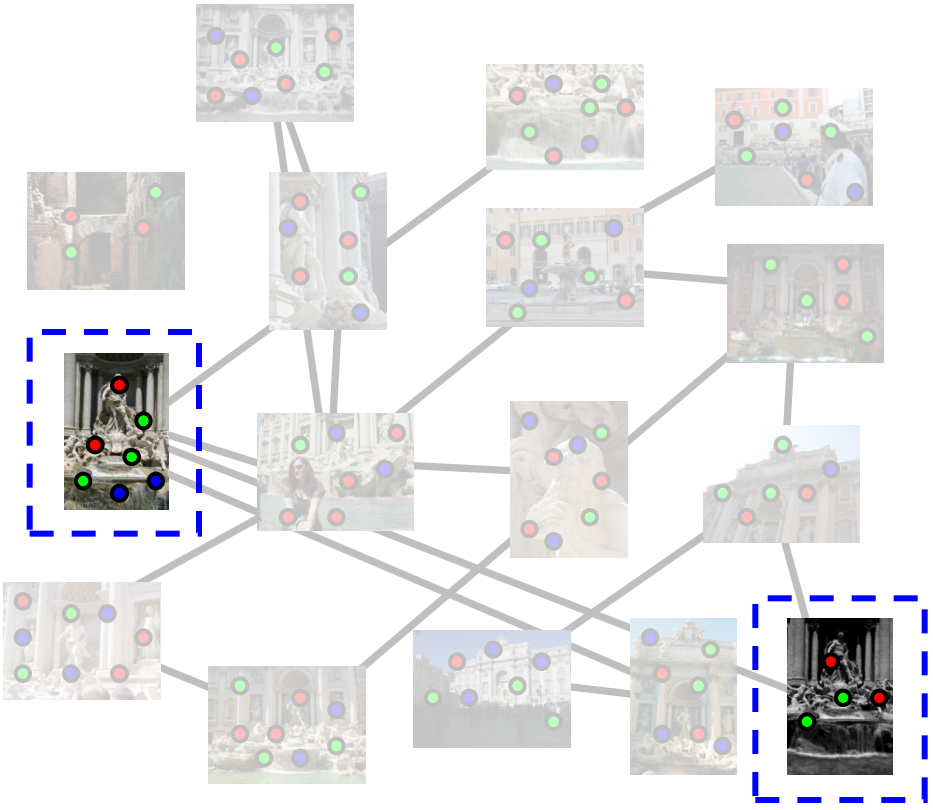
Match features between each pair of images



Structure from motion



Incremental structure from motion



- Automatically select an initial pair of images

Incremental structure from motion

Incremental structure from motion

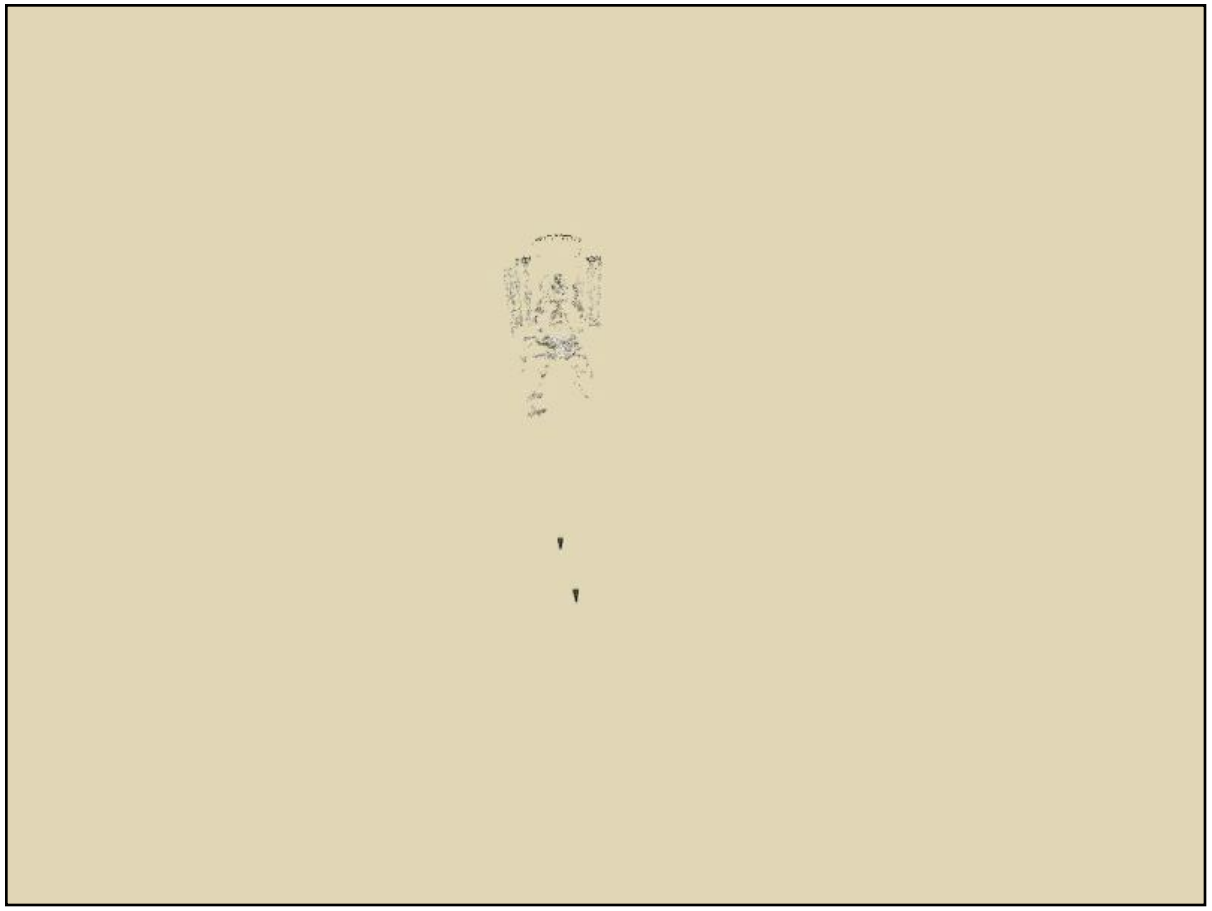
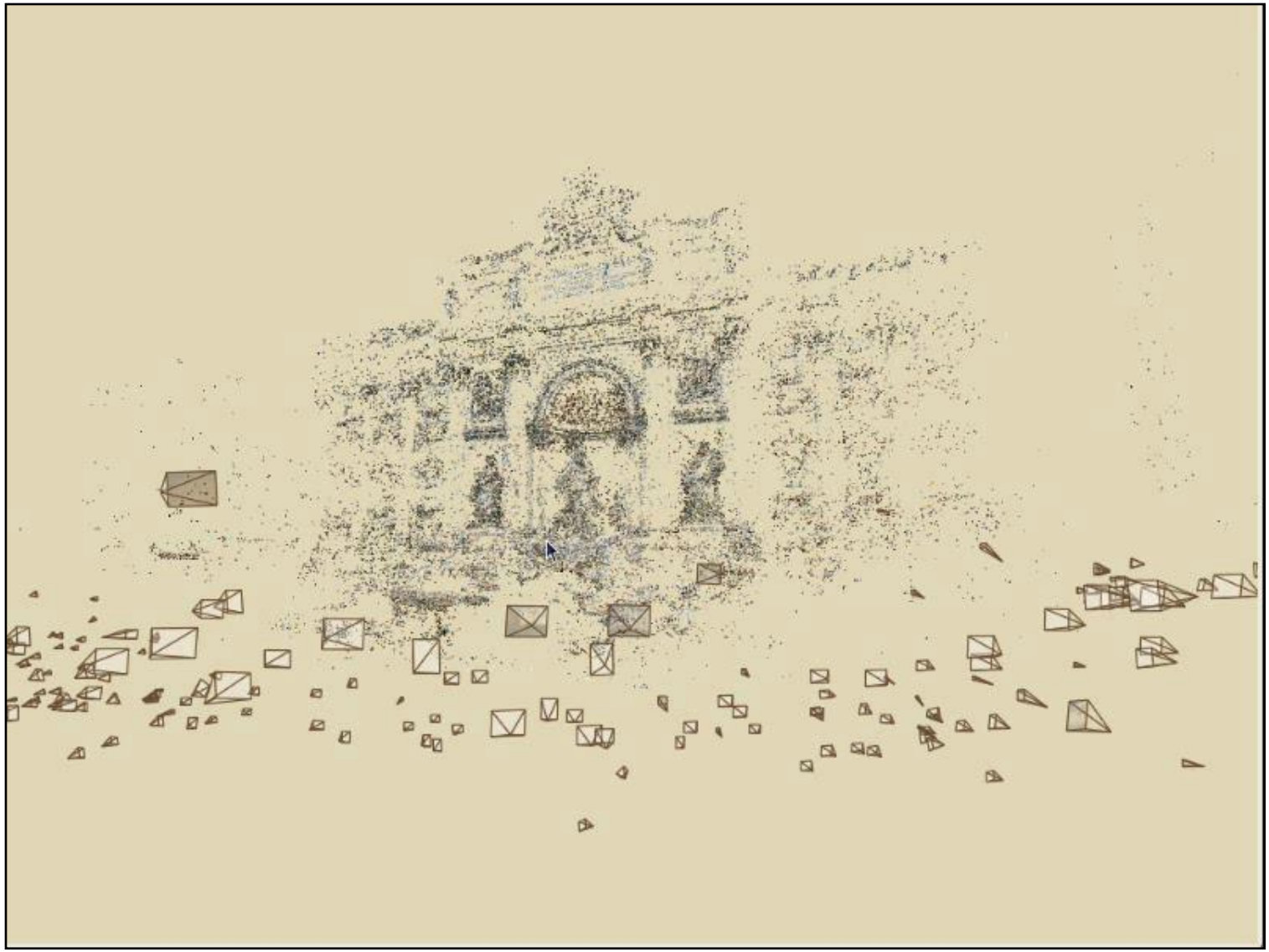
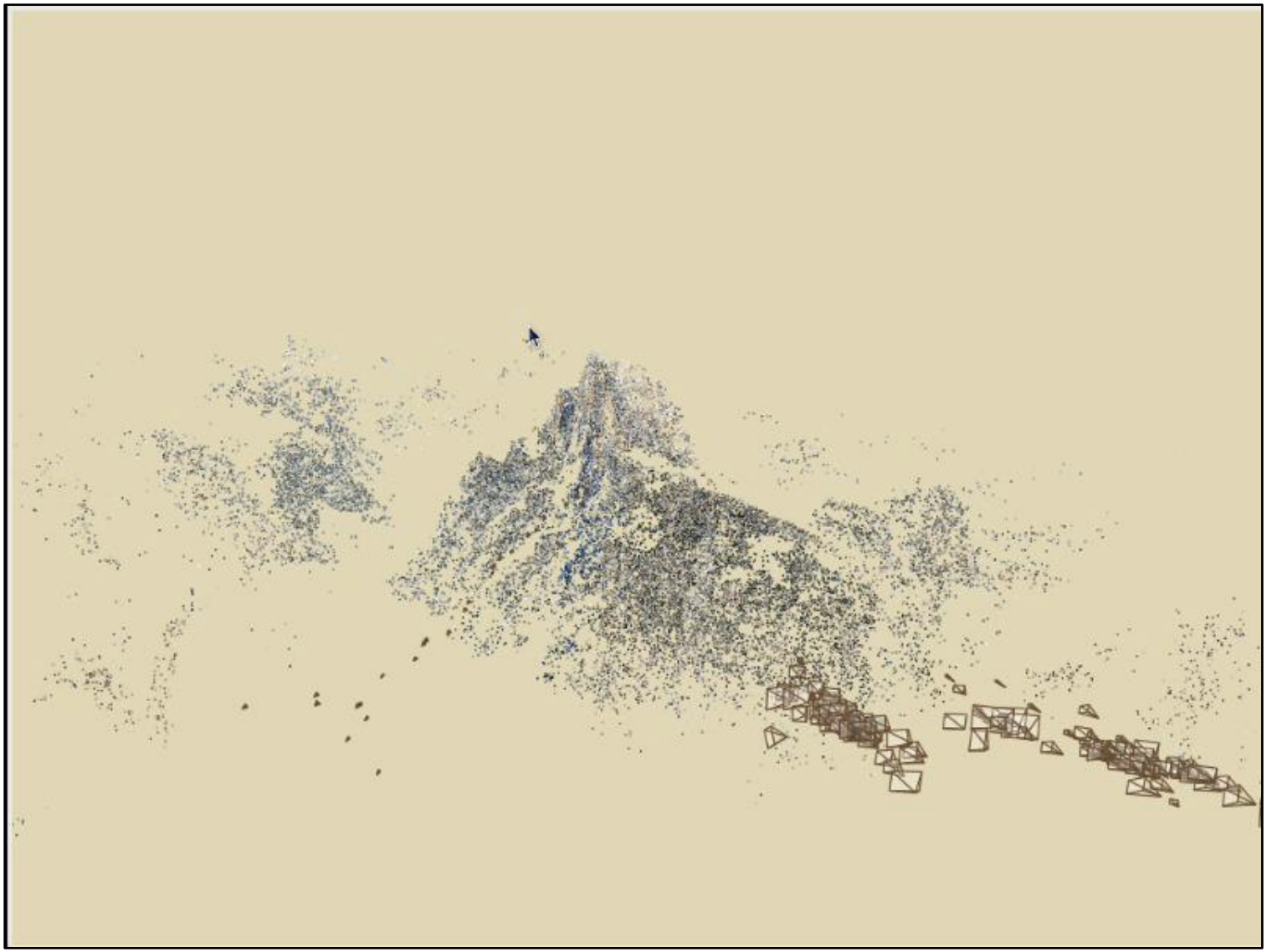


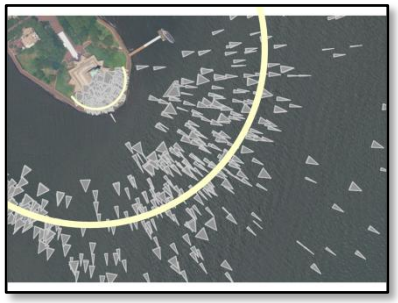
Photo Tourism



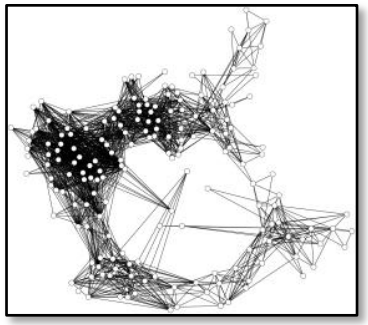
Yosemite



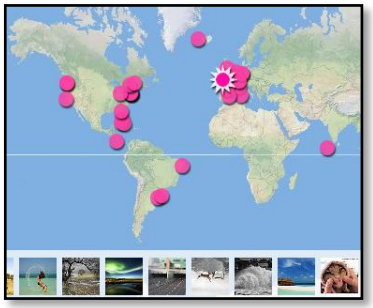
Overview



- Finding Paths through the World's Photos



- Large-scale 3D reconstruction



- Ongoing and future projects

Libration

From Wikipedia, the free encyclopedia

Not to be confused with Liberation or Libation.

