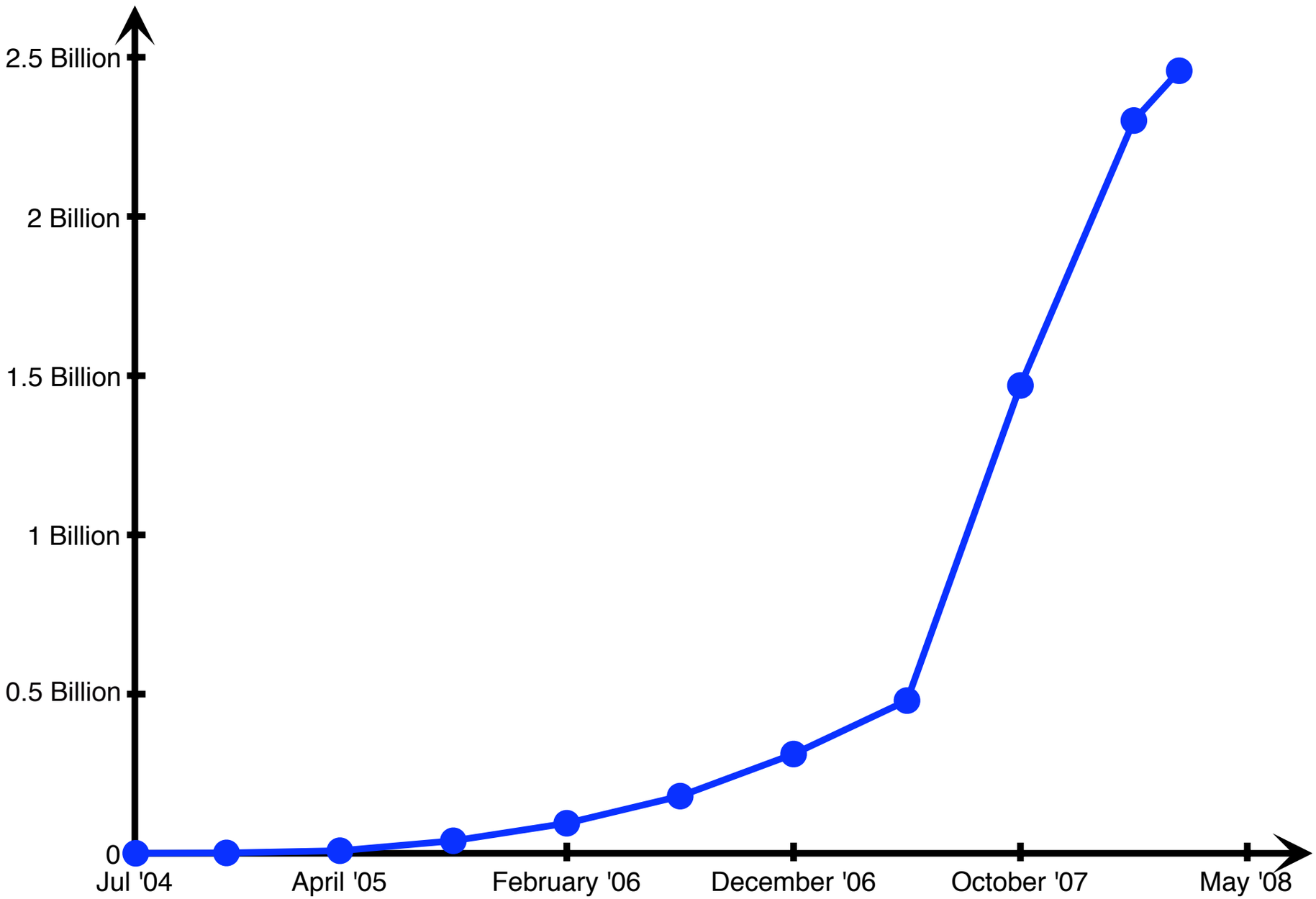
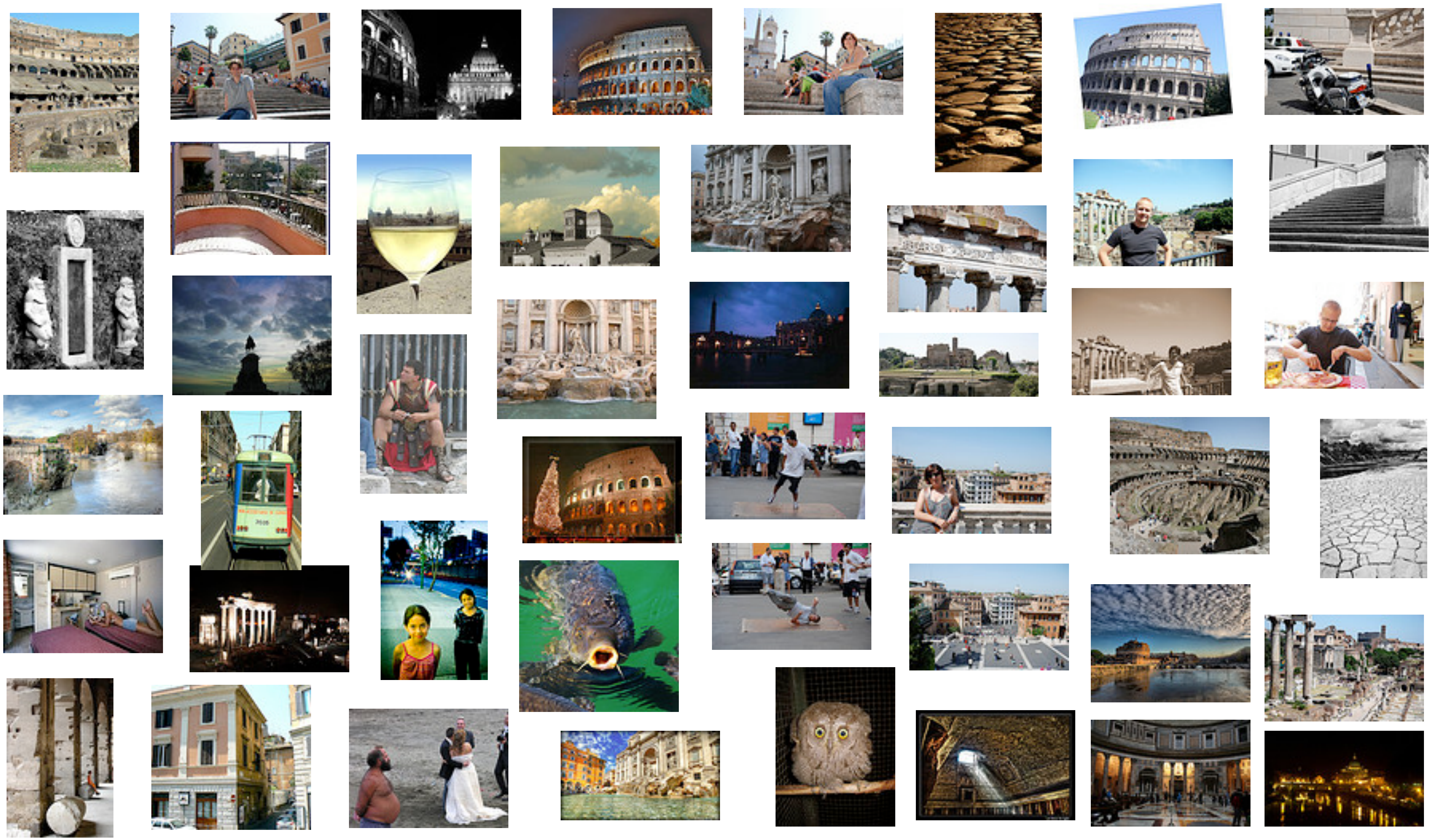
Building Rome in a Day

Sameer Agarwal, Google

Flickr growth



Rome via Flickr



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: **Most relevant** • Most recent • Most interesting Show: Details • **Thumbnails**

View as slideshow

From NupurB

From BRUNO MÉNDEZ...

From * Toshio *

From *Maqik &...

From Christopher...

From Nastrina1981

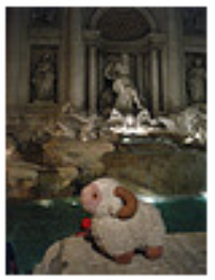
From jaki good

From Fabrizio...

From ···Skazi▪▪

From Alida's...

From hbomb1947

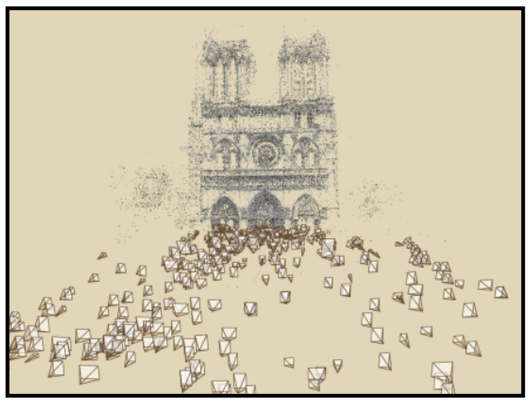


From David d'O

Photo Tourism

Input photographs

Scene reconstruction



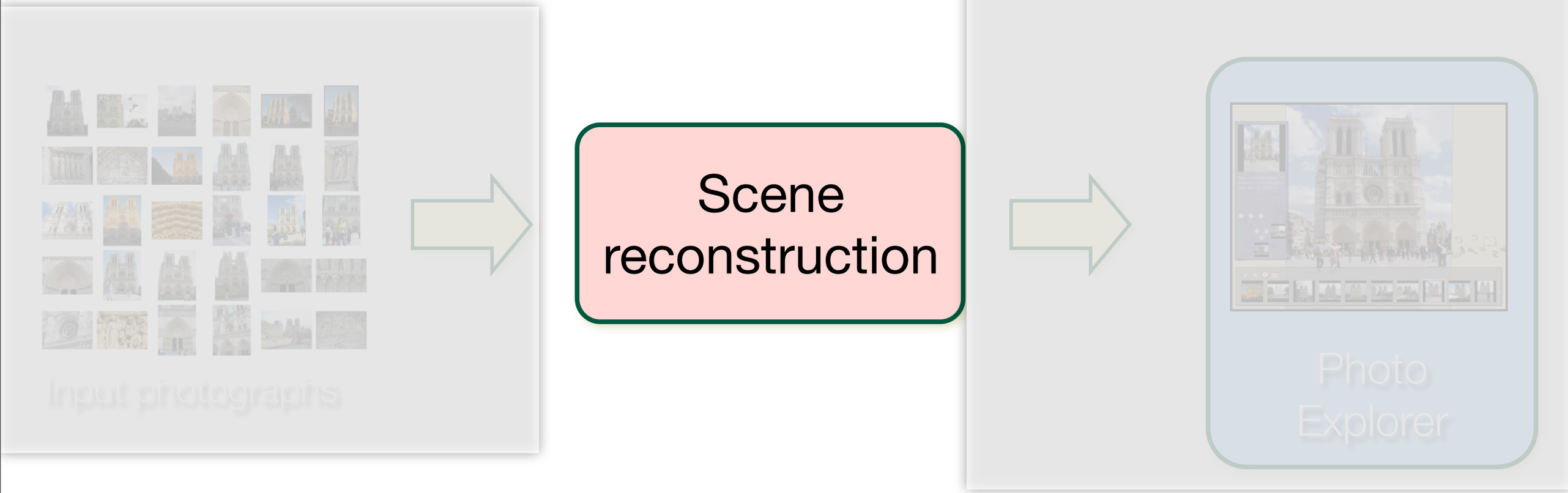
Relative camera positions and orientations

Point cloud

Sparse correspondence

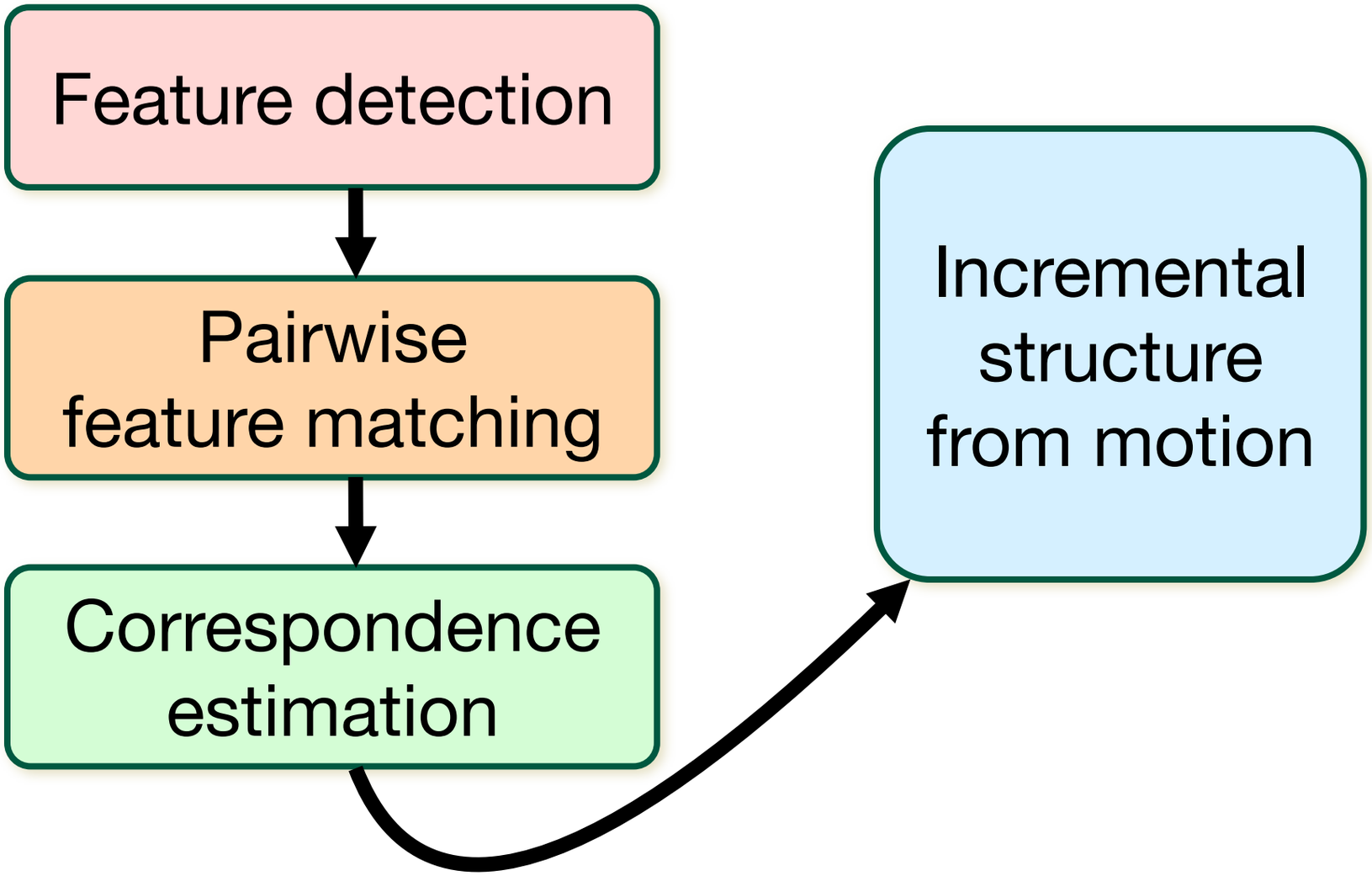
Photo Explorer

Photo Tourism overview



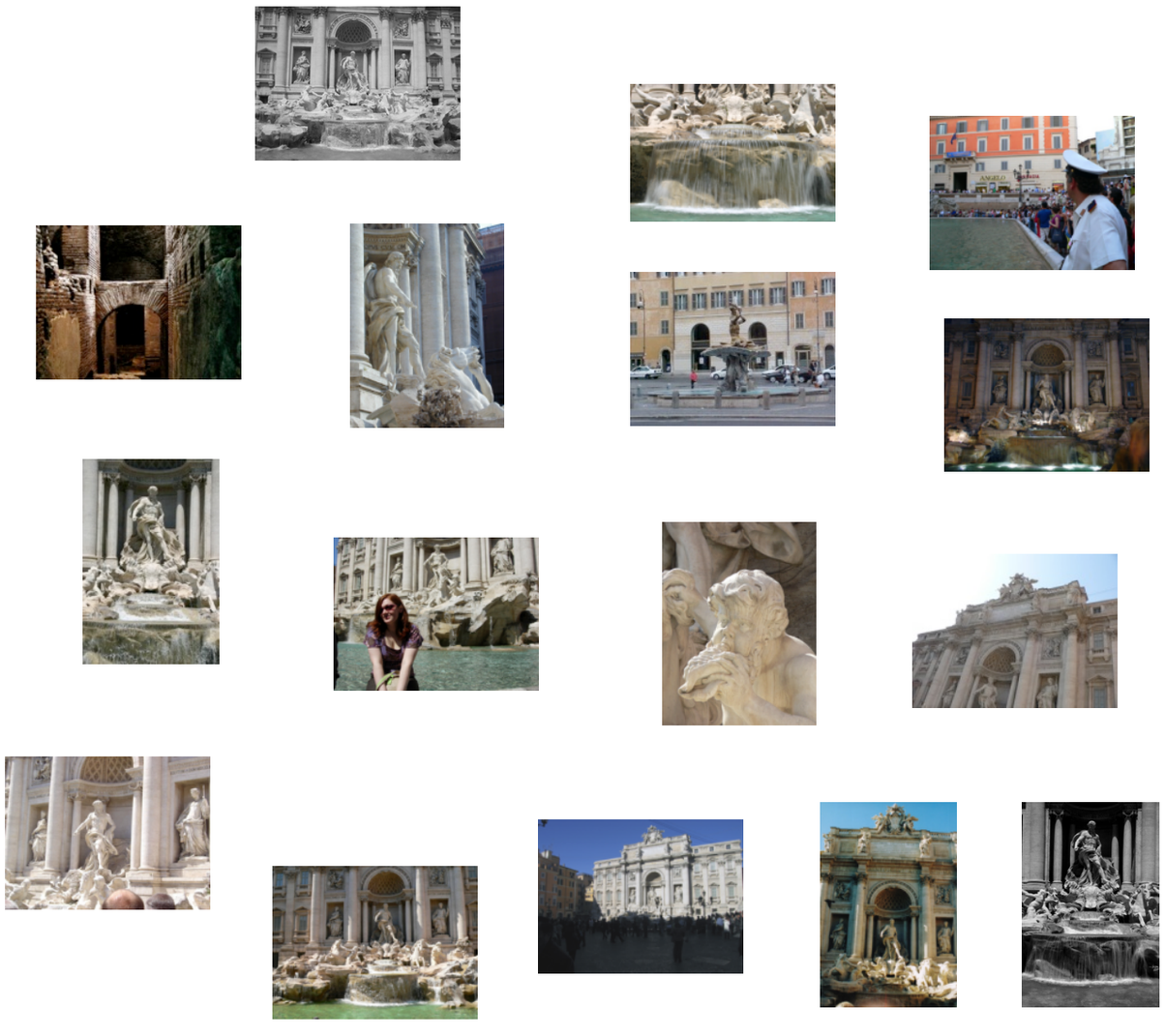
Scene reconstruction

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



Feature detection

Detect features using SIFT [Lowe, IJCV 2004]

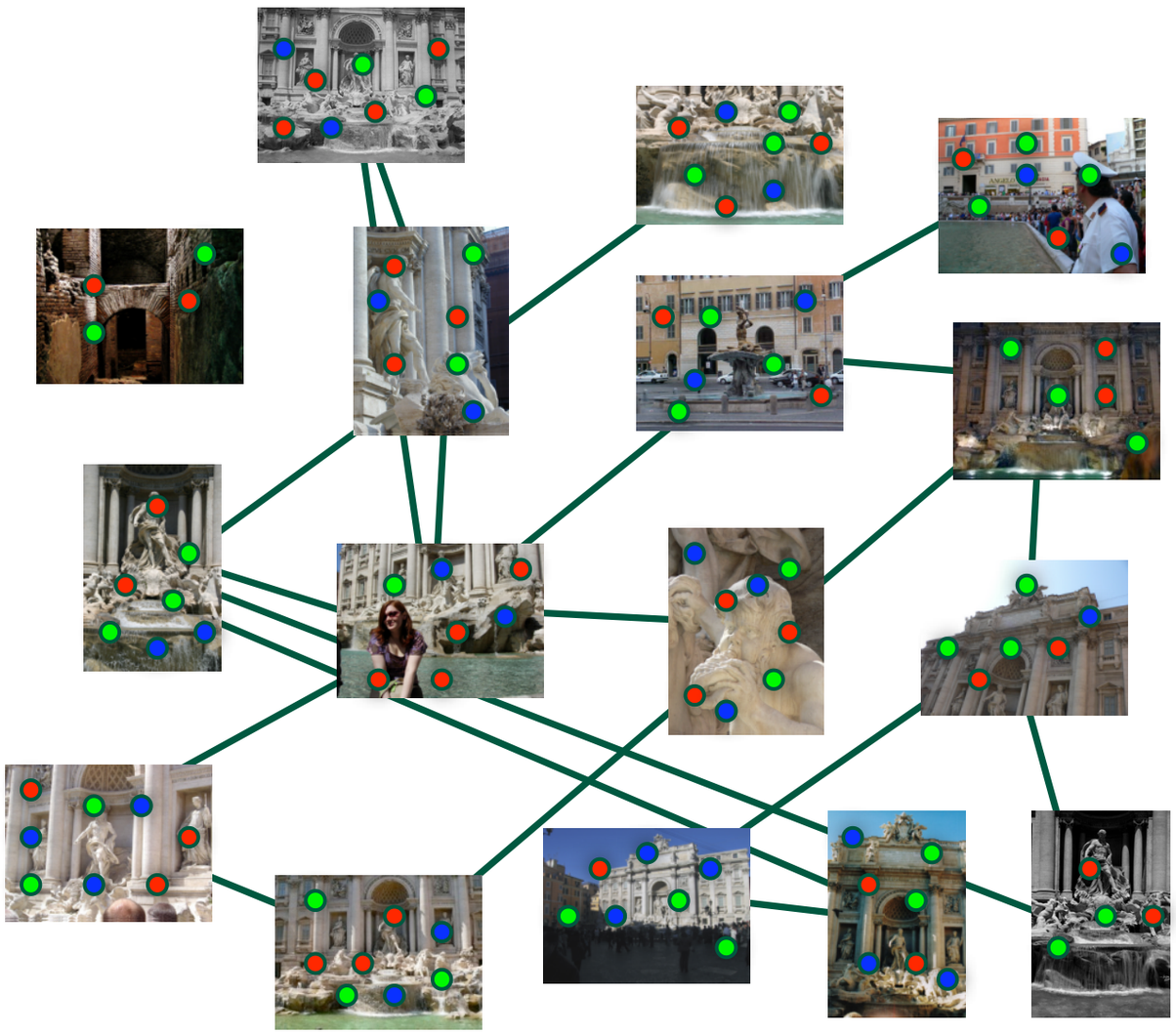


Feature detection

Detect features using SIFT [Lowe, IJCV 2004]