In astronomy **libration** (from the Latin verb *librare* "to balance, to sway", cf. *libra* "scales") refers to the various orbital conditions which make it possible to see more than 50% of the moon's surface over time, even though the front of the Moon is tidally locked to always face towards Earth. By extension, libration can also be used to describe the same phenomenon for other orbital bodies that are nominally locked to present the same face. As the orbital processes are repetitive, libration is manifested as a slow rocking back and forth (or up and down) of the face of the orbital body as viewed from the parent body, much like the rocking of a pair of scales about the point of balance.

Simulated views of the Moon over one month, demonstrating librations in latitude and longitude.

In the specific case of the Moon's librations, this motion permits a terrestrial observer to see slightly differing halves of the Moon's surface at different times. This means that a total of 59% of the Moon's surface can be observed from Earth.

There are three types of libration:

- *Libration in longitude* is a consequence of the Moon's orbit around Earth being somewhat eccentric, so that the Moon's rotation sometimes leads and sometimes lags its orbital position.
- *Libration in latitude* is a consequence of the Moon's axis of rotation being slightly inclined to the normal to the plane of its orbit around Earth. Its origin is analogous to the way in which the seasons arise from Earth's revolution about the Sun.
- *Diurnal libration* is a small daily oscillation due to the Earth's rotation, which carries an observer first to one side and then to the other side of the straight line joining Earth's center to the Moon's center, allowing the observer to look first around one side of the Moon and then around the other. This is because the observer is on the surface of the Earth, not at its centre.

WIKIPEDIA
The Free Encyclopedia

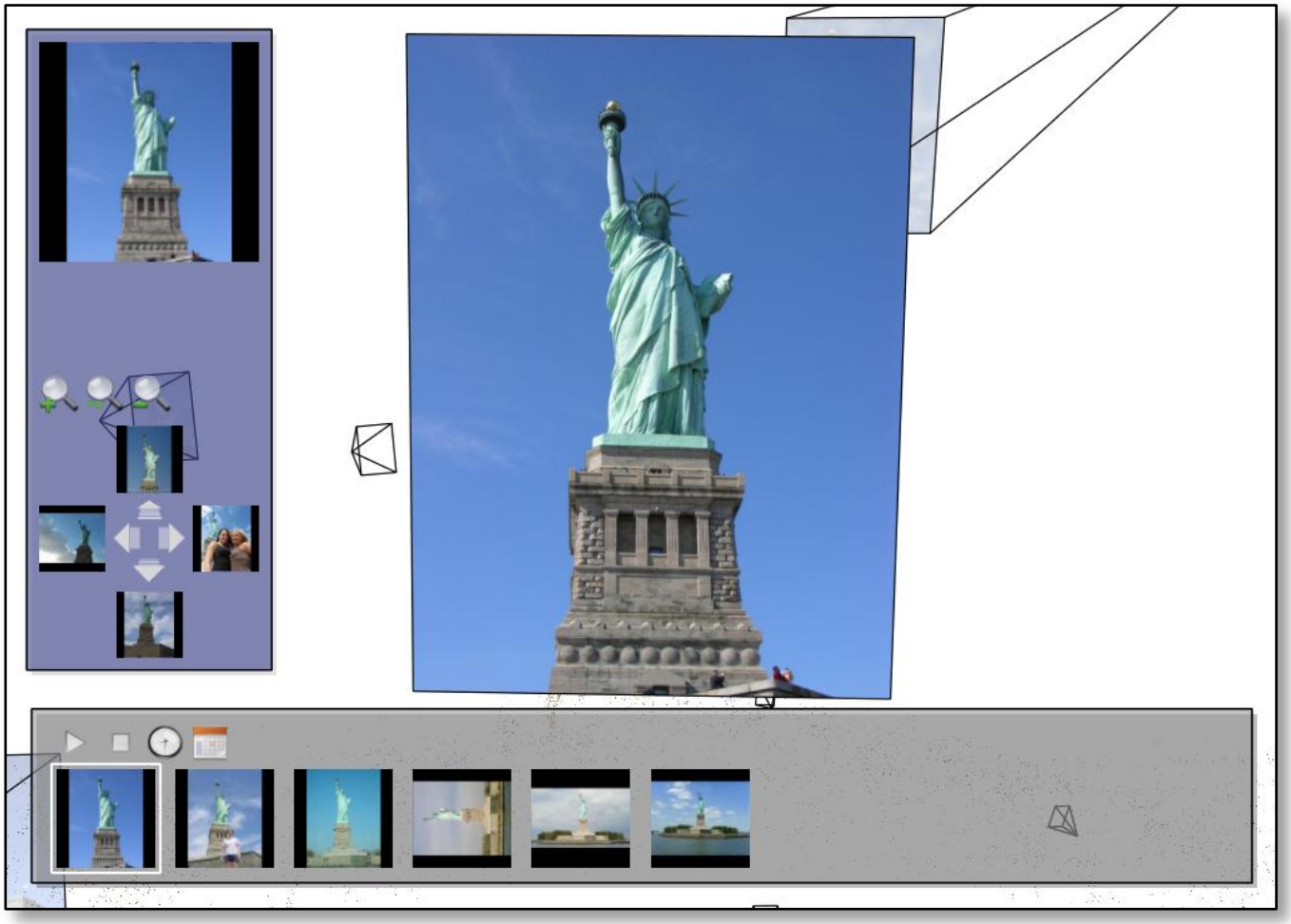
navigation
- Main page
- Contents
- Featured content
- Current events
- Random article

search
Go Search

interaction
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

3D navigation – Photo Tourism

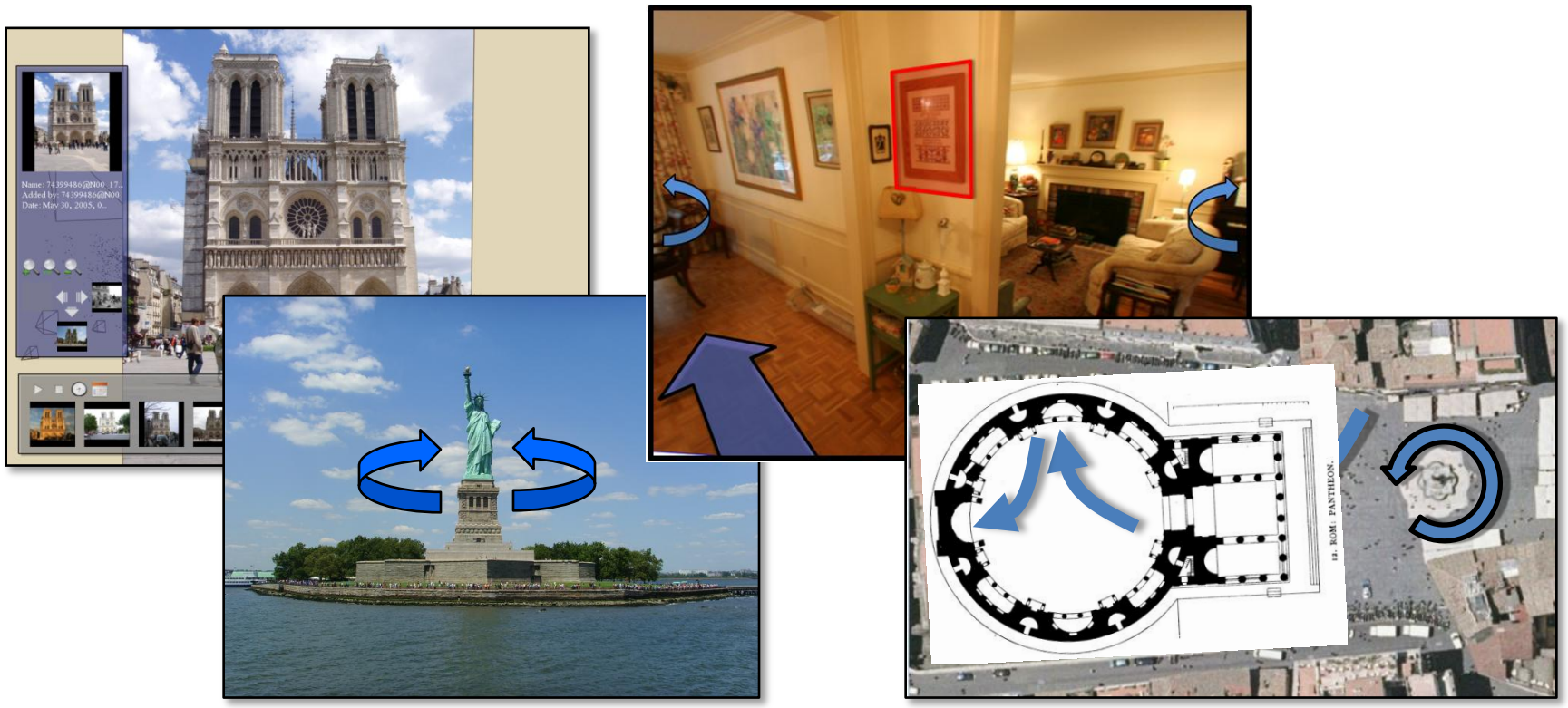


Demo

Continuous navigation

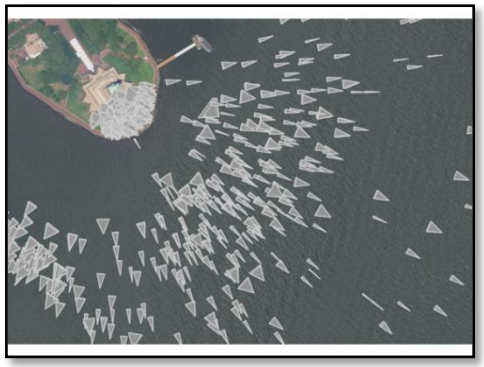
Demo

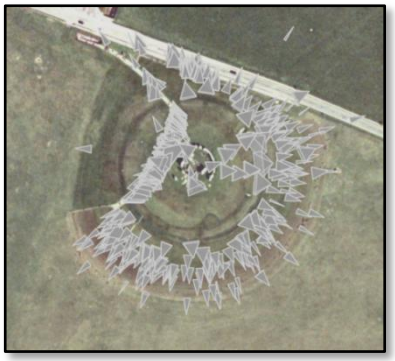
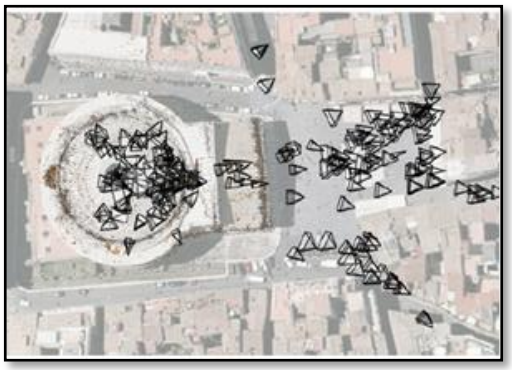
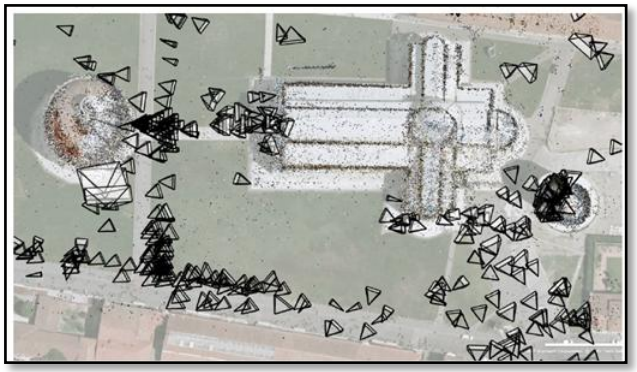
- What are good controls for exploring a given 3D scene?