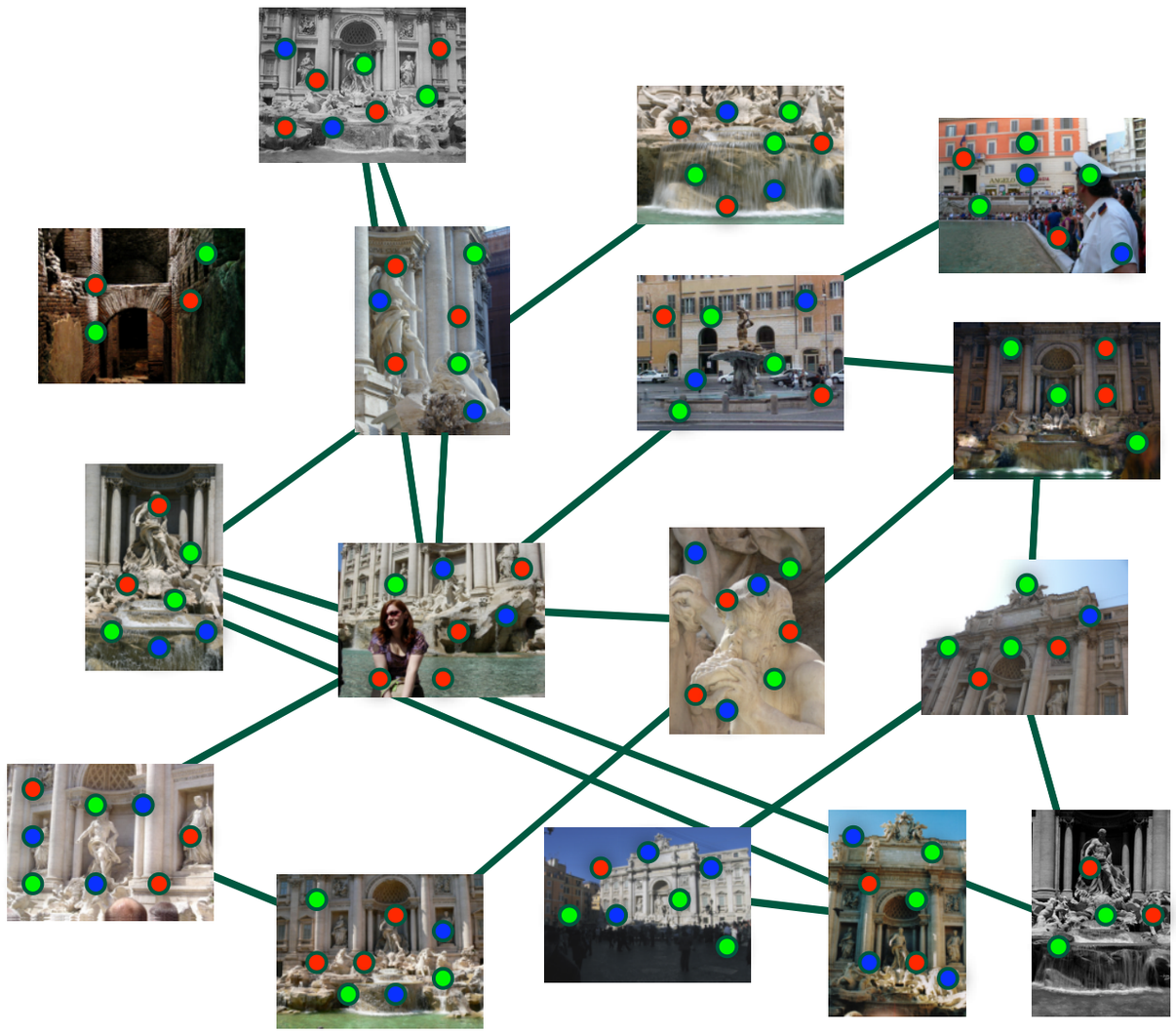
Feature matching

Match features between each pair of images



Feature matching

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



Correspondence estimation

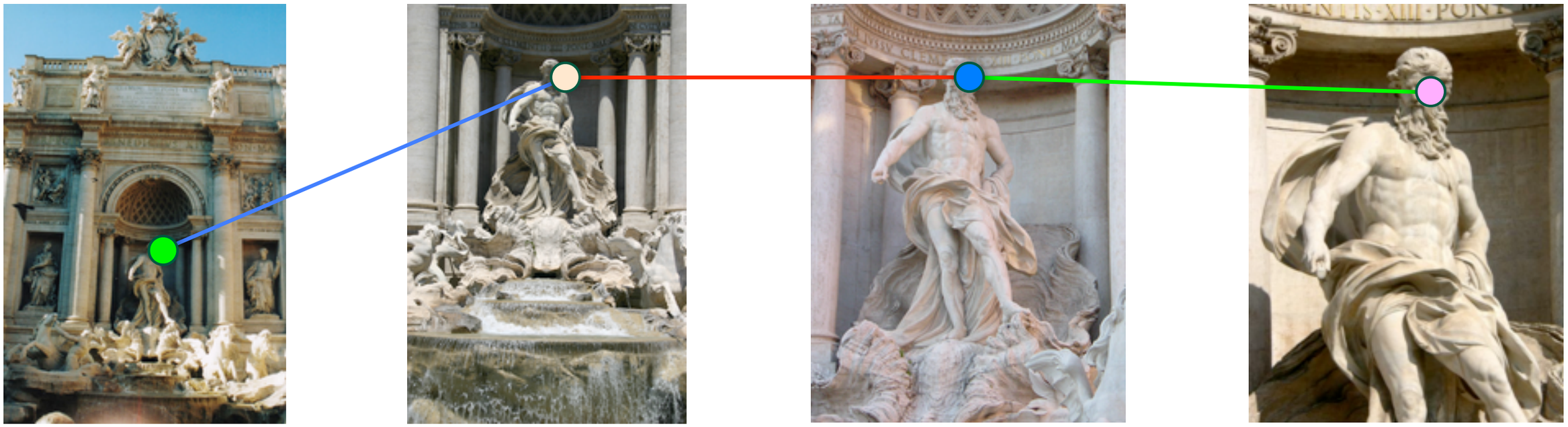


Image 1 Image 2 Image 3 Image 4

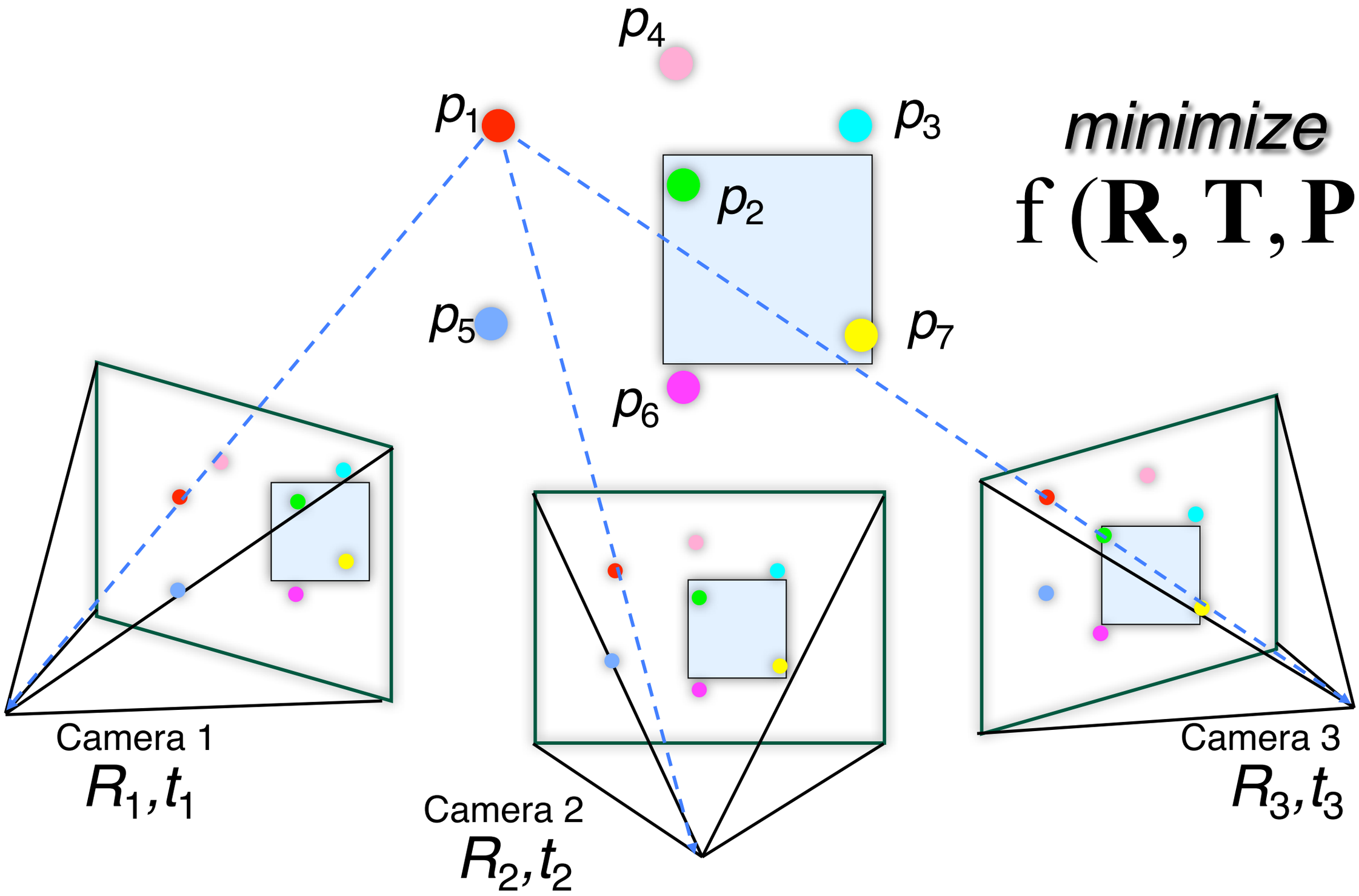
Link up pairwise matches to form connected components of matches across several images

Structure from motion

- Given many points in *correspondence* across several images, $\{(u_{ij},v_{ij})\}$, simultaneously compute the 3D location $\mathbf{p}_i$ and camera (or *motion*) parameters ($\mathbf{R}_j$, $\mathbf{t}_j$)

$$\begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} = f(R_j, t_j, p_i)$$

Structure from motion



minimize

$$\mathrm{f}(\mathbf{R}, \mathbf{T}, \mathbf{P})$$

Photo Tourism

Exploring photo collections in 3D

Noah Snavely Steven M. Seitz Richard Szeliski

University of Washington *Microsoft Research*

SIGGRAPH 2006

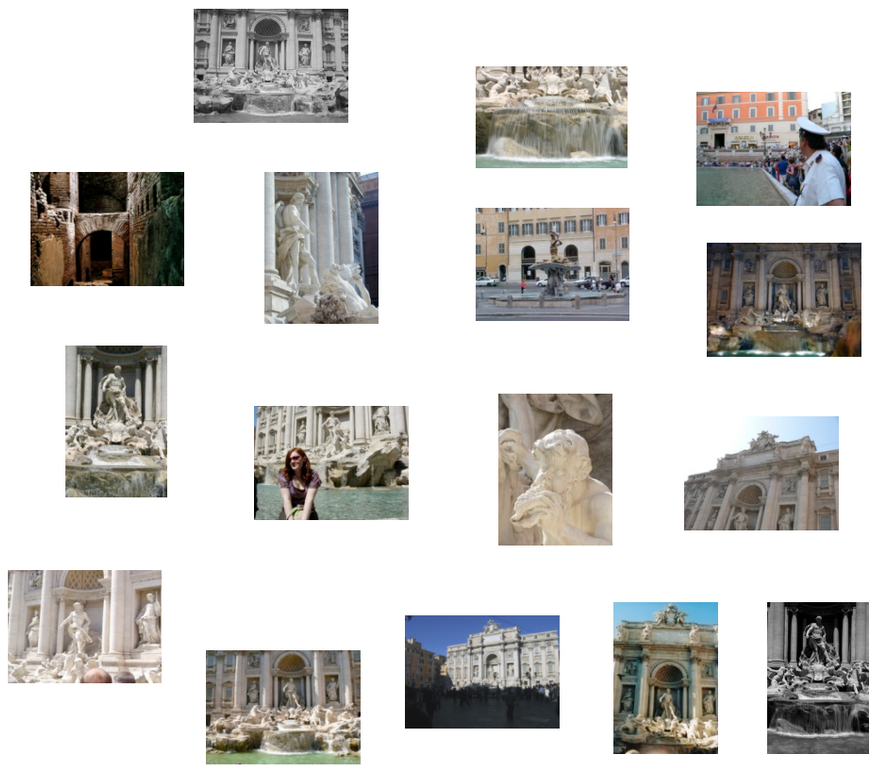
Microsoft Photosynth

Cities on the web

	Flickr	Picasa	images.google
Venice	1,300,000	7,800,000	12,000,000
Rome	2,600,000	26,000,000	20,000,000
Tokyo	3,200,000	12,000,000	19,800,000
New York	6,500,000	41,000,000	290,000,000
London	7,200,000	40,800,000	89,000,000

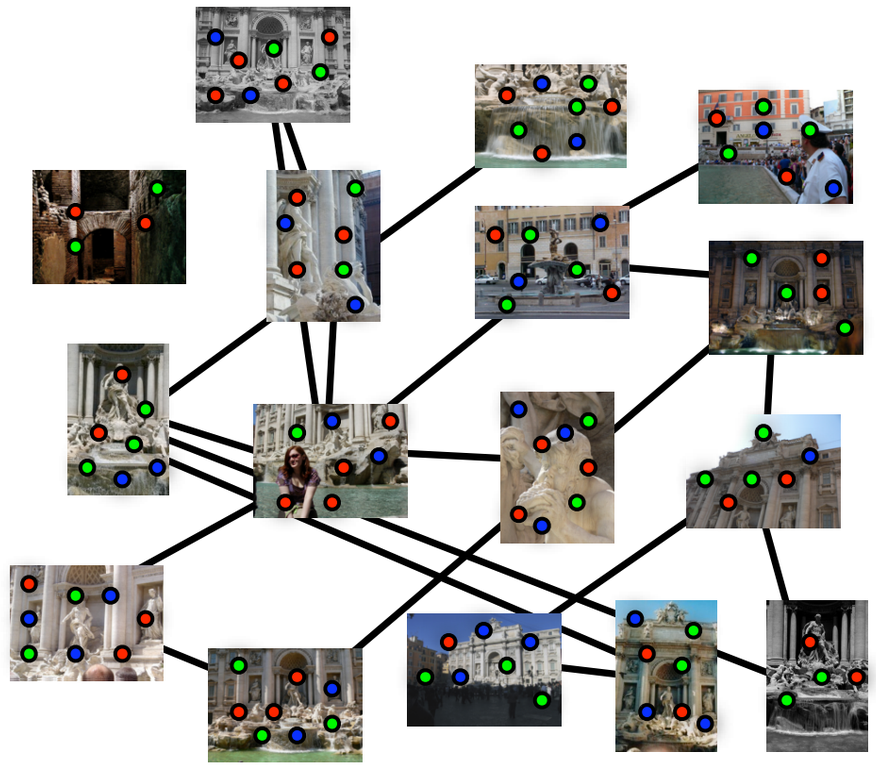
The Task

- Download a million images of Rome
- Match the images
- Build a 3D model of the city



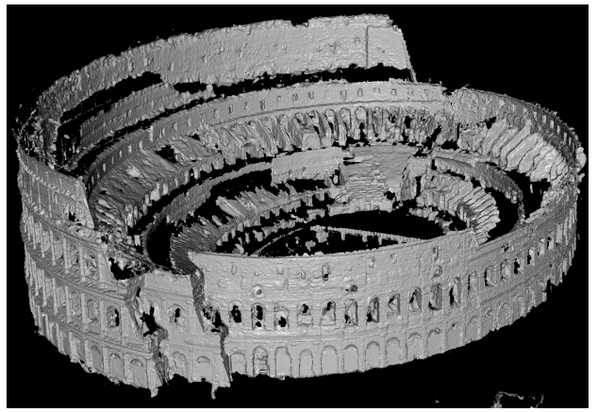
The Task

- Download a million images of Rome
- Match the images
- Build a 3D model of the city



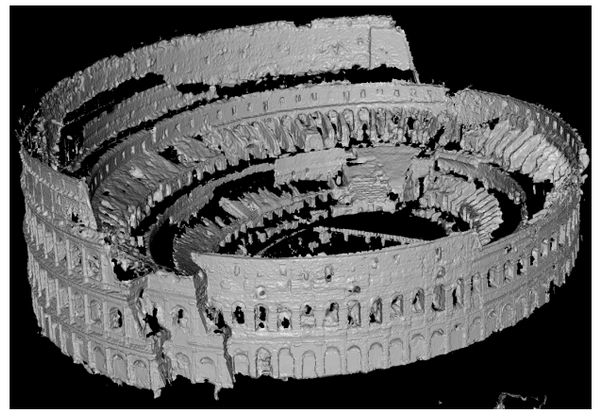
The Task

- Download a million images of Rome
- Match the images
- Build a 3D model of the city



The Task

- Download a million images of Rome
- Match the images
- Build a 3D model of the city



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

Why?

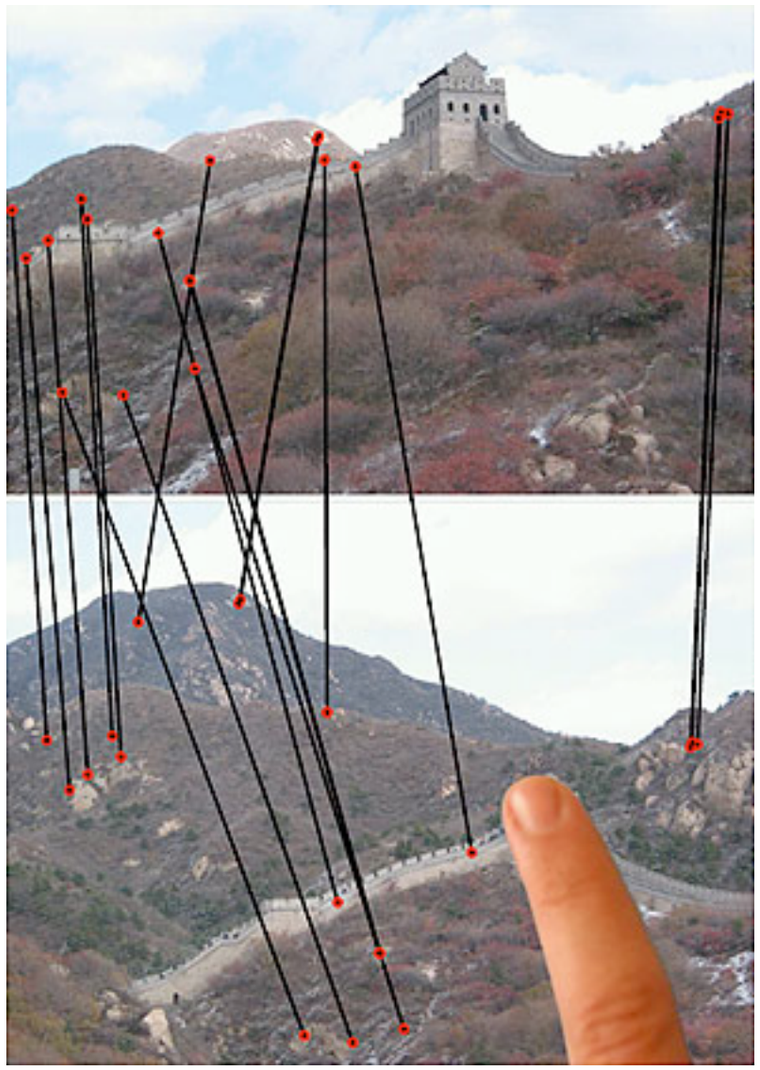
- Interiors, high level of geometric detail, texture maps.
- Better models for Google/Virtual Earth, GPS, virtual sets for movie production.
- Historical preservation.
- Urban geography.
- Games set in the real world, Photocity, Grand Theft Auto “Roma”.