Navigation controls

- Our approach: *Derive* good controls from the distribution of viewpoints in a large photo collection




Orbit point
Proxy plane
Virtual camera

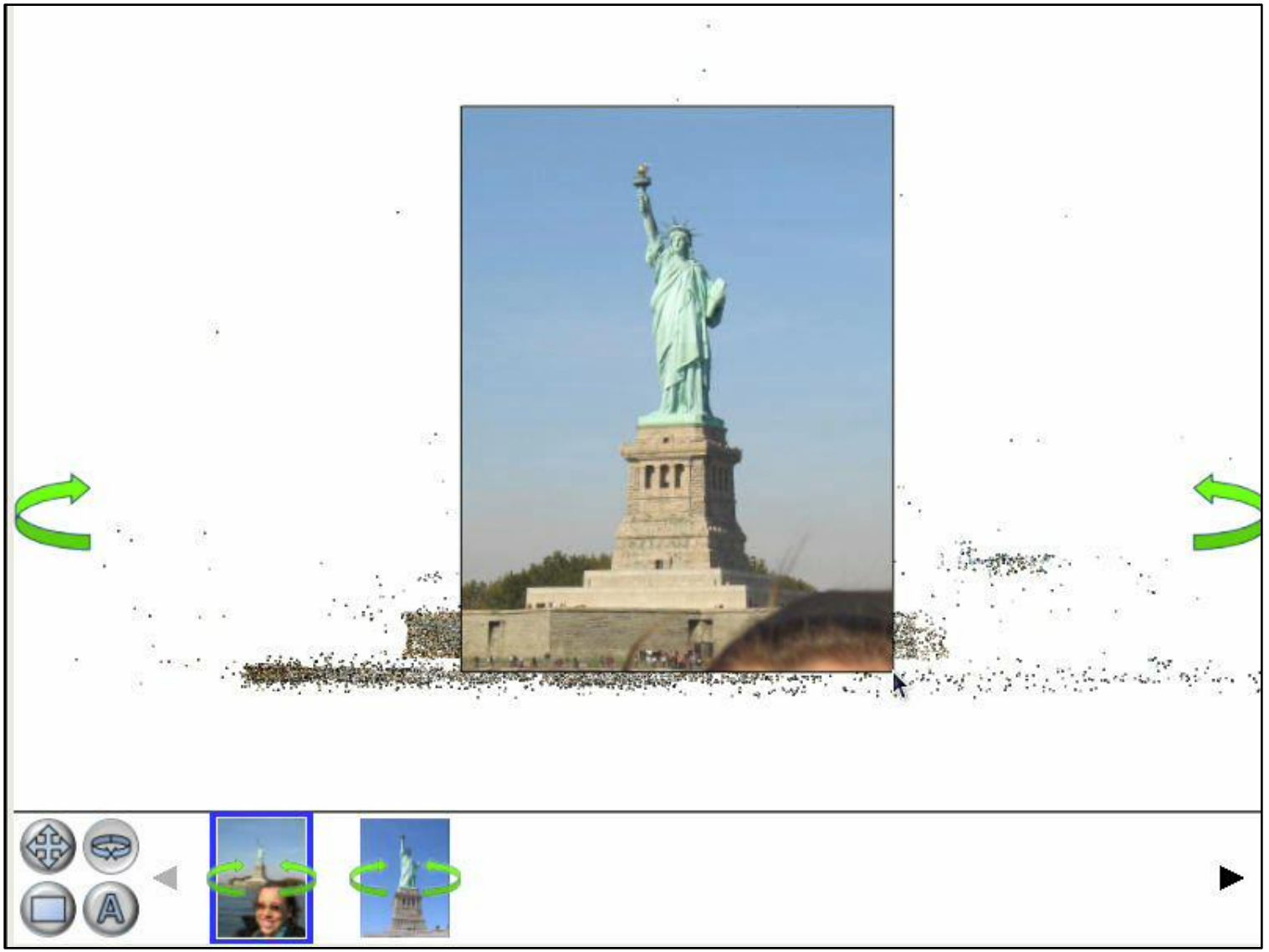
3D navigation controls

Problem: 3D scenes are difficult to navigate

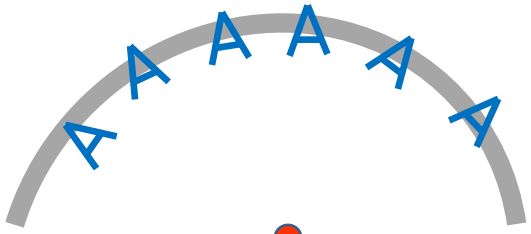
- How does the user know where to go?
- How does the user get there?
- Good controls are scene dependent

Solution: exploit the distribution of photos to derive good controls

Finding paths through photo collections



Scene-specific navigation controls



- Orbits

- Panoramas

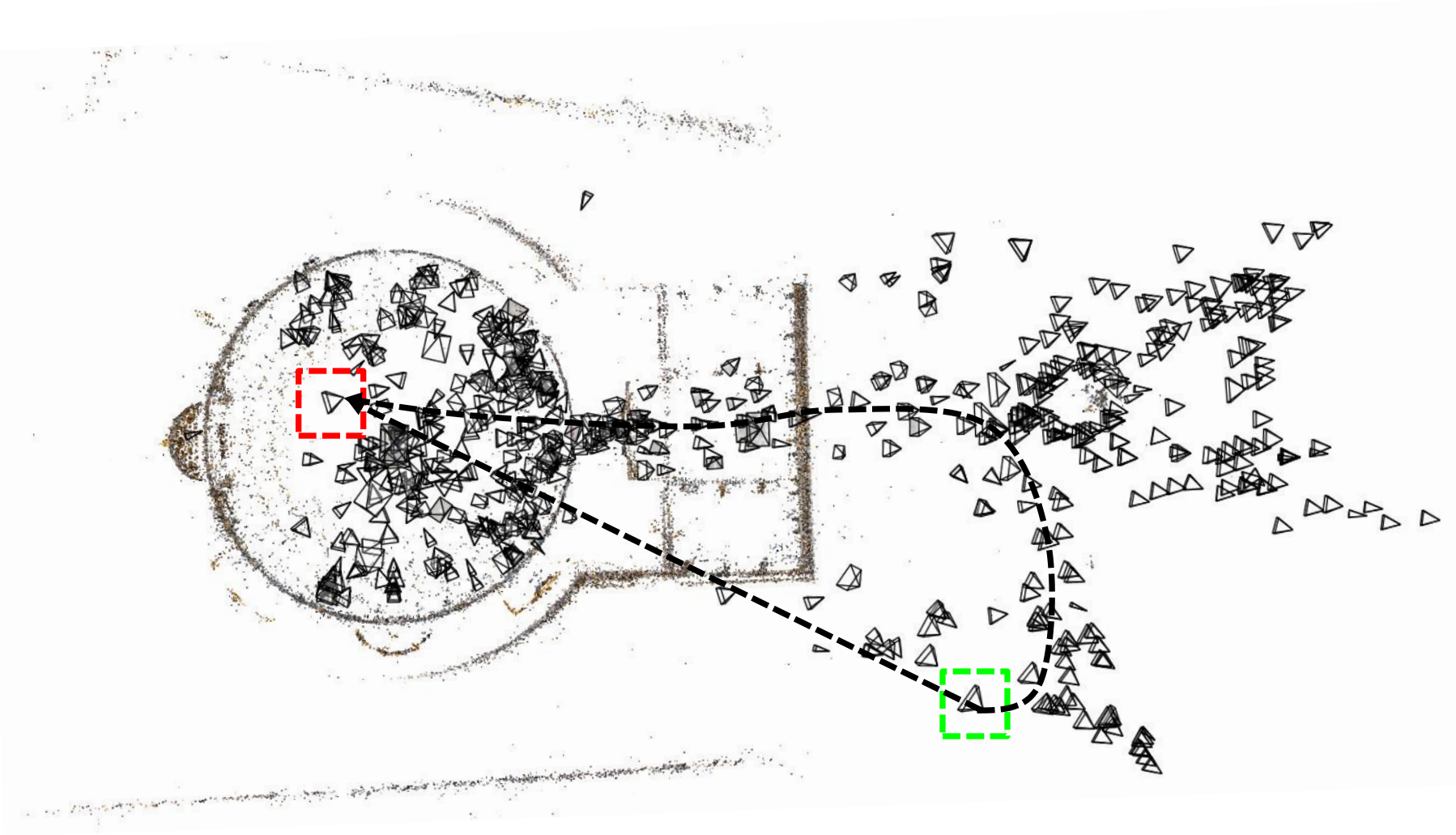
- Representative viewpoints

- Optimized paths between views

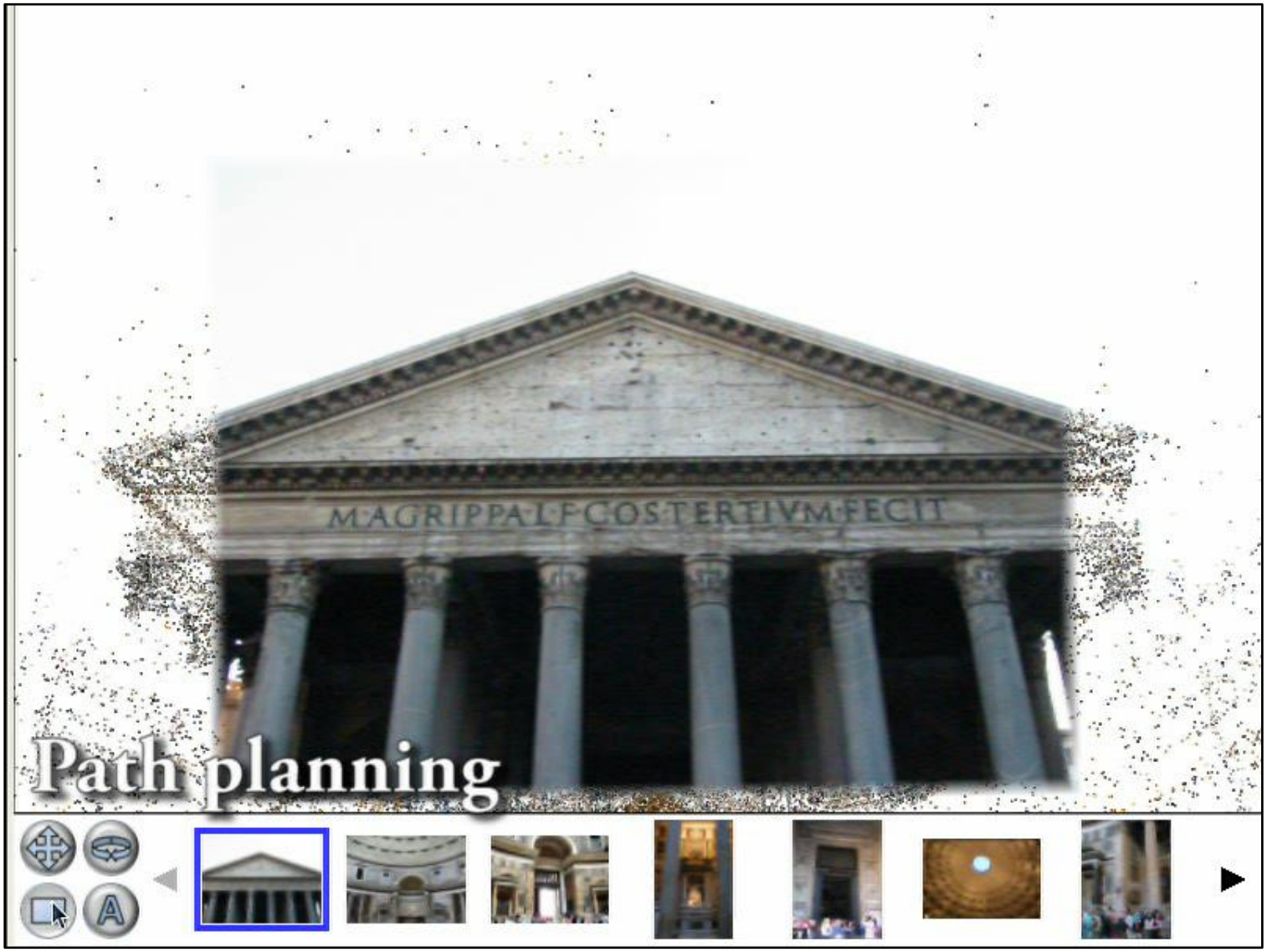
Pantheon

[Simon, et al., **Scene summarization for Online Image Collections**, ICCV ’07]