Our Approach

1. Scrape images
2. Extract Features
3. Match Images
4. Reconstruct sparse image set (Skeletal Sets)
5. Reconstruct full image set (Bundle Adjustment)

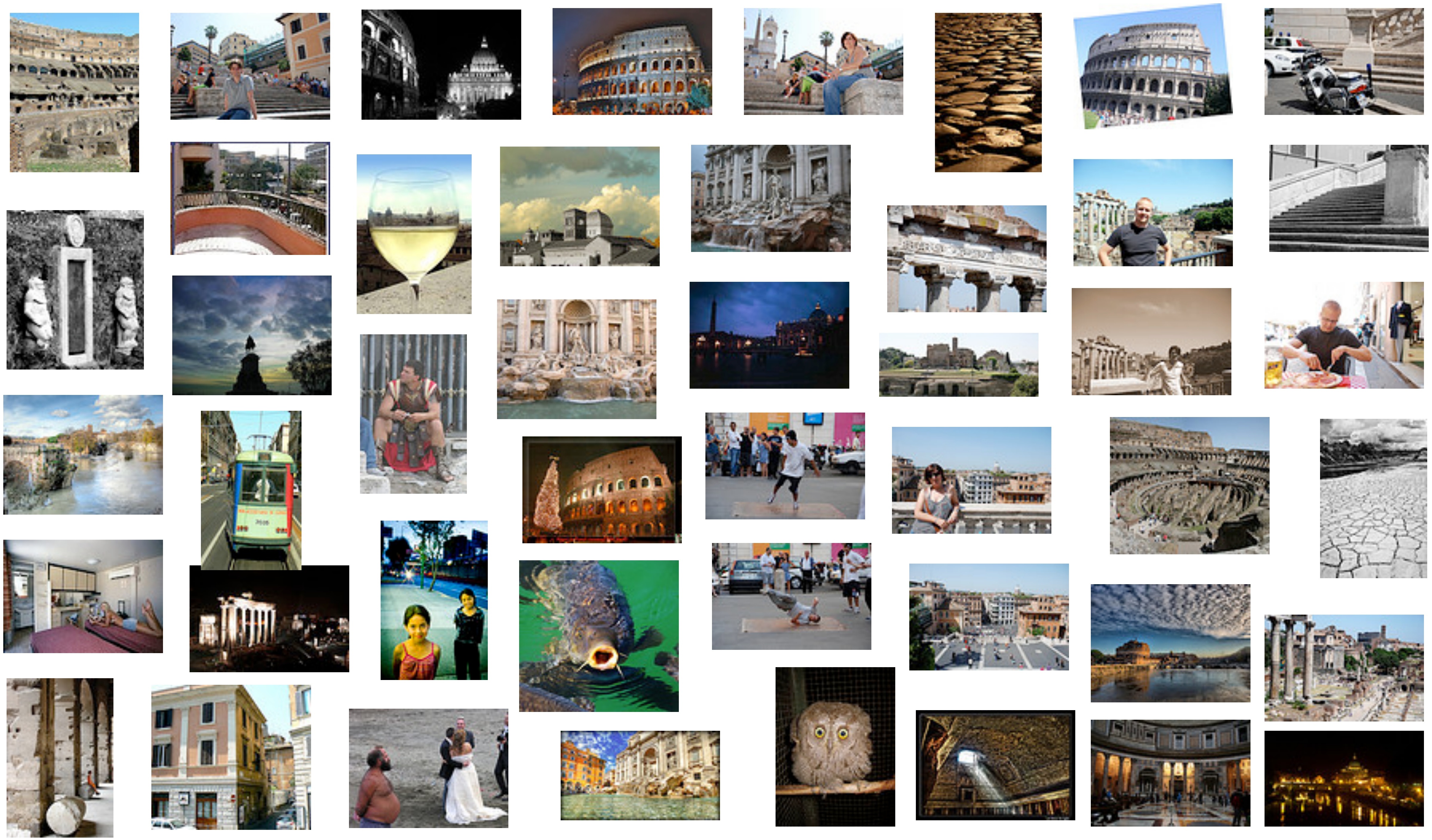
Image Matching

Find points across images which correspond to the same point in the world.

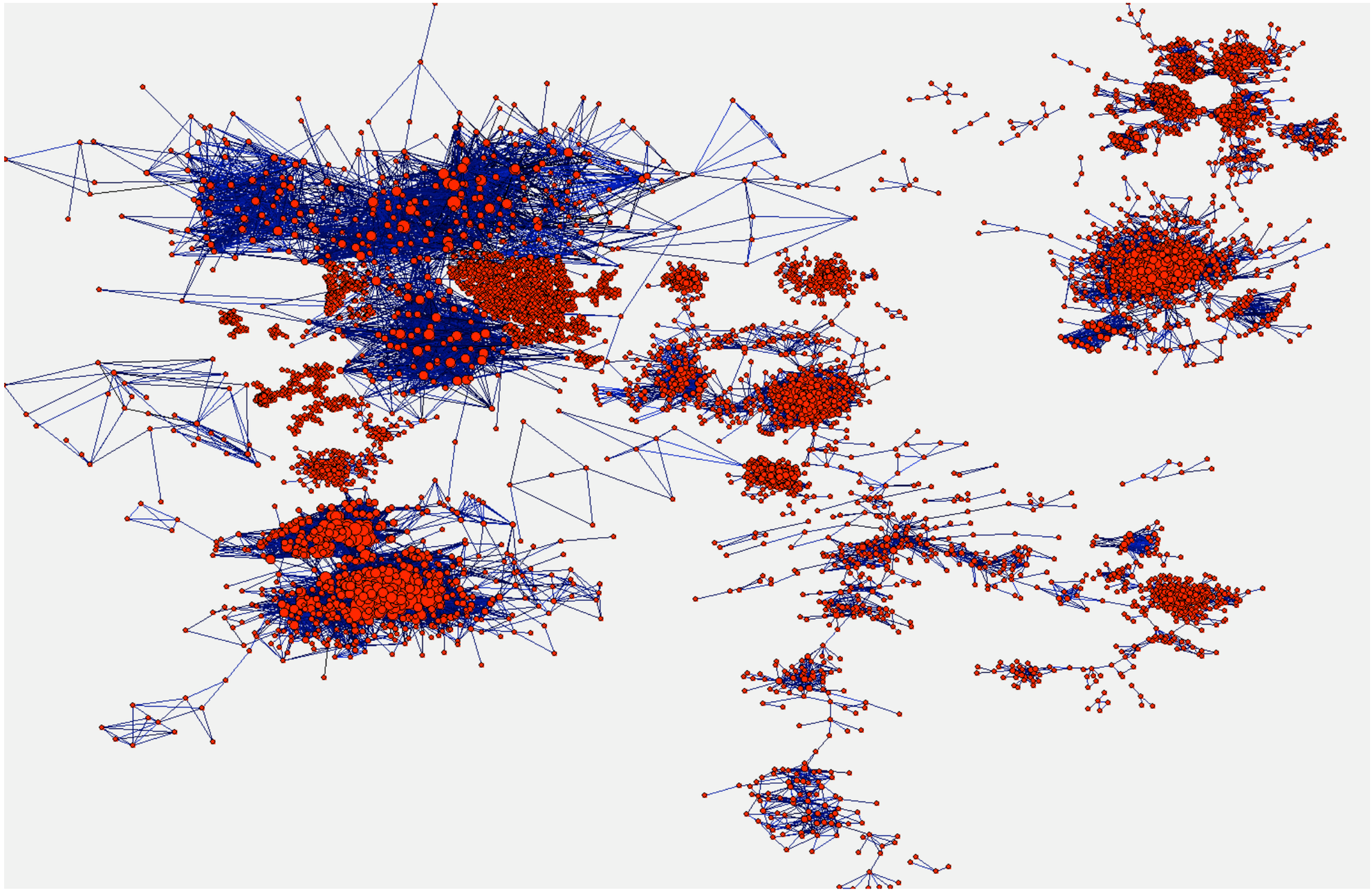


- All pairs matching is data parallel, but expensive in CPU and network bandwidth (~10TB).
- 0.5 Trillion pairwise comparisons.
- 10k matches/sec = 1.5 years.

Rome via Flickr

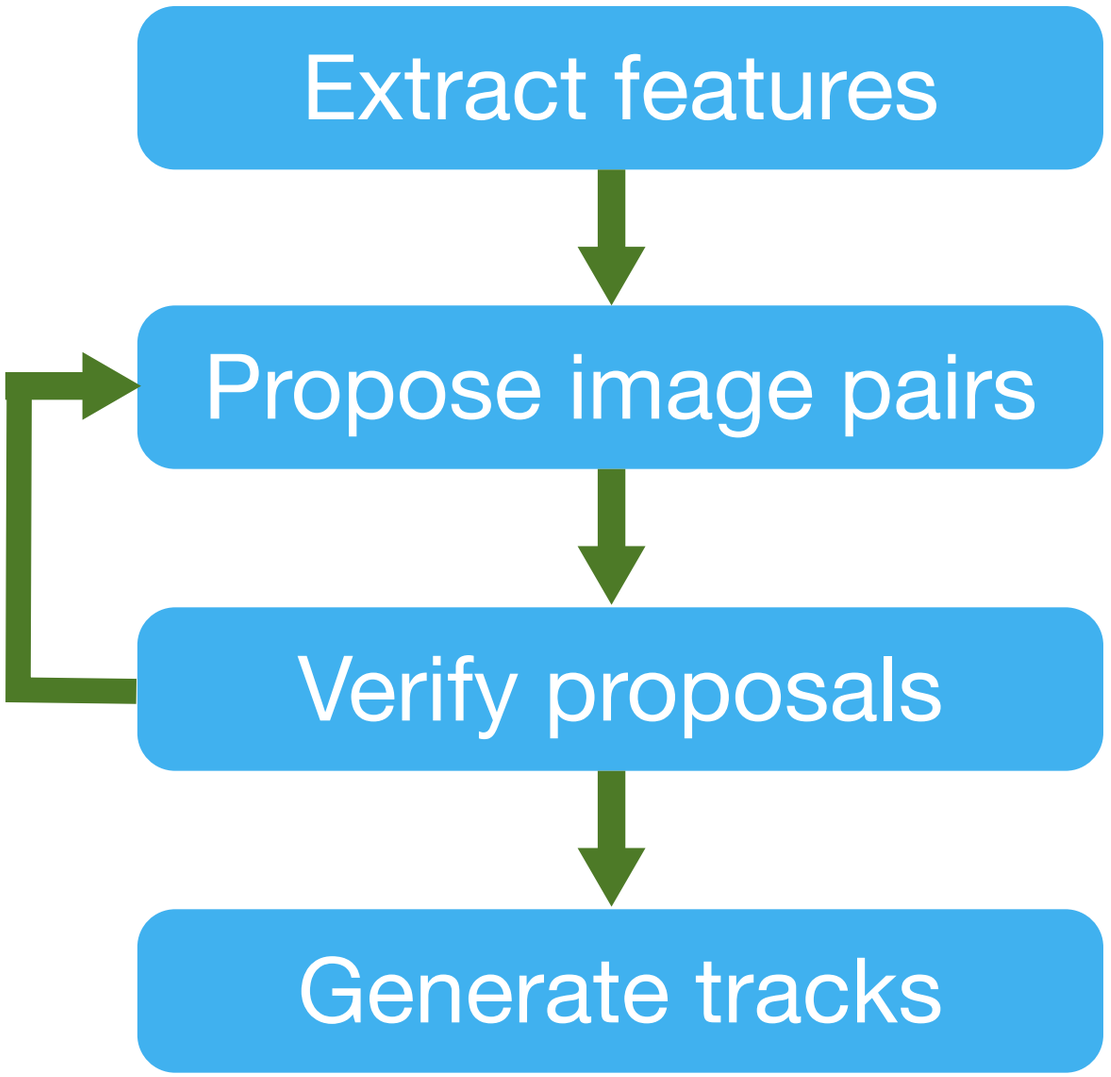


Rome



Matching Algorithm

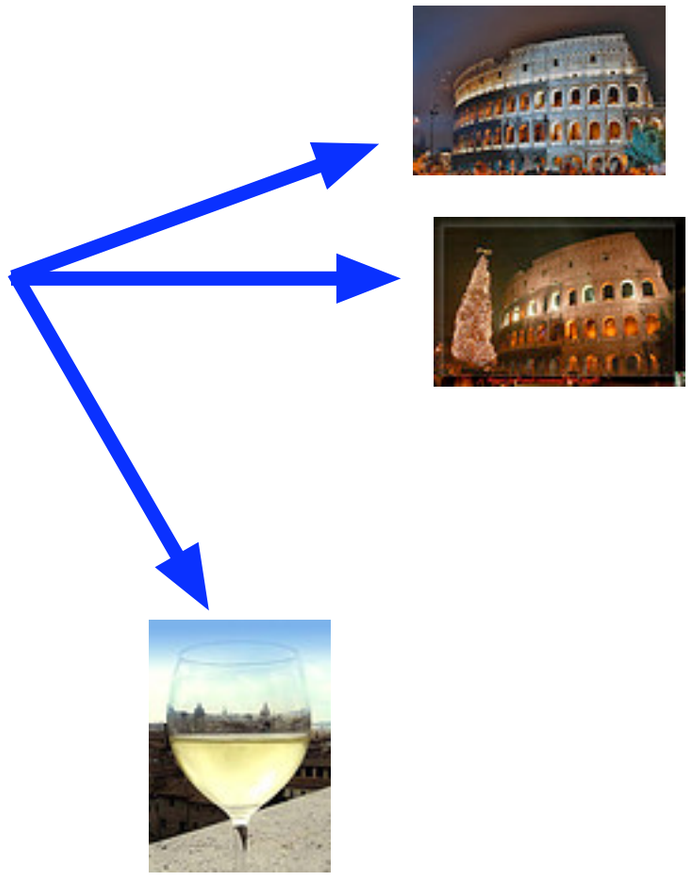
- Multi round propose and verify scheme
  - 2 rounds based on whole image similarity.
  - 4 rounds based on query expansion
- Verification = SIFT feature matching + RANSAC.