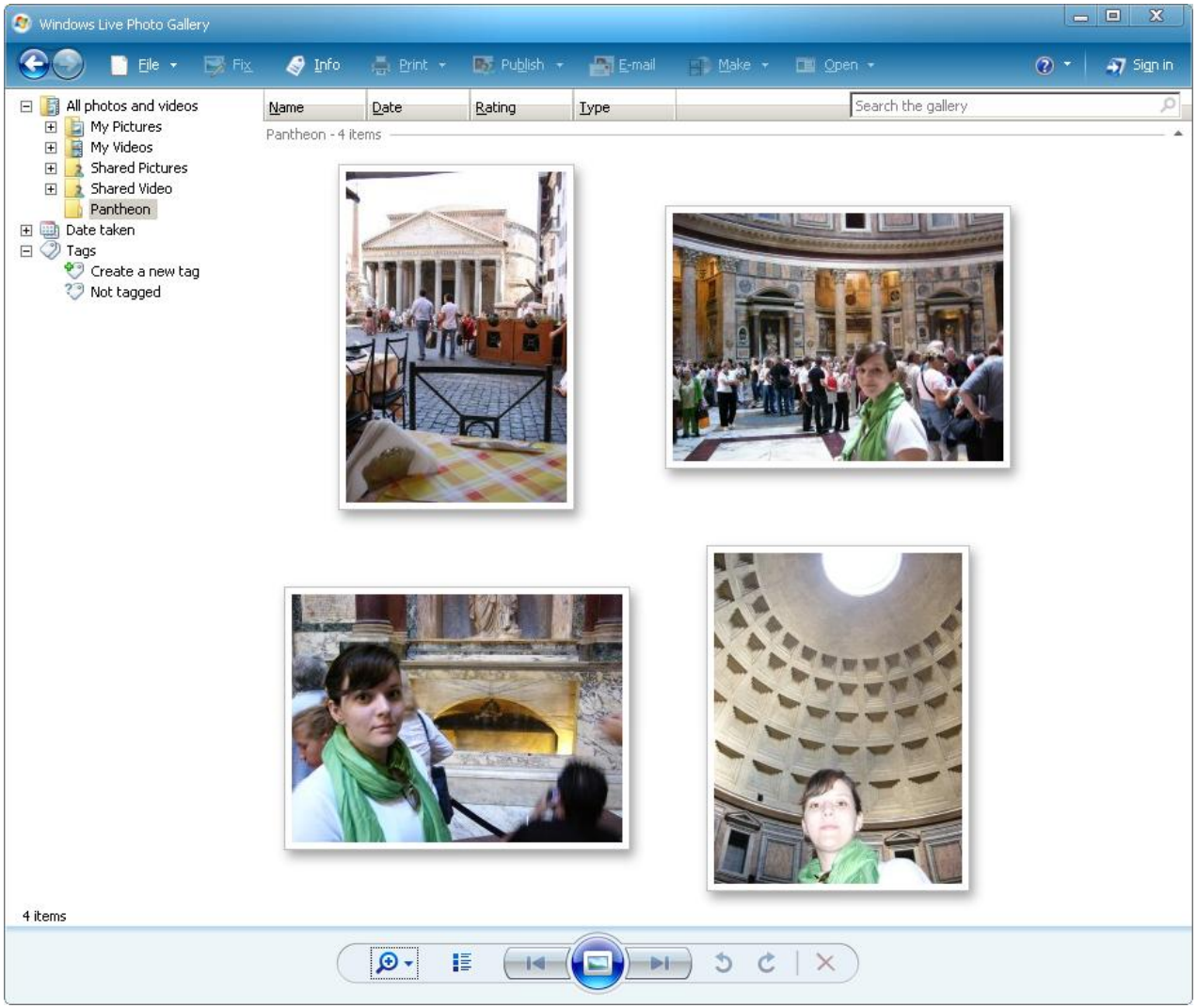
Optimized paths between views




M·AGRIPPA·L·F·COS·TERTIVM·FECIT


M·AGRIPPA·L·F·COS·TERTIVM·FECIT
Path planning

Personal photo tour




M AGRIPPA L F COS TERTIVM FECIT

Trevi Fountain

Microsoft® Live Labs™

Photosynth™

Sign In | Search Synths

National Geographic - Sphinx

232 photos, 100% synthy

Go

Christchurch Cathedral by GoldingArts
525 Photos - 94% Synthy

Create your Synth

Synth and upload your own photos

Explore Synths

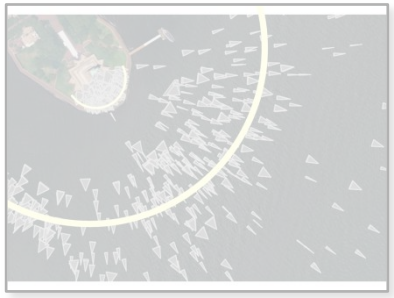
See what other people are synthing

About Photosynth

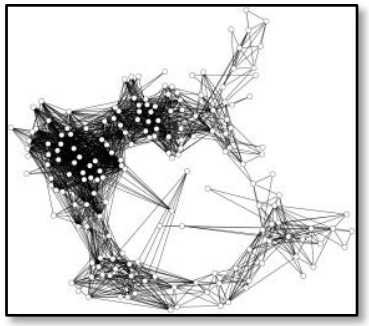
Learn about Photosynth and our team

Read the latest Photosynth news and updates on the blog

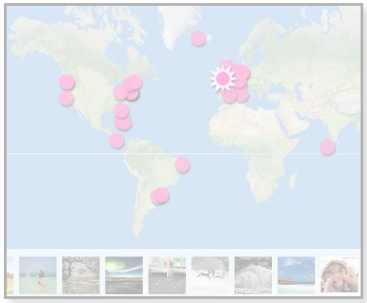
Overview



- Finding Paths through the World's Photos



- Large-scale 3D reconstruction



- Ongoing and future projects

Large-scale reconstruction

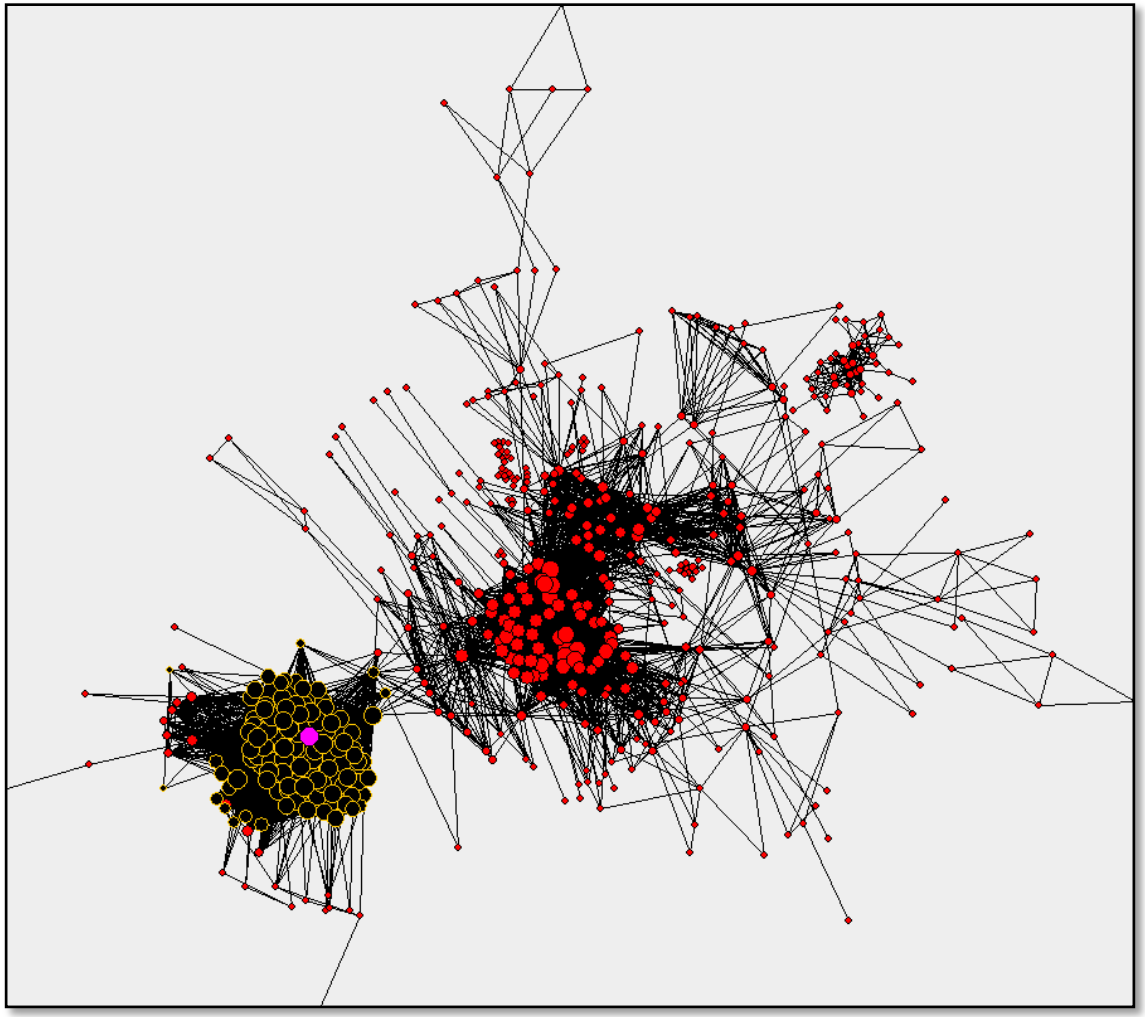
- Most of the models shown so far have had ~500 images

- How do we scale from 100s to 10,000s of images?

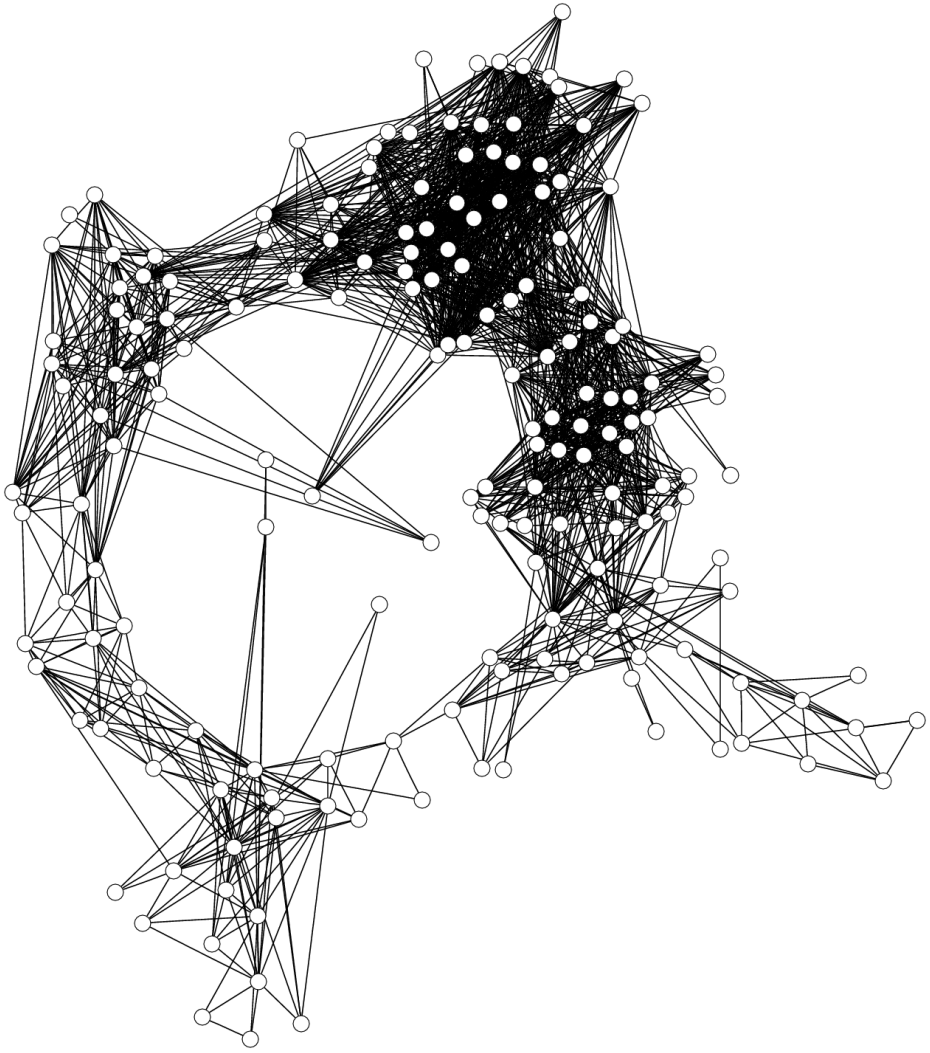
- Observation: Internet collections represent very non-uniform samplings of viewpoint

  [Snavely, Seitz, Szeliski, *CVPR* 2008]

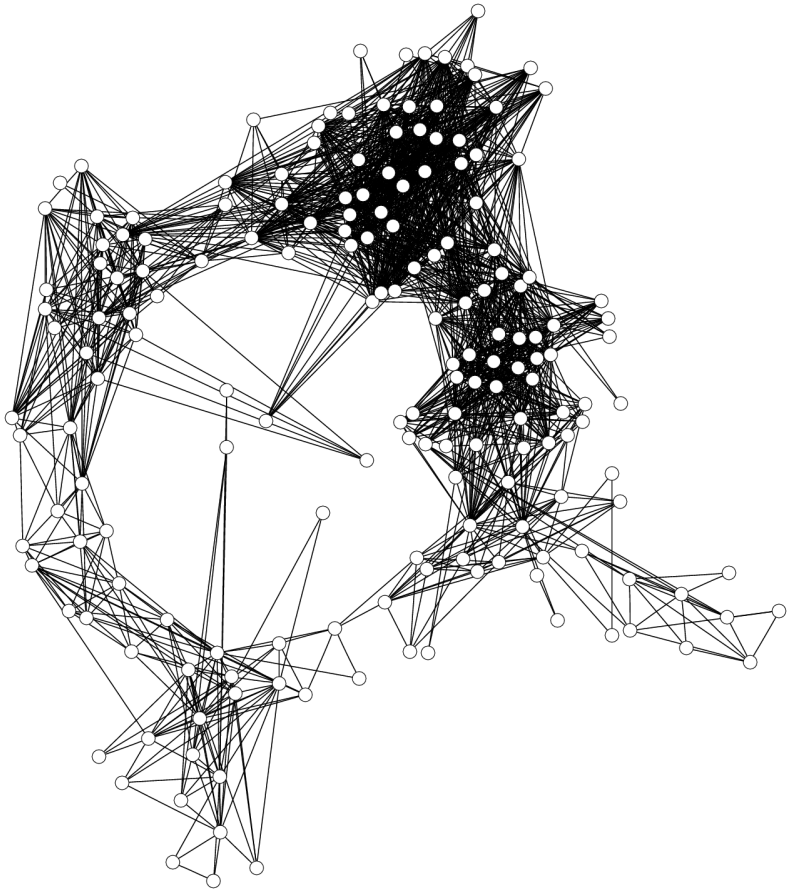
The Pantheon



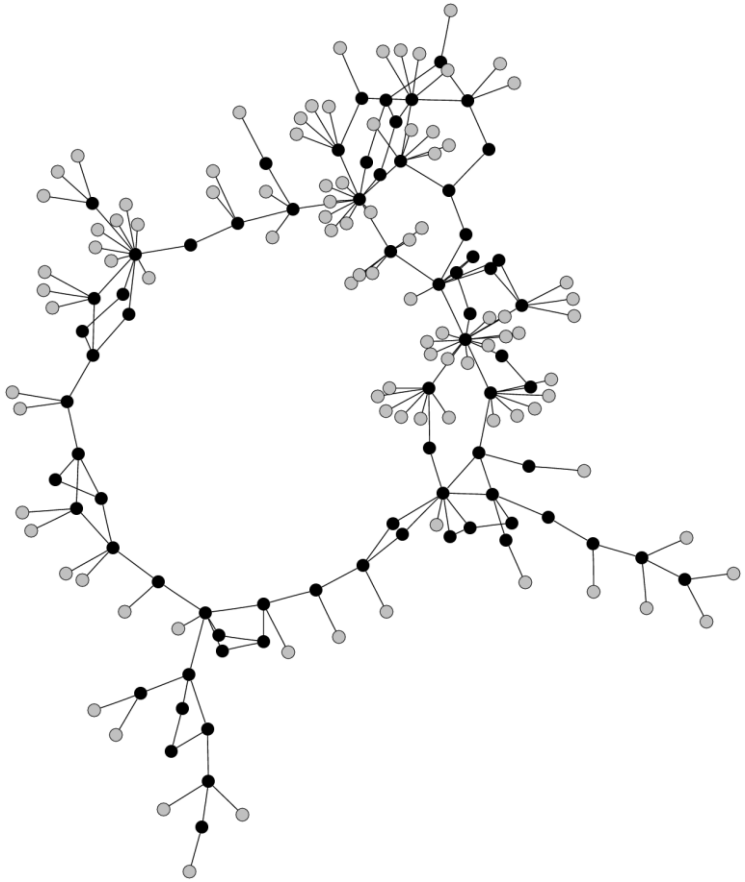
Stonehenge



Stonehenge



Full graph



Skeletal graph

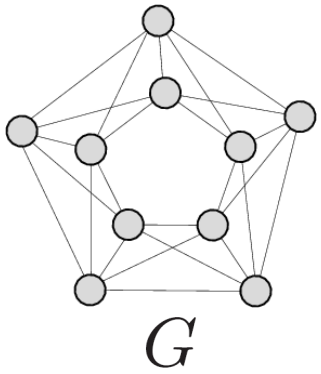
Skeletal set

- Goal: given an image graph $G_{\mathcal{I}}$,

  select a small set *S* of ***important*** images to reconstruct, bounding the loss in ***quality*** of the reconstruction

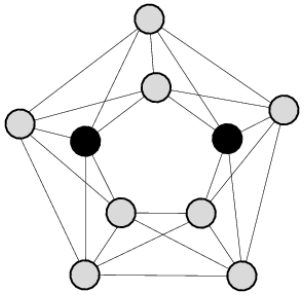
- Reconstruct the skeletal set *S*
- Estimate the remaining images with much faster pose estimation steps

Properties of the skeletal set



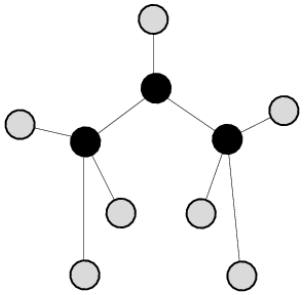
- Should touch all parts of G

  Dominating set



- Should form a single reconstruction

  Connected dominating set



- Should result in an *accurate* reconstruction

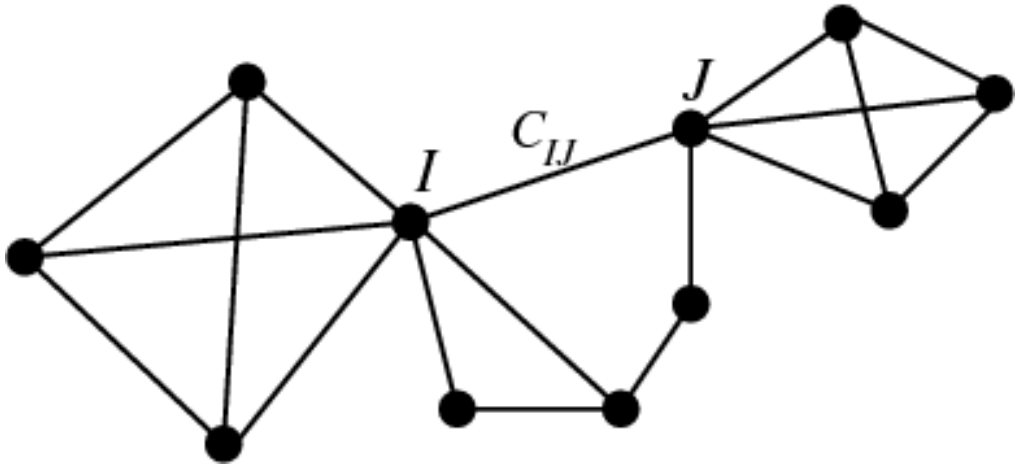
?

Representing information in a graph

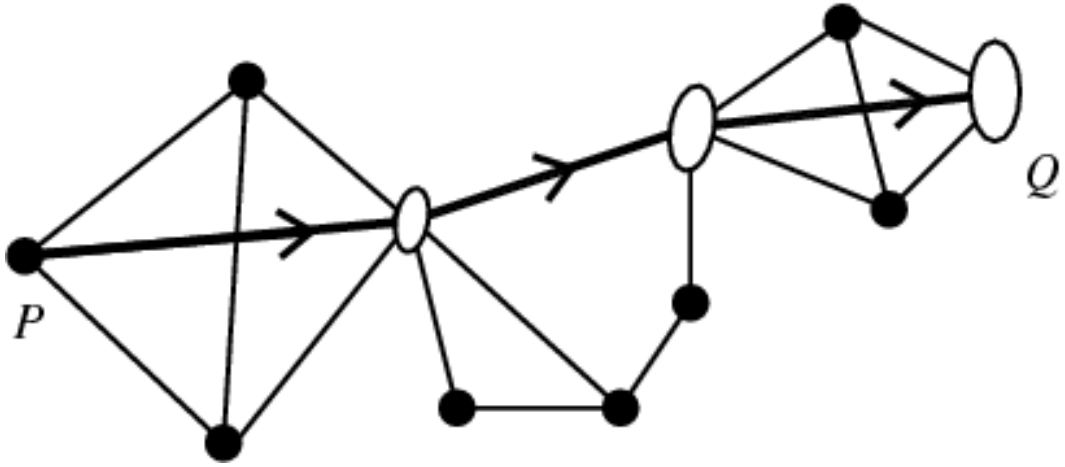
- What kind of information?
  - No absolute information about camera positions
  - Each edge provides information about the relative positions of two images

  - … but not all edges are equally informative
- We model information with the uncertainty (covariance) in pairwise camera positions

Representing information in a graph

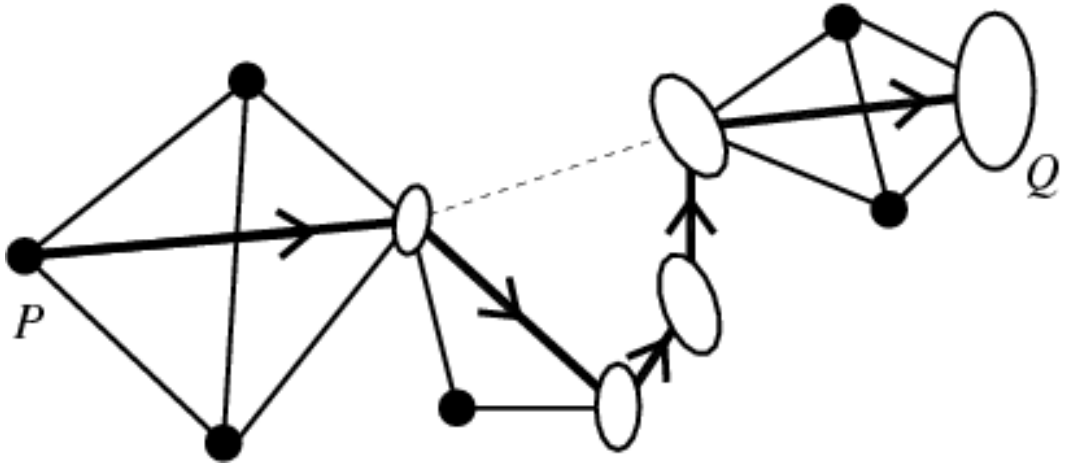


Representing information in a graph



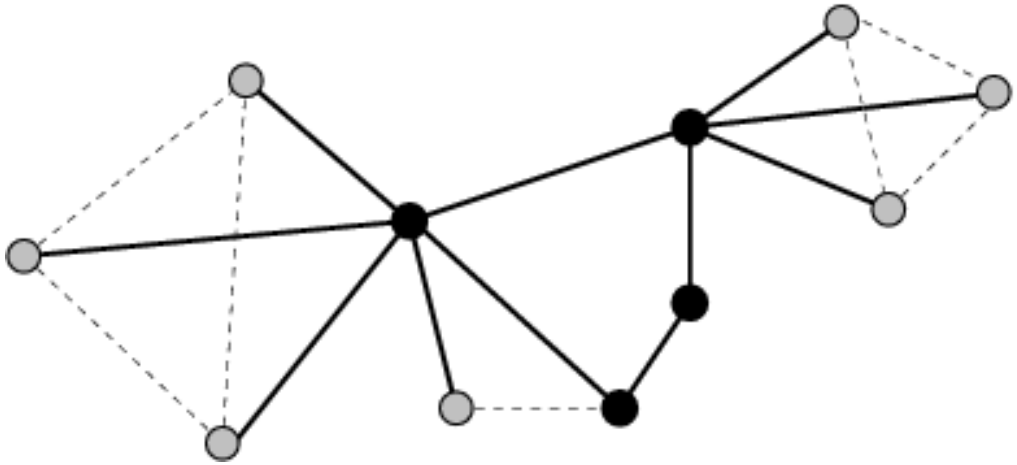
- Uncertainty grows with the length of the path
- Shortest path gives a bound on expected uncertainty
- We use the trace of the covariance matrix as our scalar edge weights

Representing information in a graph



- Uncertainty grows with the length of the path
- Shortest path gives a bound on expected uncertainty
- We use the trace of the covariance matrix as our scalar edge weights

Representing information in a graph

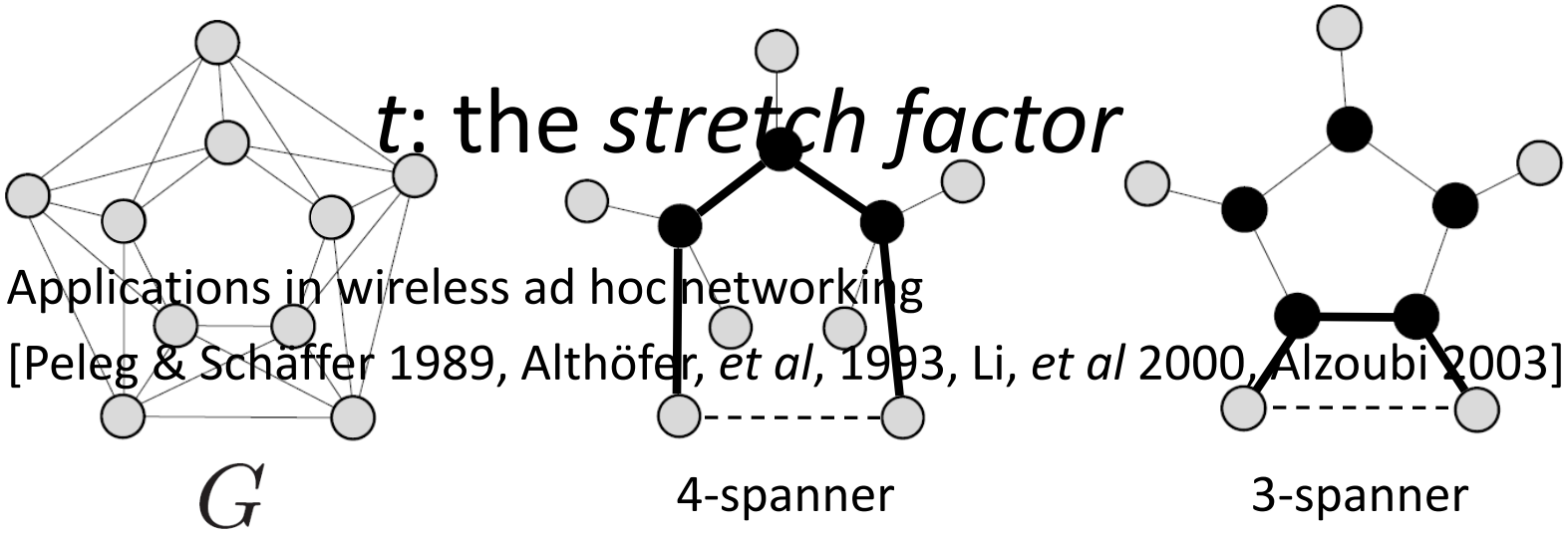


Want to find a subgraph with:

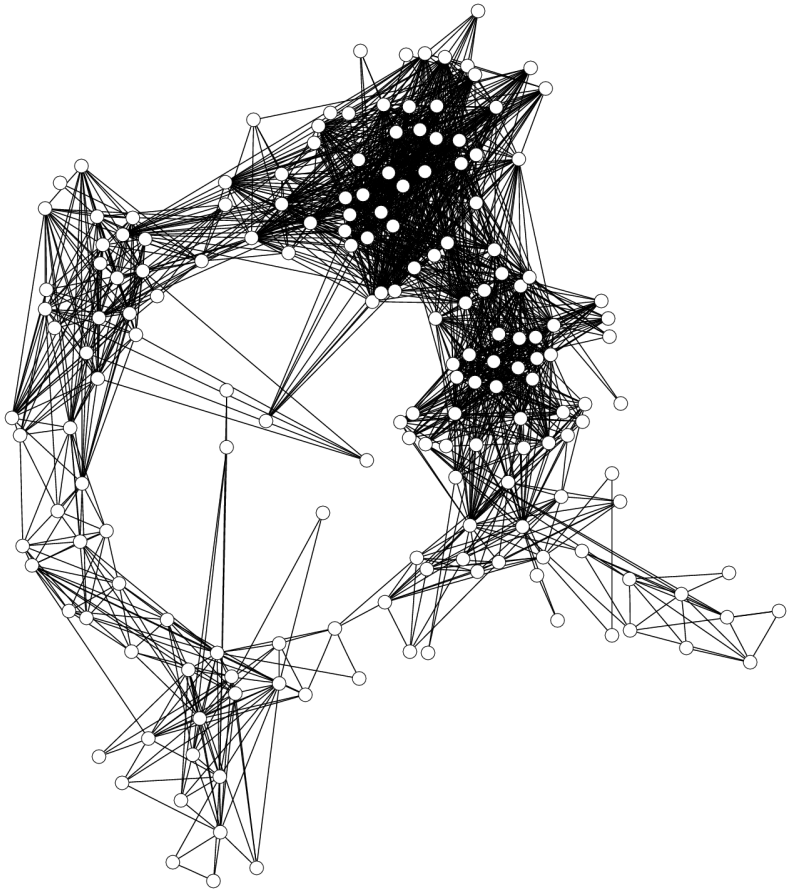
a) many leaves
b) small growth in estimated uncertainty between any pair of nodes

t-spanner problem

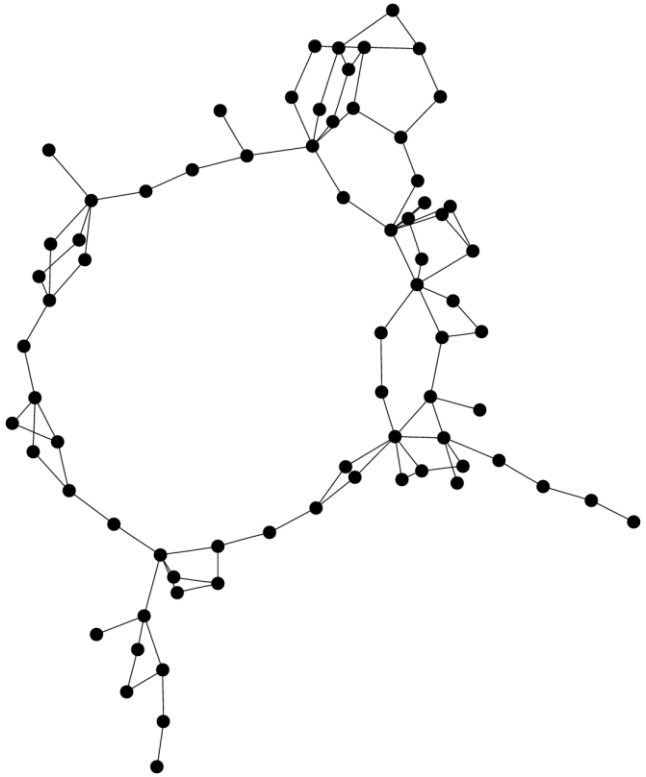
- Given a graph G, find a spanning subgraph G' such that, for every pair of vertices (*P*,*Q*), the distance between *P* and *Q* in G' is at most *t* times the distance between *P* and *Q* in G



Stonehenge



Full graph



Skeletal graph (t=16)
(leaves omitted)

Maximum-leaf *t*-spanner problem

- Given a graph $G_{\mathcal{I}}$, and a stretch-factor *t*, find a subgraph $G_{\mathcal{S}}$ which:
  a) is a *t*-spanner
  b) has the largest number of leaves