Whole Image Similarity

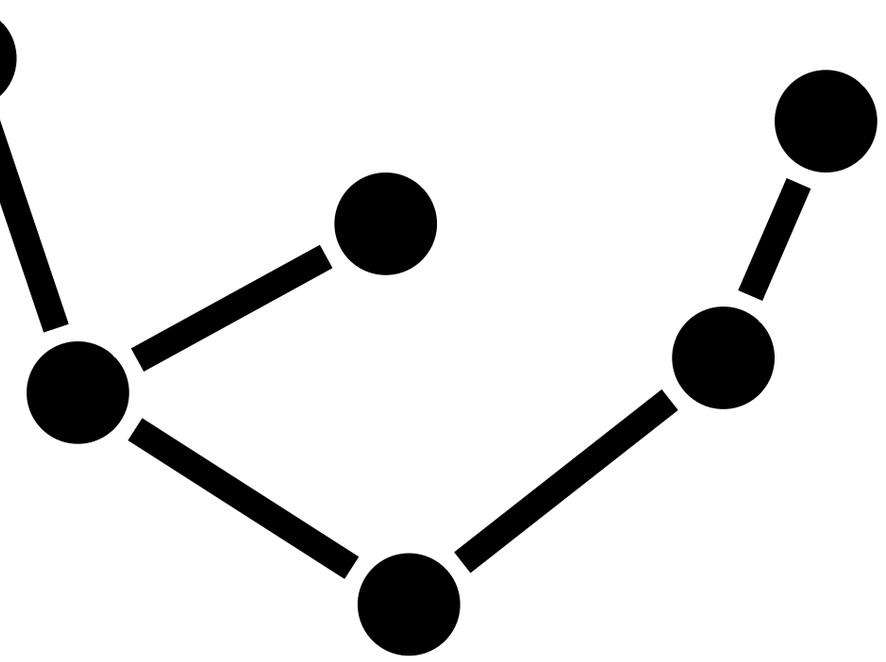
Text retrieval inspired approach.

- Represent images as high dimensional vectors using a vocabulary tree.
- Inner product between of vectors is the similarity between images.
- Top k scoring images are potential matches.