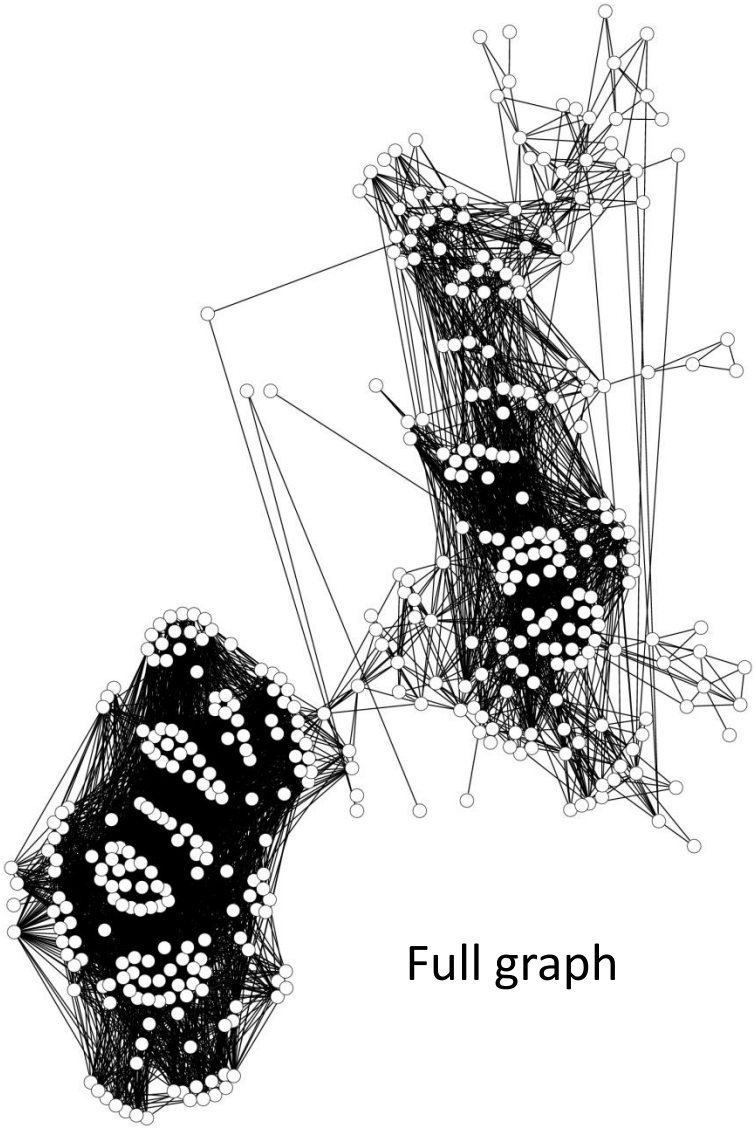
- We construct $G_{\mathcal{S}}$ by adding edges one at a time to an empty graph, until a) is satisfied

Properties of approach

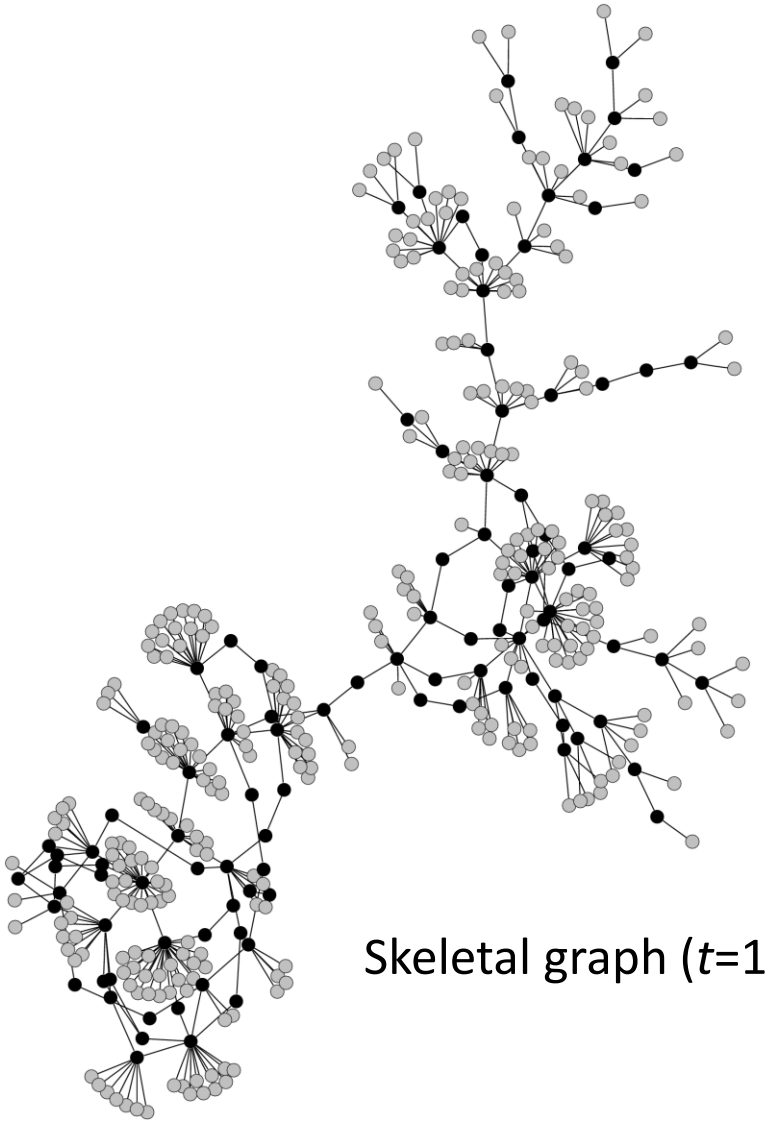
- Results in a connected reconstruction (when possible)
- Bounds expected increase in uncertainty of reconstructed model (bound is defined by t)
- Remaining information can be used to refine the model after the initial reconstruction

Results

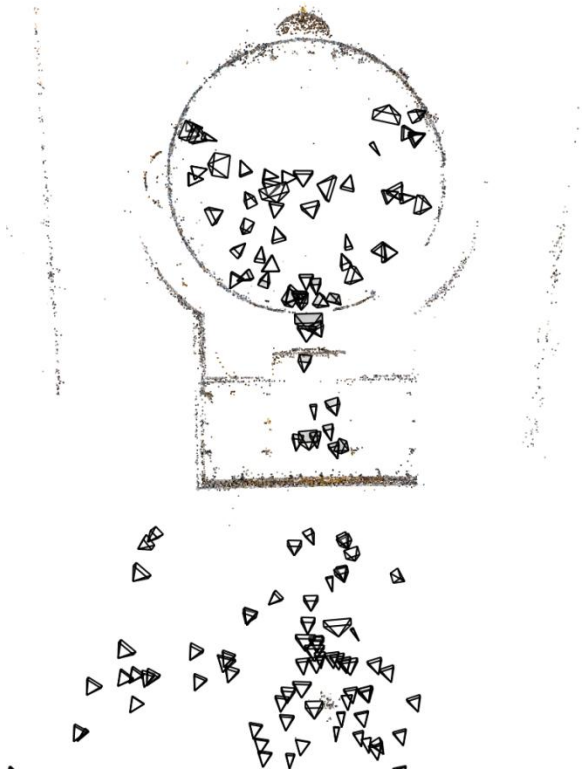
Pantheon



Full graph

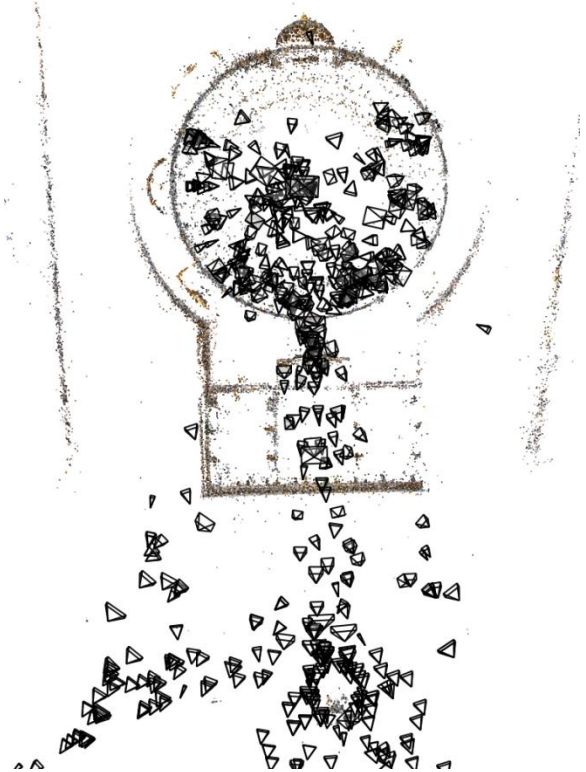


Skeletal graph (t=16)



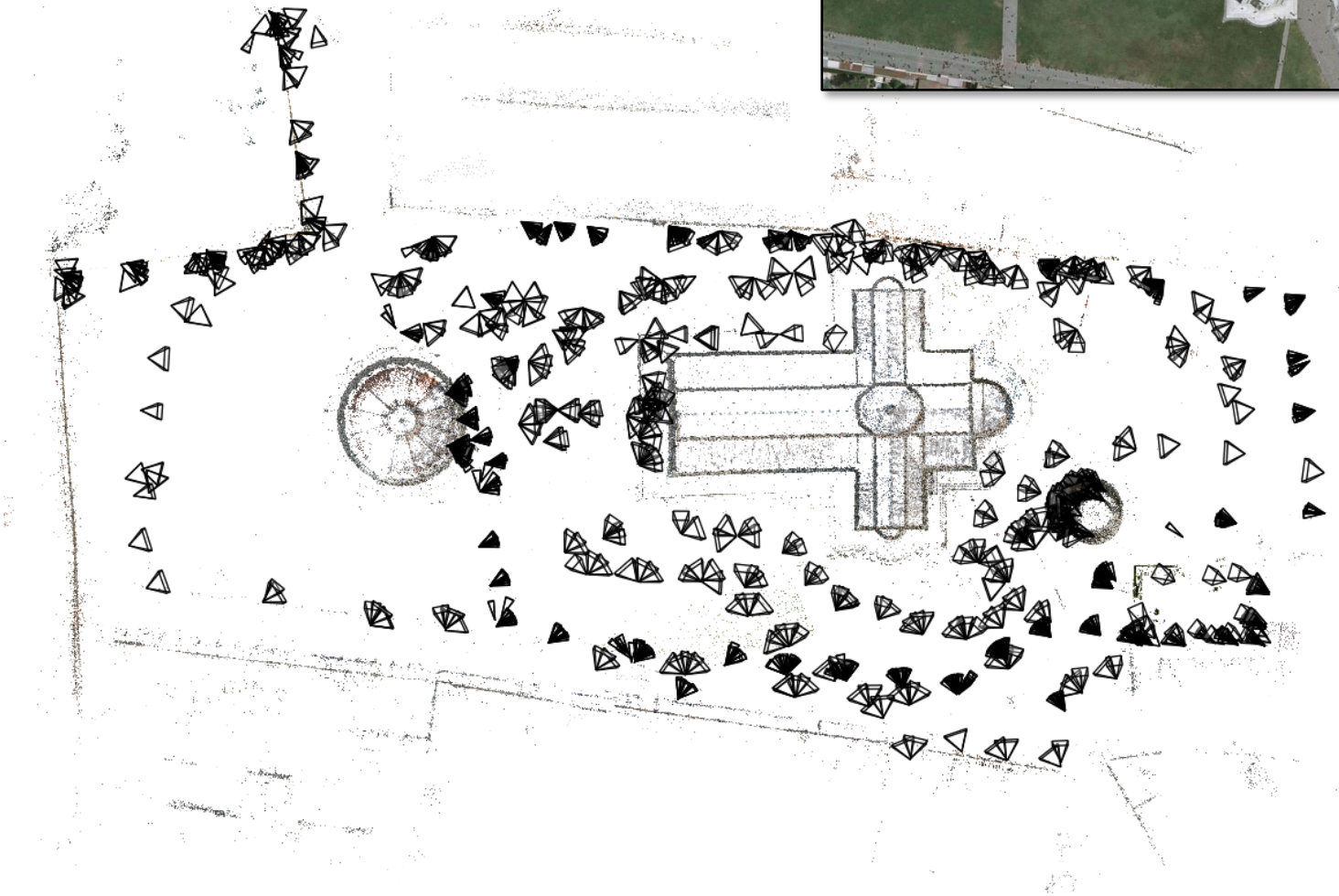
Skeletal reconstruction
101 images

After adding leaves
579 images



After final optimization
579 images

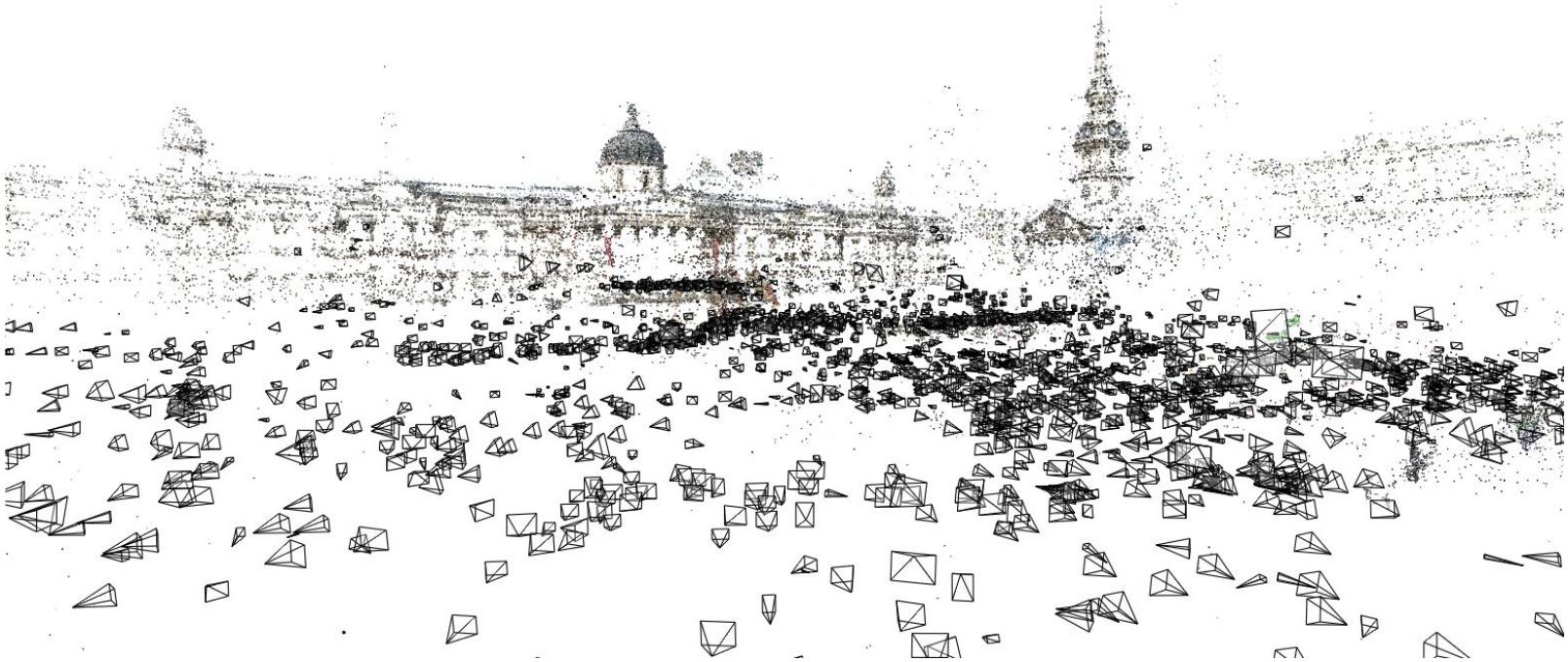
Pisa



1093 images registered (352 in skeletal set)

Trafalgar Square

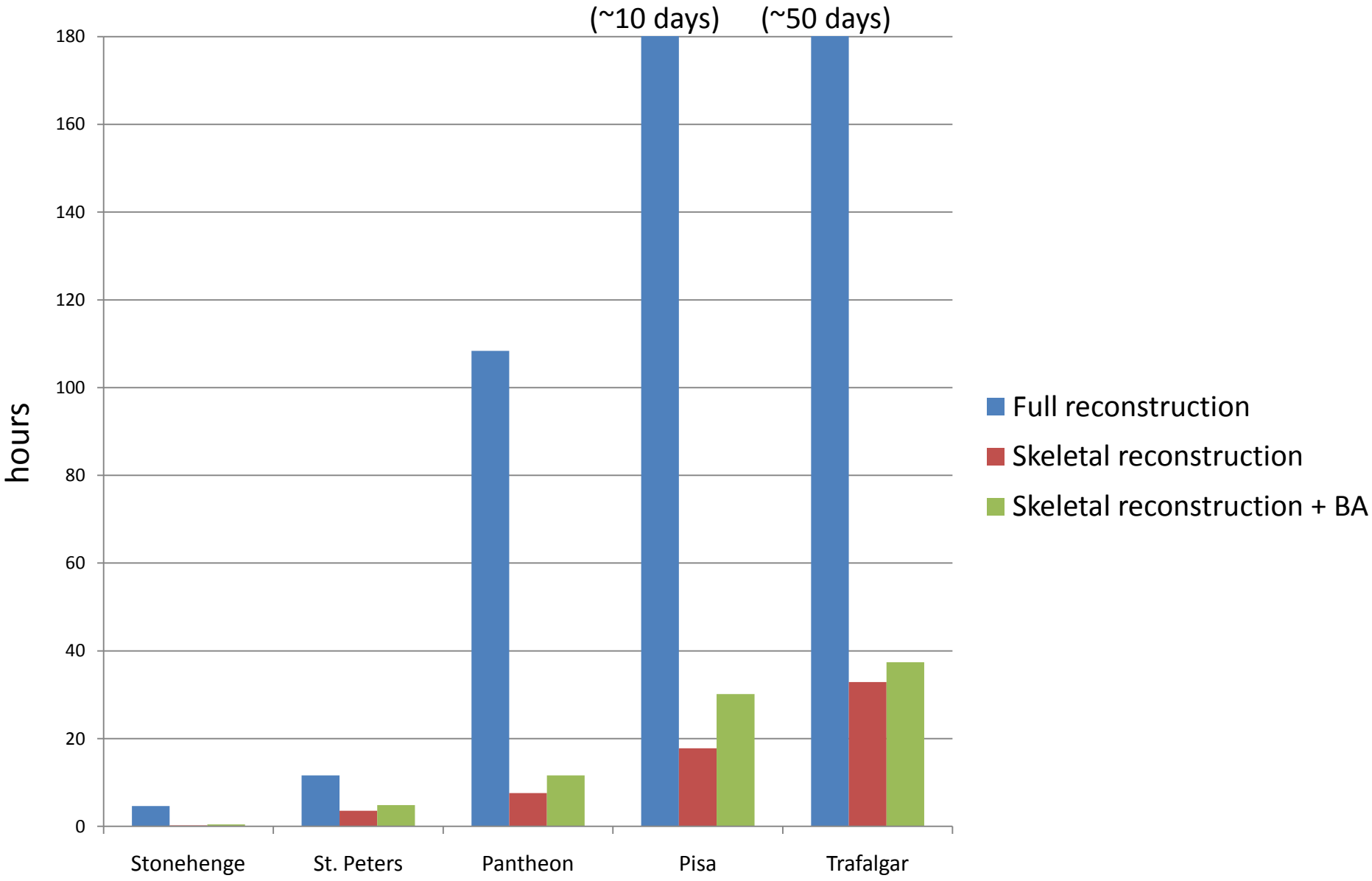
2973 images registered (277 in skeletal set)



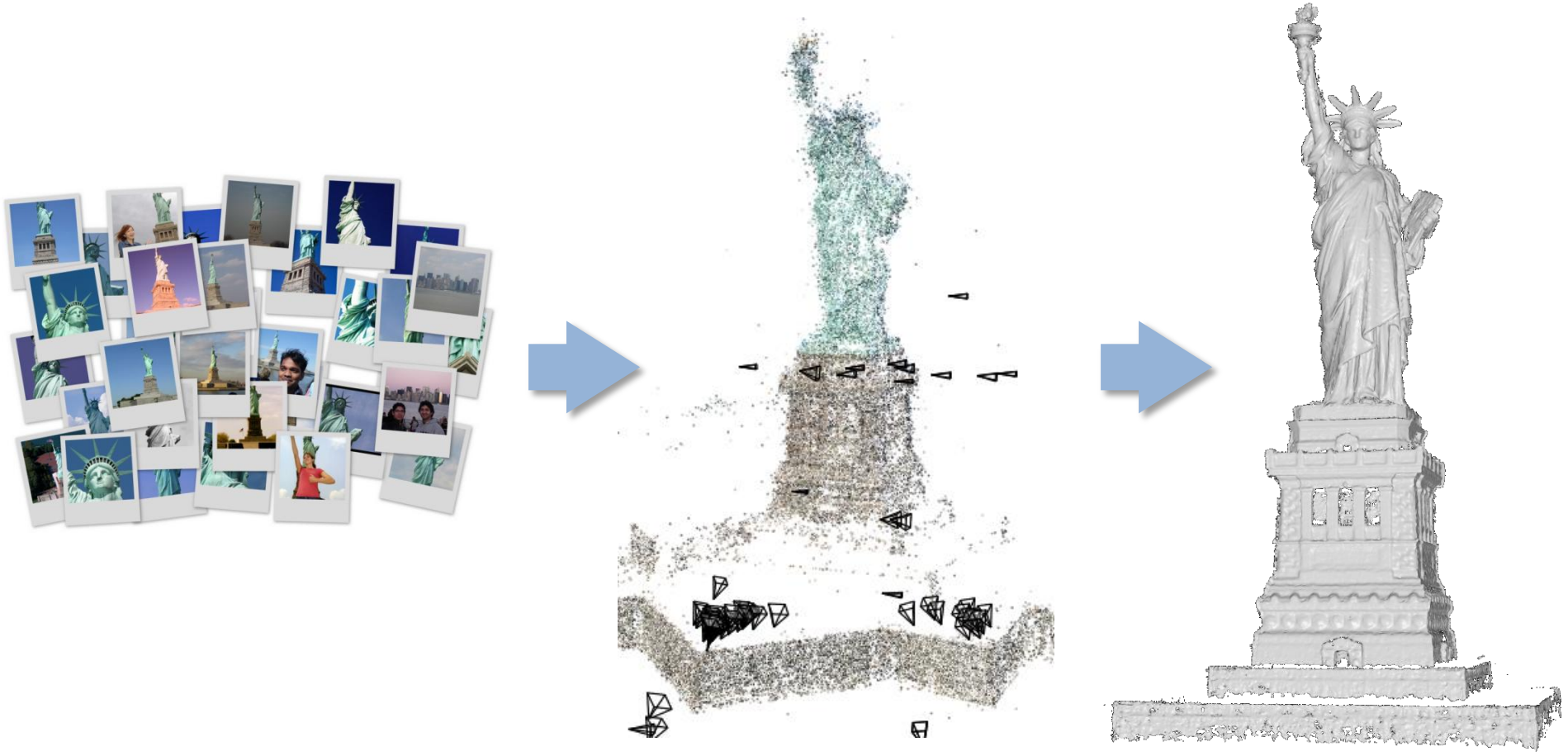
Statue of Liberty

7834 images registered (322 in skeletal set)

Running time

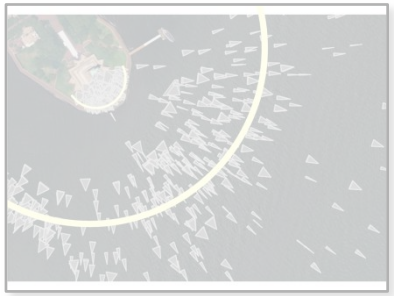


Dense 3D Modeling

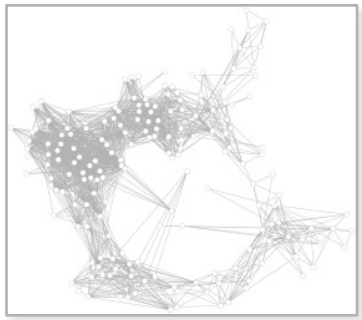


Michael Goesele, Noah Snavely, Brian Curless, Hugues Hoppe, Steve Seitz, ICCV 2007

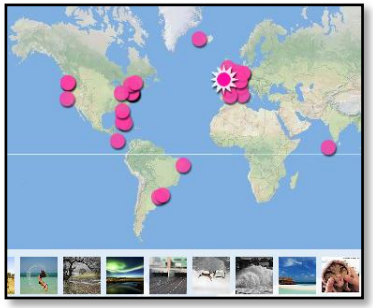
Overview



- Finding Paths through the World's Photos



- Large-scale 3D reconstruction



- Ongoing and future projects

Rebuilding Rome

- How much of the city of Rome can be reconstructed from Internet photos?

flickr

Signed in as Jimantha | Help | Sign Out

Home | You | Organize | Contacts | Groups | Explore

Search everyone's phot | Search

Can't see your photos? Find out why...

Search

Photos | Groups | People

Everyone's Photos | rome or roma | SEARCH | Advanced Search | Search by Camera

Full text | Tags only

We found **2,379,801 results** matching rome or roma.

View as slideshow

View: Most relevant • Most recent • Most interesting

Show: Details • Thumbnails

From Giampaolo...

From Kipourax

From MrMass

From rocdam

From alessandro...

From * Toshio *

From Optical...

From egold

From egold

From egold

From donato.chiru...

From cuellar

From egold

From Aquilant

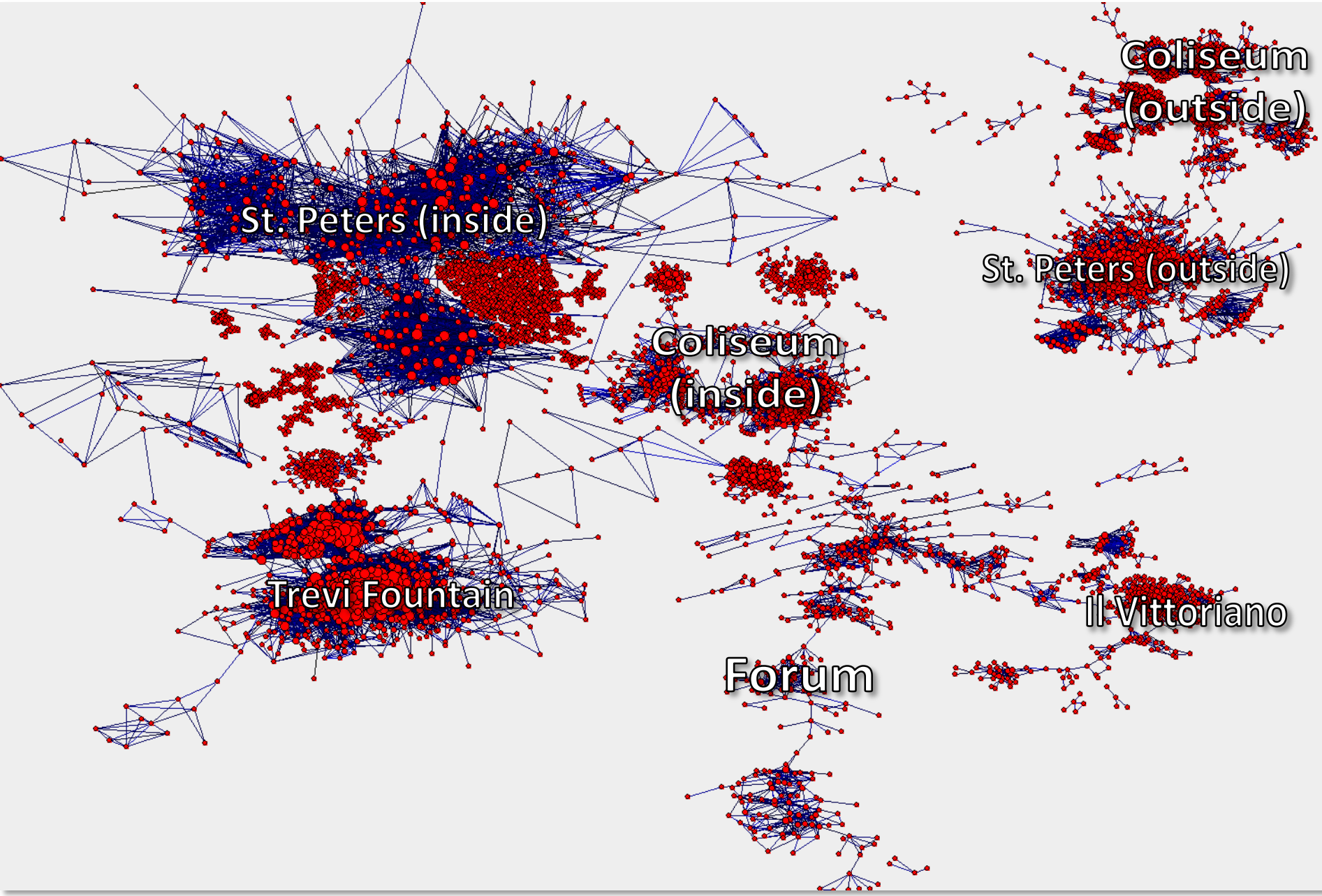
From Peter...