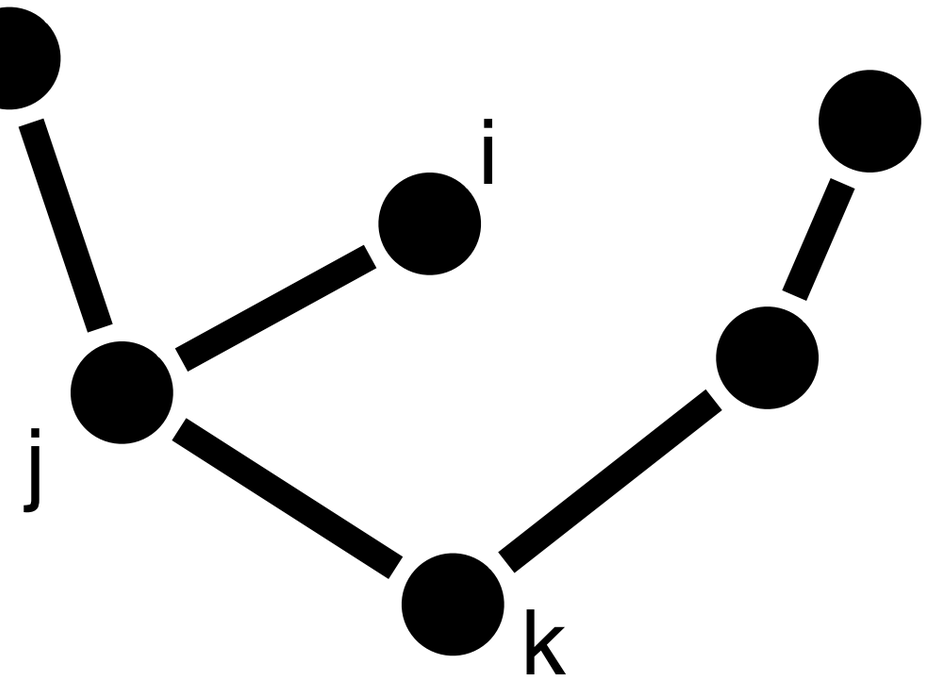
Nister and Stewenius, CVPR 2006

Query Expansion



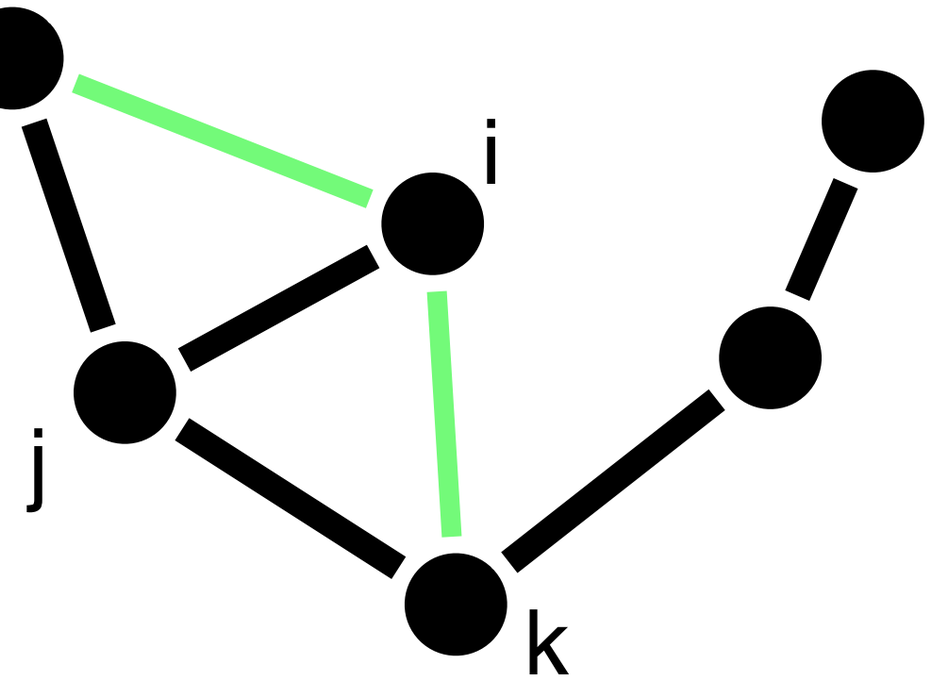
Chum et al, ICCV 2007

Query Expansion



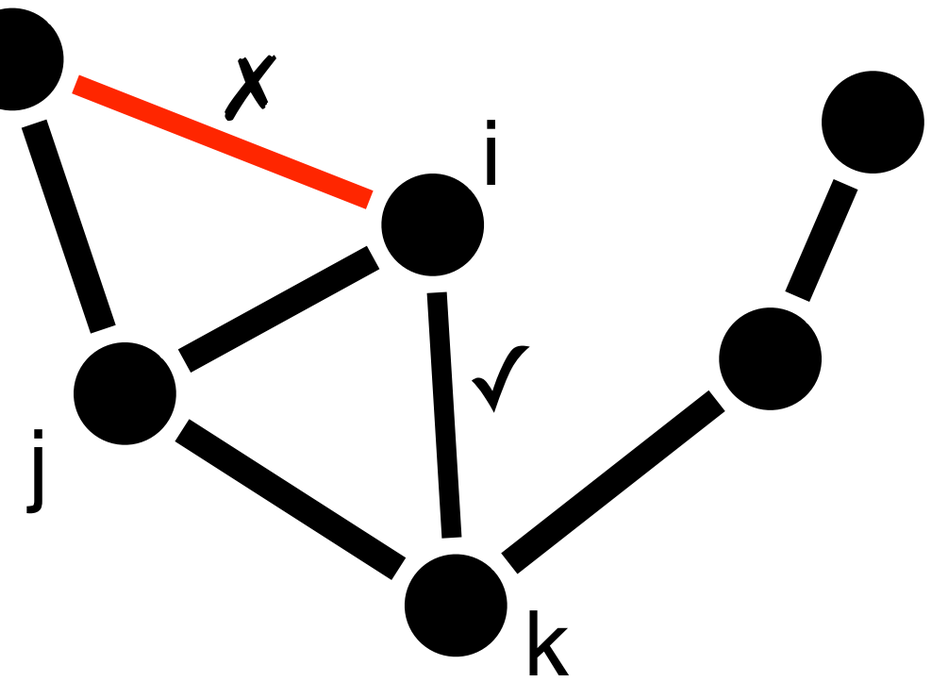
Chum et al, ICCV 2007

Query Expansion



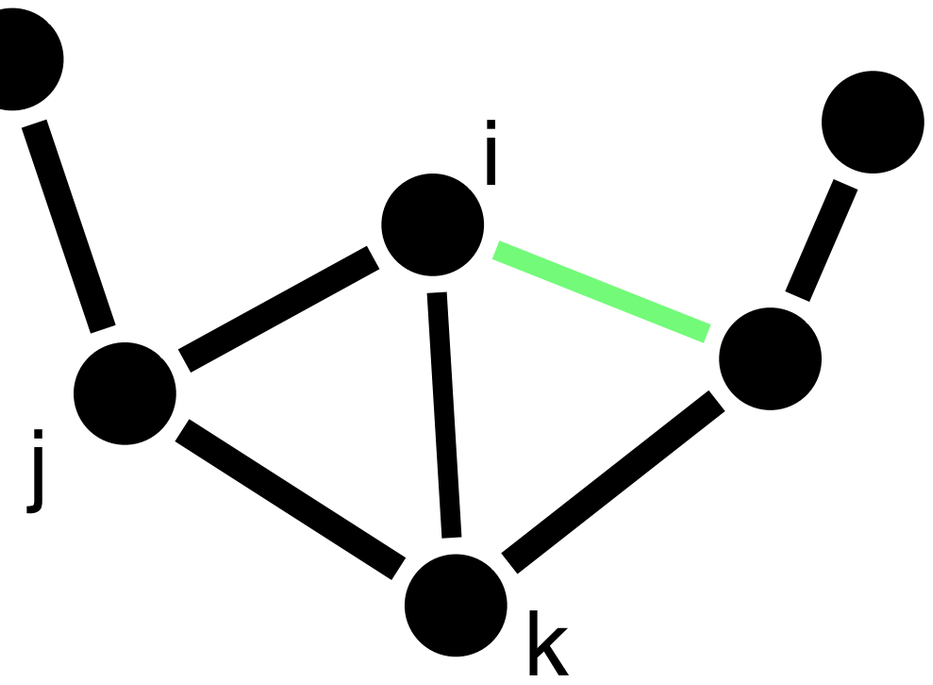
Chum et al, ICCV 2007

Query Expansion



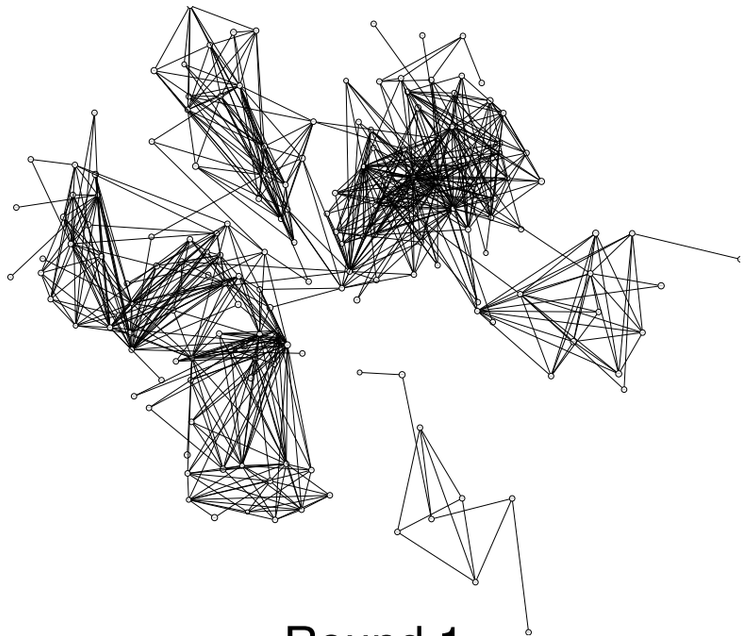
Chum et al, ICCV 2007

Query Expansion

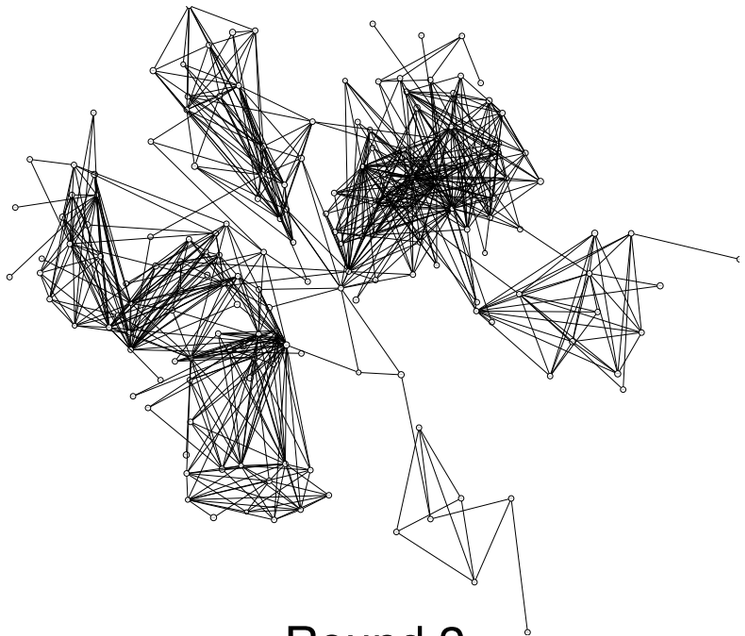


Chum et al, ICCV 2007

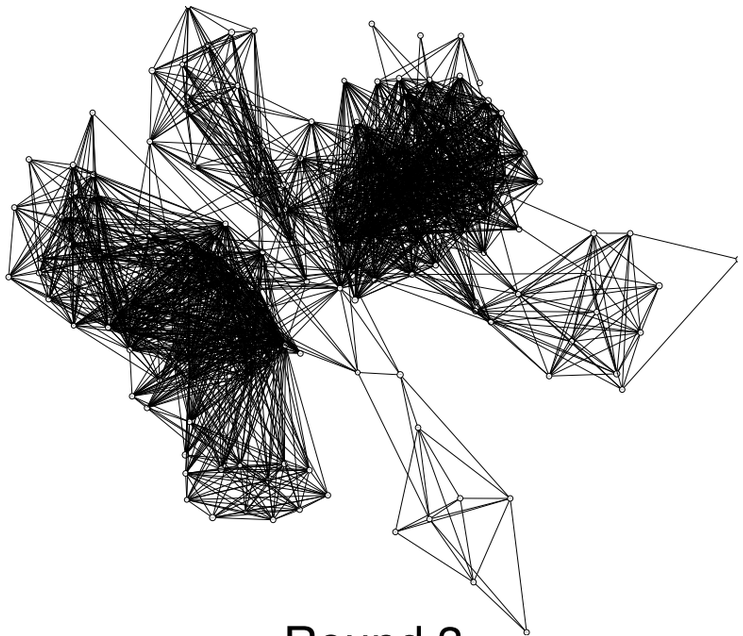
Matching Progress



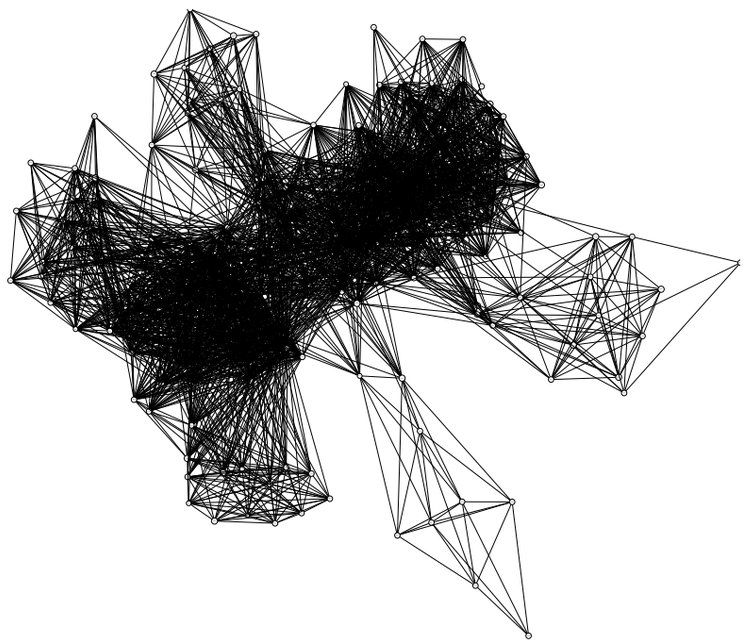
Round 1



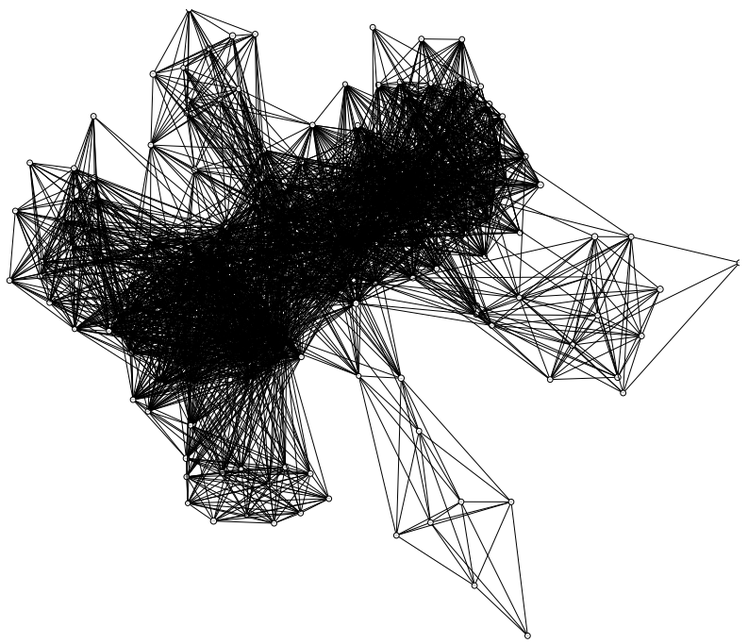
Round 2



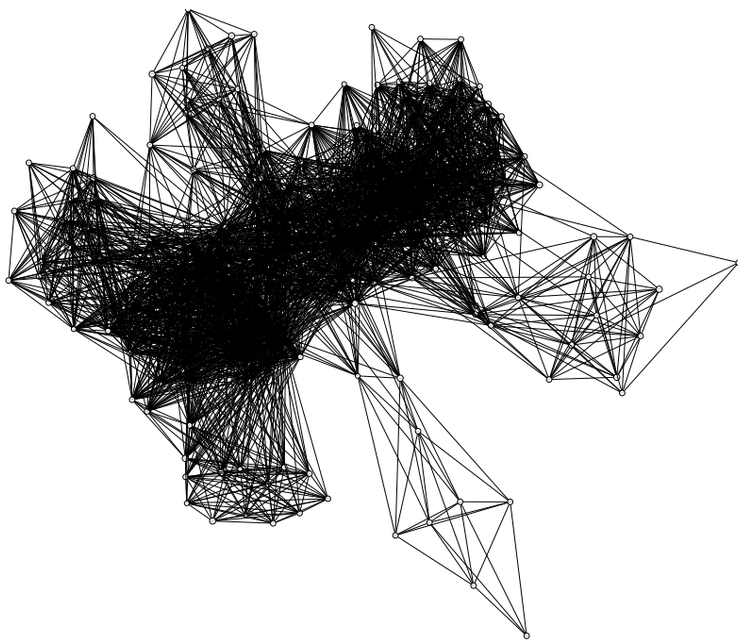
Round 3



Round 4

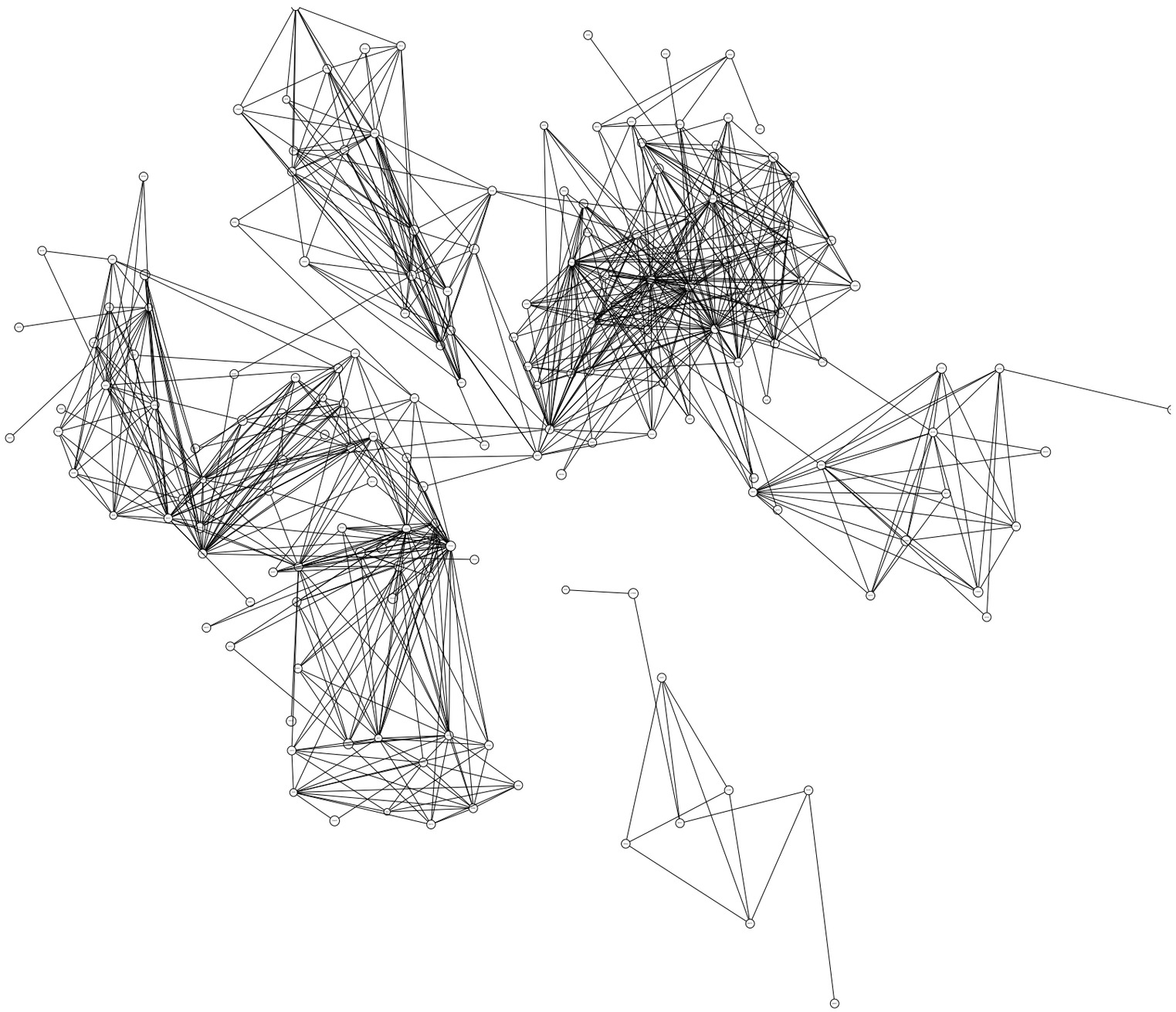


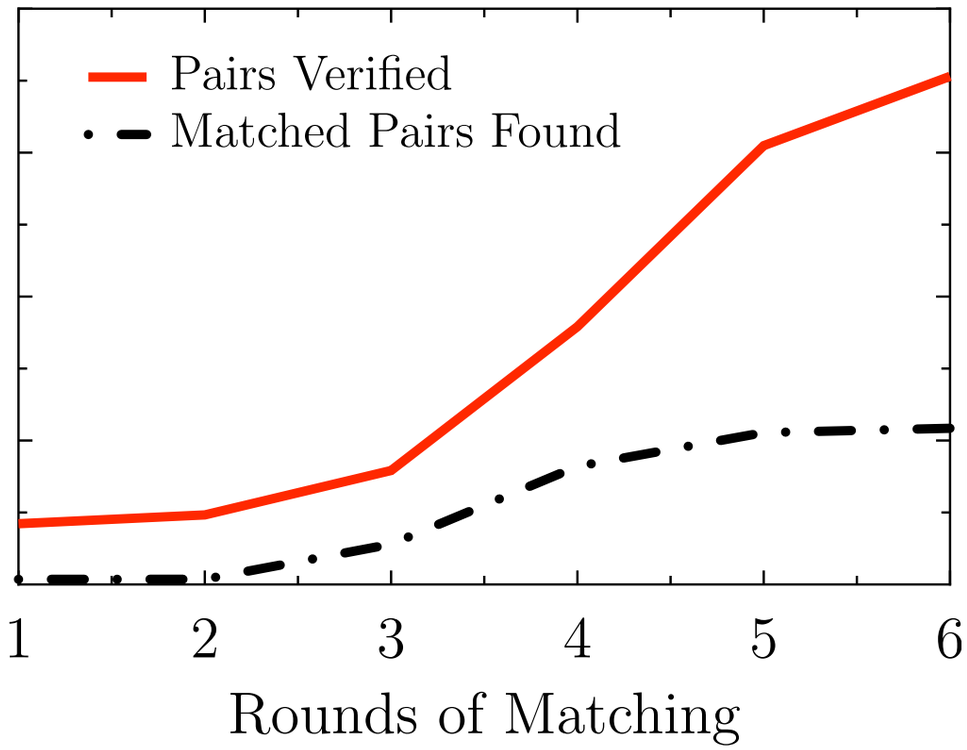
Round 5



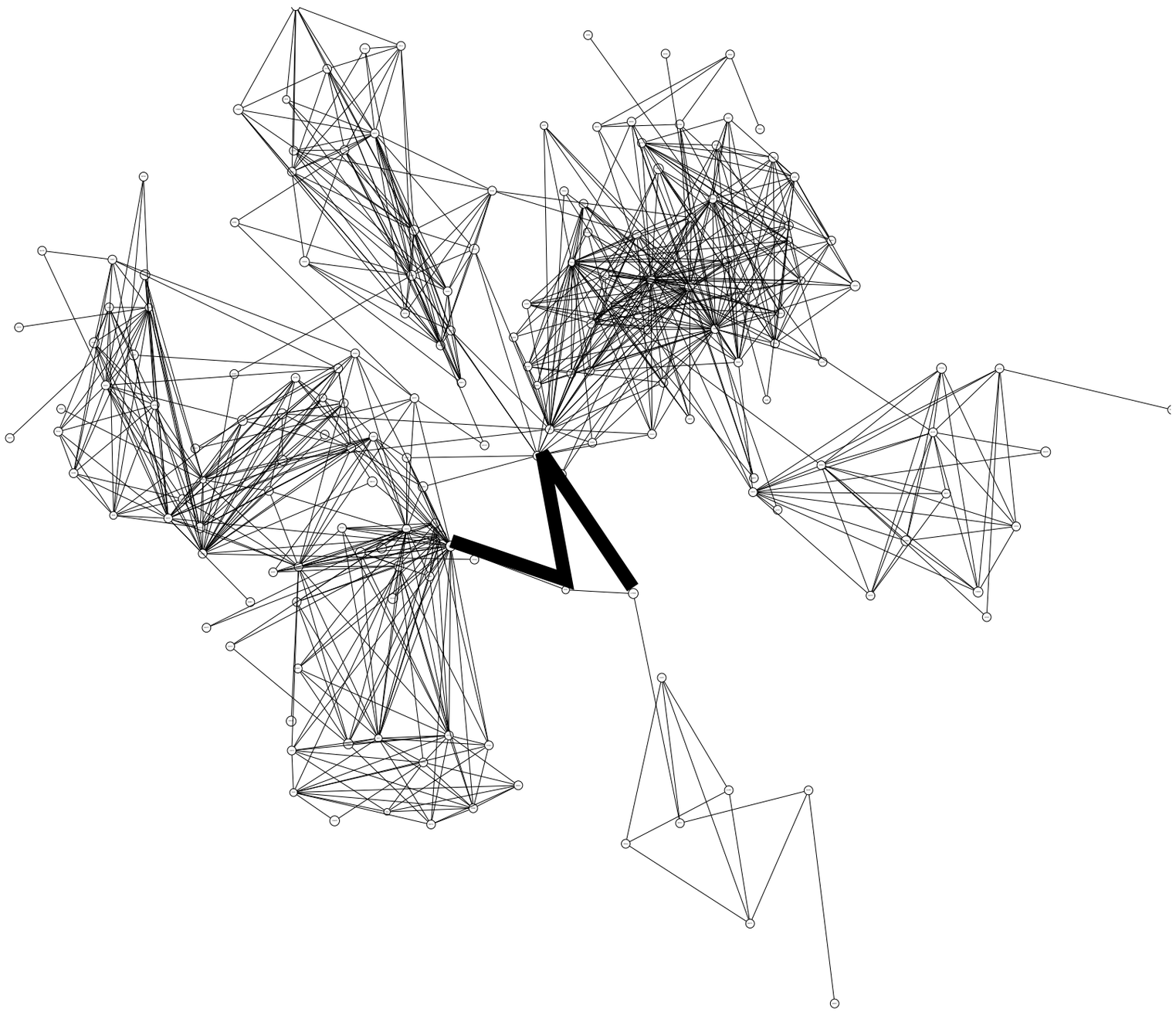
Round 6

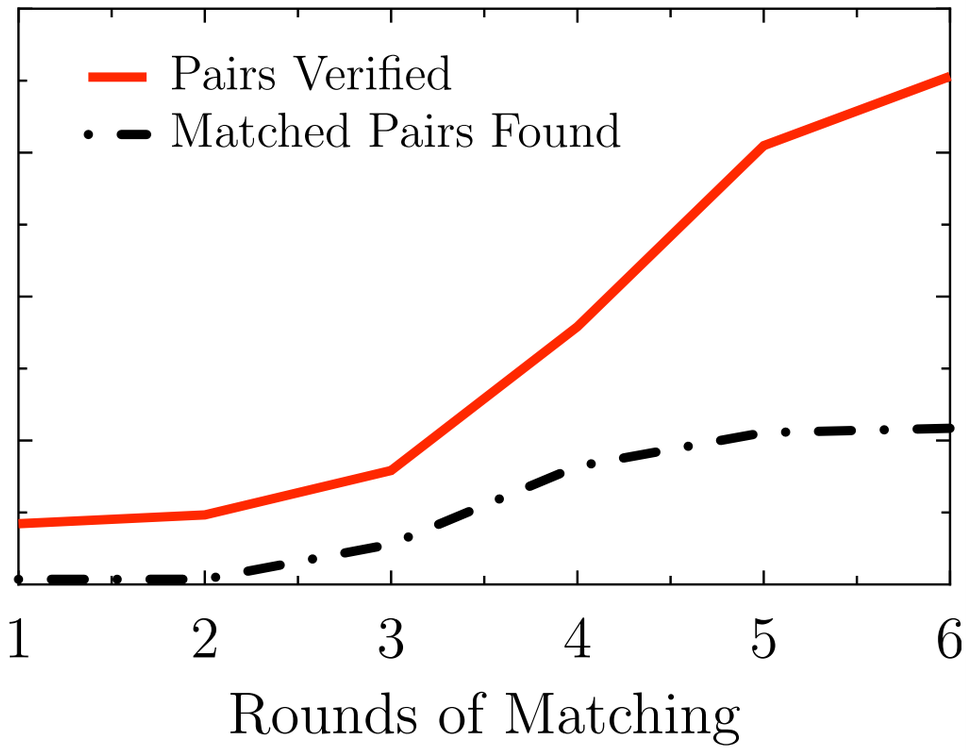
Matching - Round 1



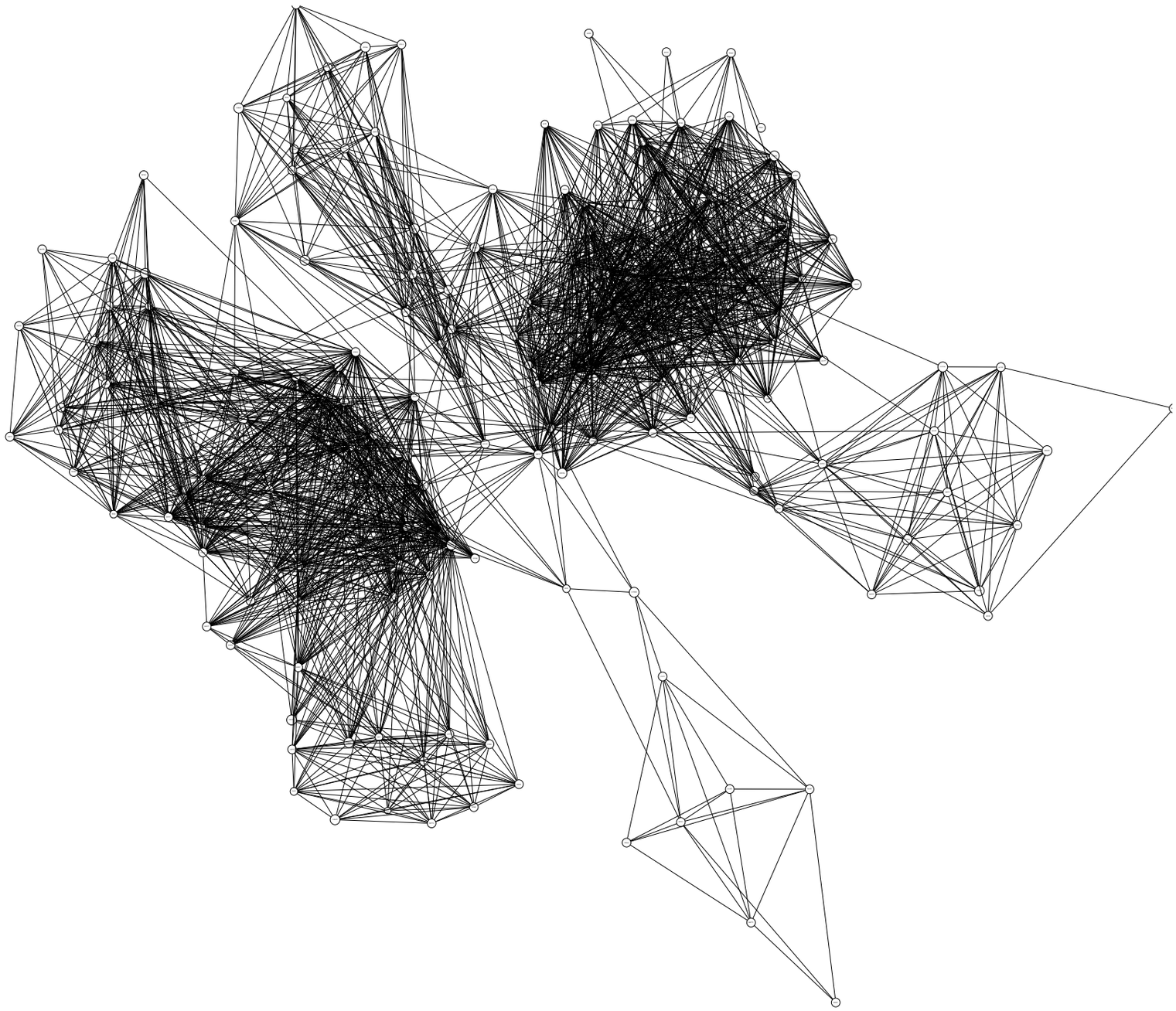


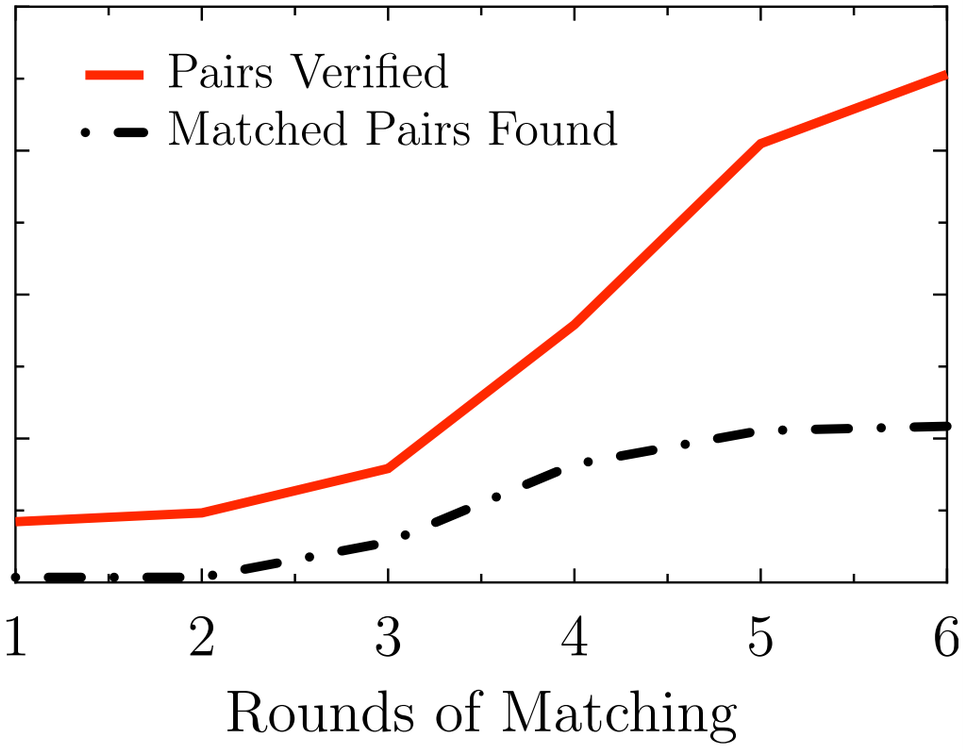
Matching - Round 2



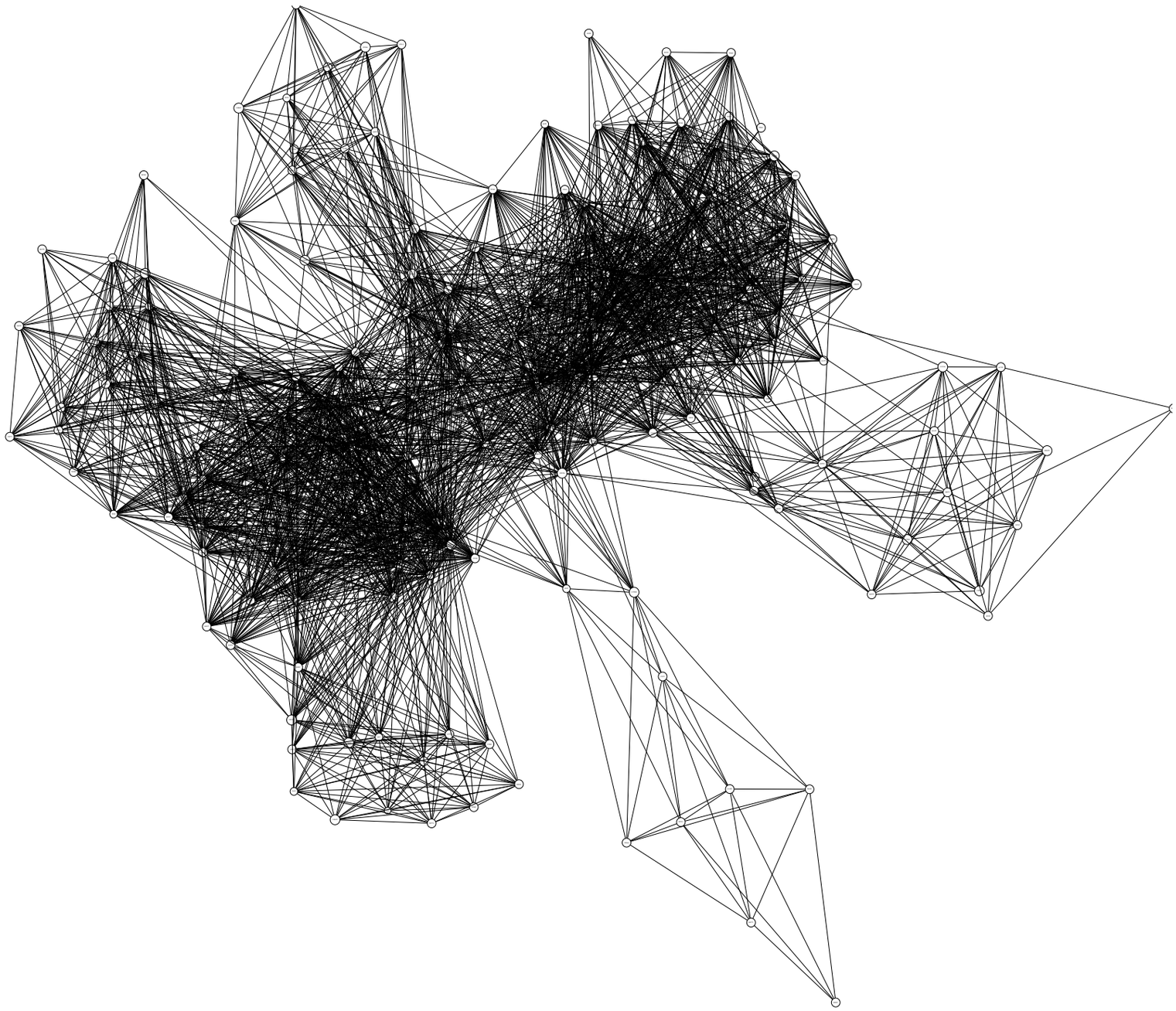