From SDBryan

From cuellar

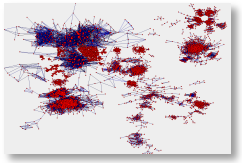
From david.bank

1,000,000 images of Rome


Coliseum
(outside)
St. Peters (inside)
St. Peters (outside)
Coliseum
(inside)
Trevi Fountain
Il Vittoriano
Forum

Rebuilding Rome in a day

- Download a million (or more) photographs of Rome from Flickr.com.

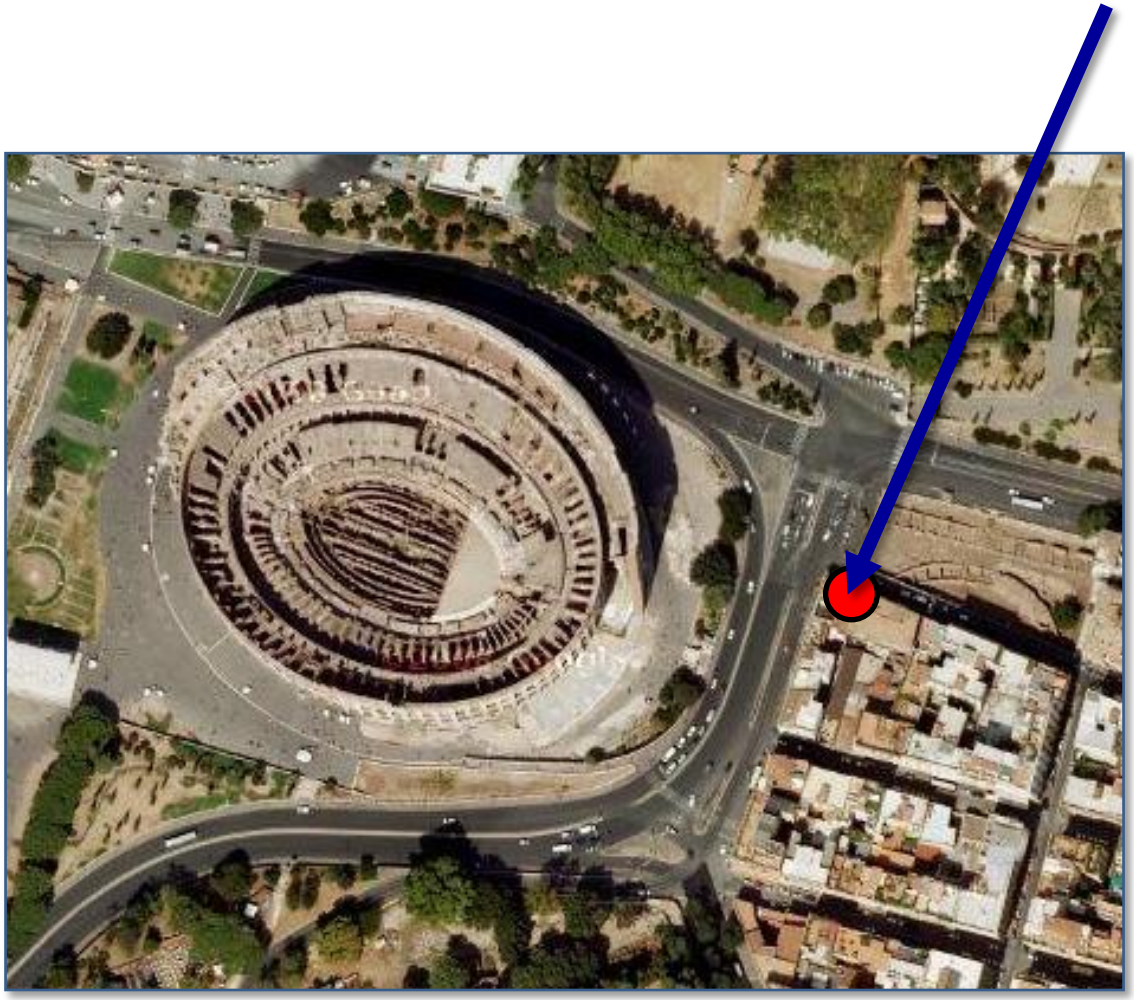


- Match the photos to find corresponding points.

- Build a three dimensional model of the city by incrementally adding photographs to it.

- Do all of the above in a fully distributed manner on a 1000 node cluster in 24 hours.

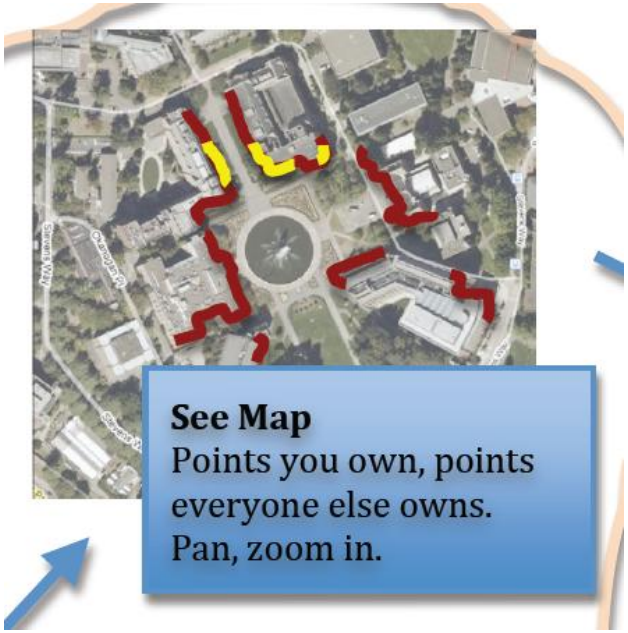
How do we fill in the gaps?



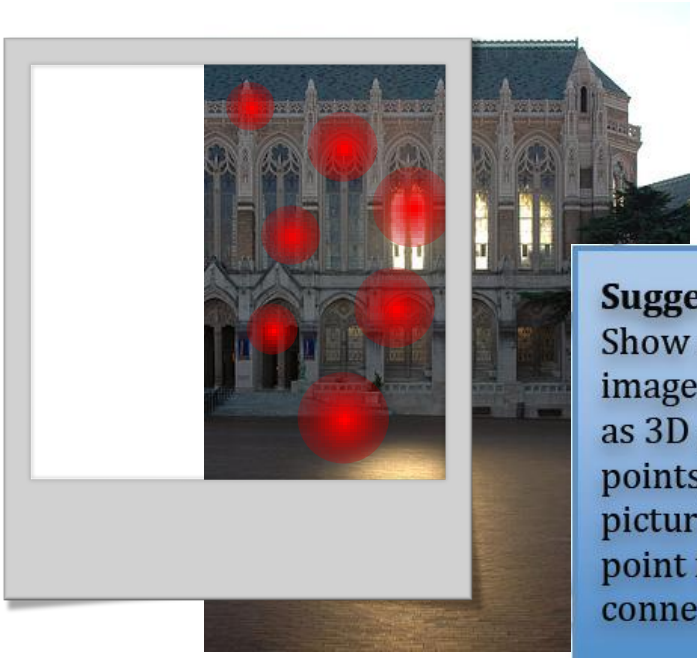
Can we turn this into a game?

- Use the millions of digital cameras / cellphones as a distributed world capture device

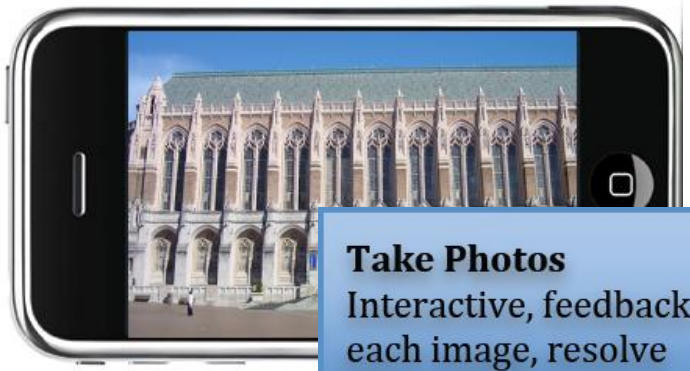
- Use humans to help where computer vision fails



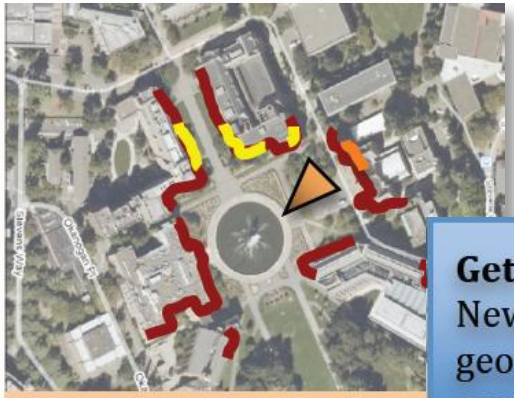
See Map
Points you own, points everyone else owns.
Pan, zoom in.



Suggestion View
Show nearby "edge" images, features marked as 3D points and non-3D points, try to take pictures of the non-point features that still connect to model.

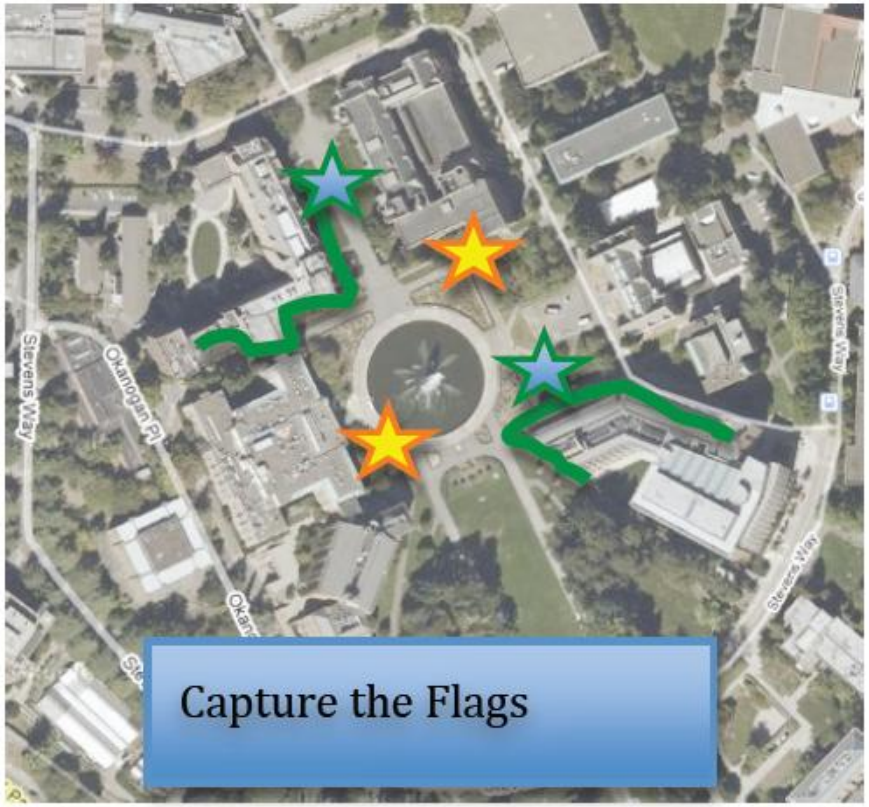


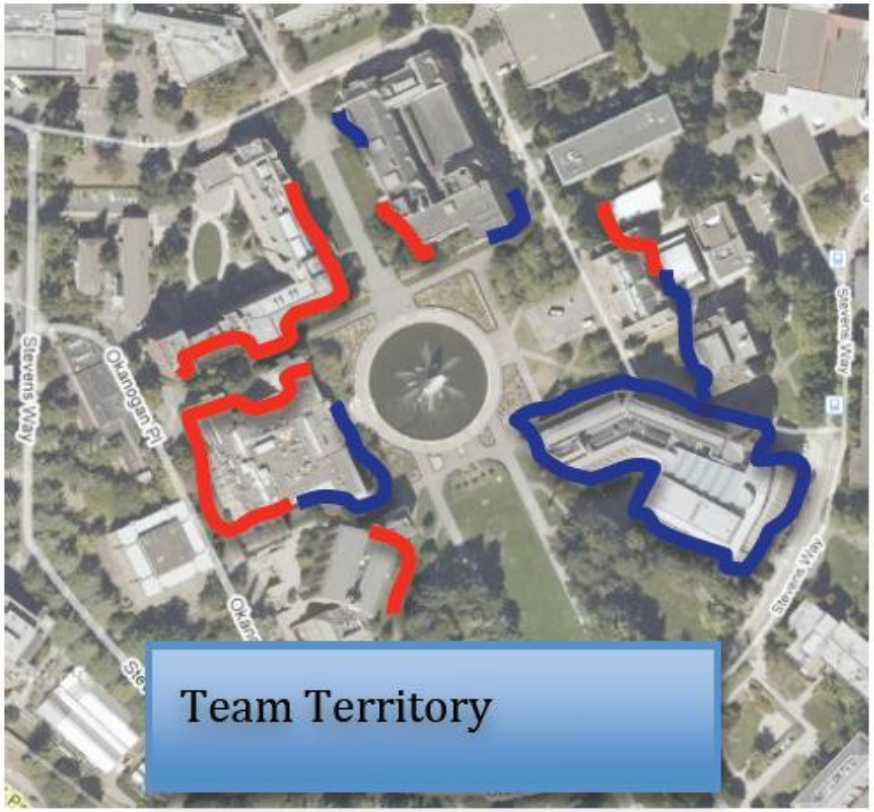
Take Photos
Interactive, feedback for each image, resolve ambiguities.



1 photo added
40 points created
150 points enhanced
New score: 4,500!

Get Back Report
New score, show added geometry and new camera positions.


Stevens Way
Okanogan Pl
Stevens Way
Capture the Flags


Stevens Way
Okanogan Pl
Stevens Way
Team Territory


Thank you!