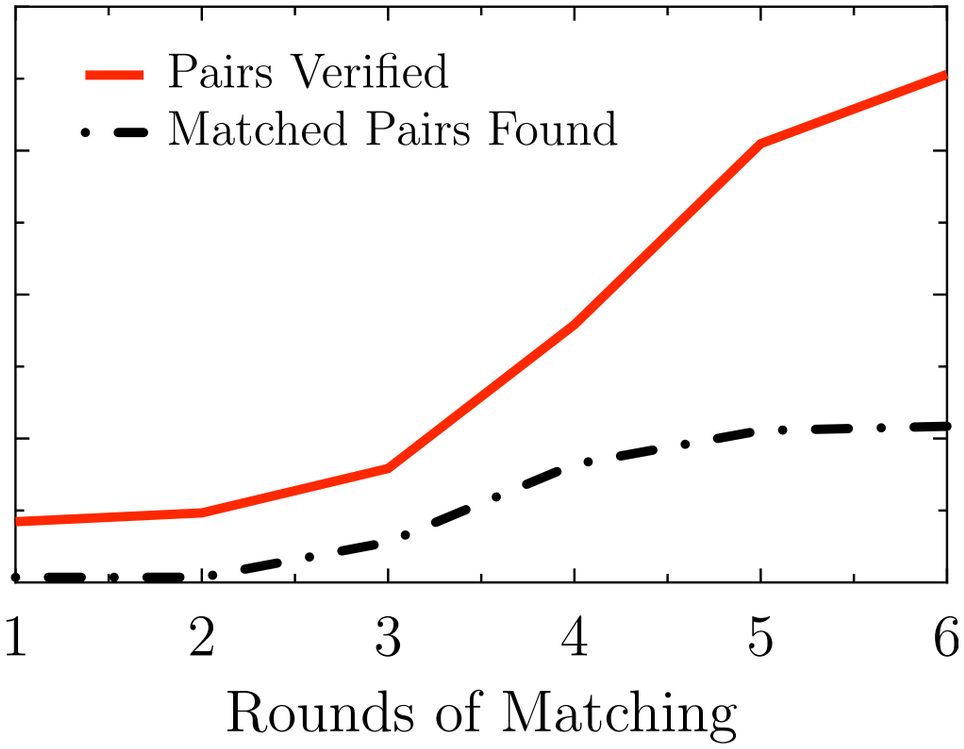
Matching - Round 3



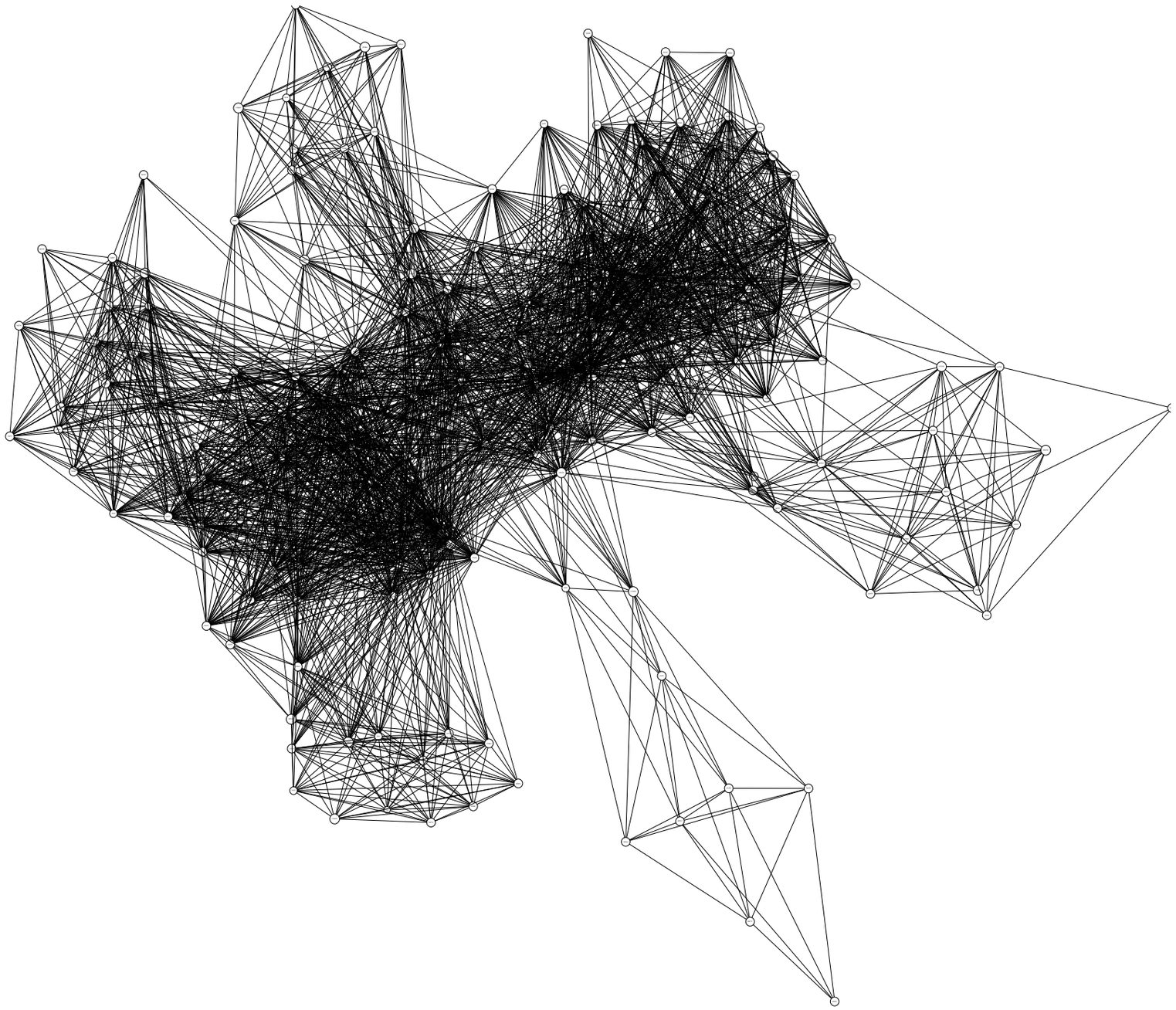


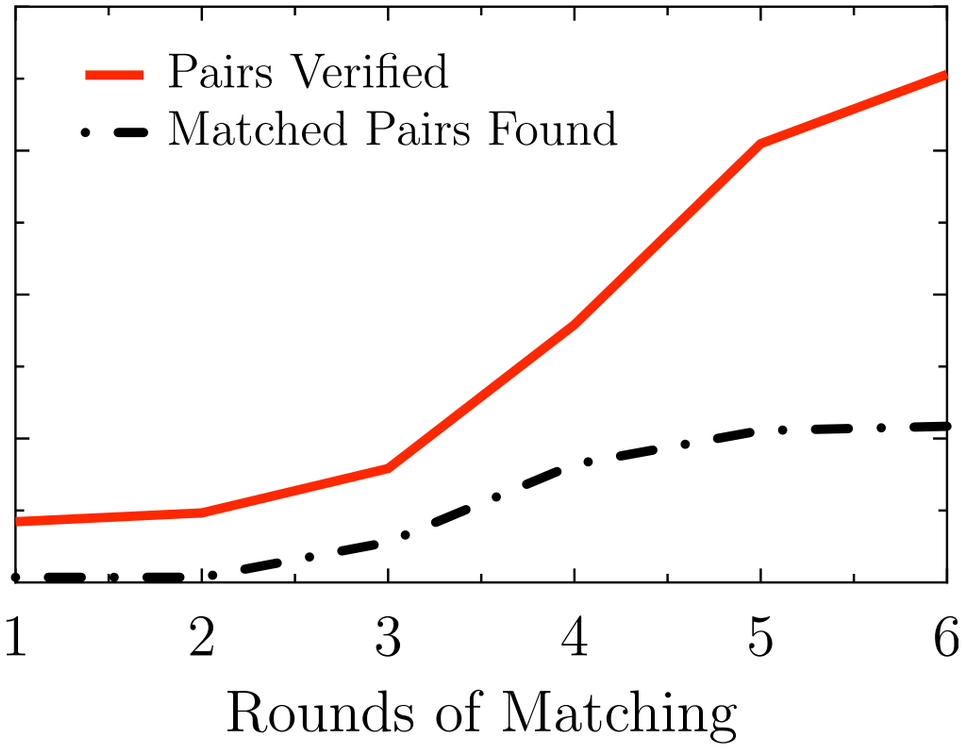
Matching - Round 4



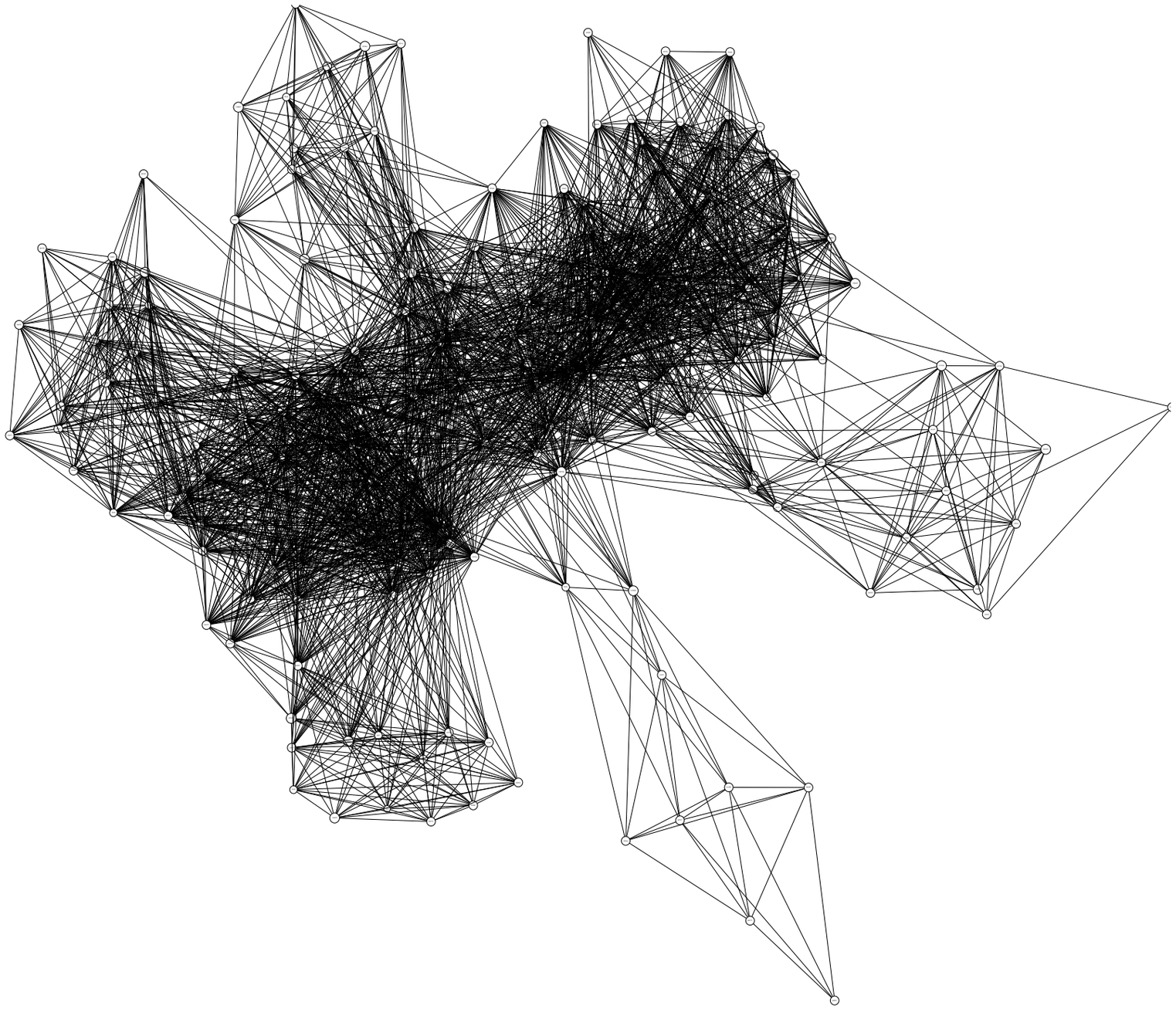


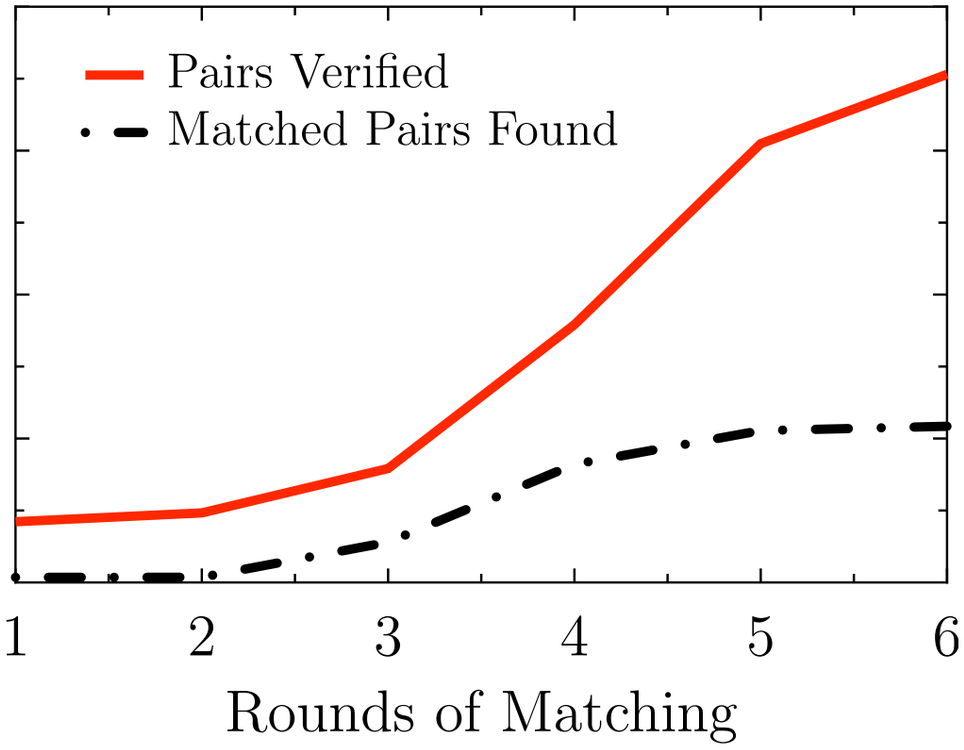
Matching - Round 5





Matching - Round 6





Matching Statistics

Dataset	Size	Matches possible	Matches Tried	Matches Found	Time
Dubrovnik	58K	1.6 Billion	2.6M	0.5M	5 hrs
Rome	150K	11.2 Billion	8.8M	2.7M	13 hrs
Venice	250K	31.2 Billion	35.5M	6.2M	27 hrs

How good are we?

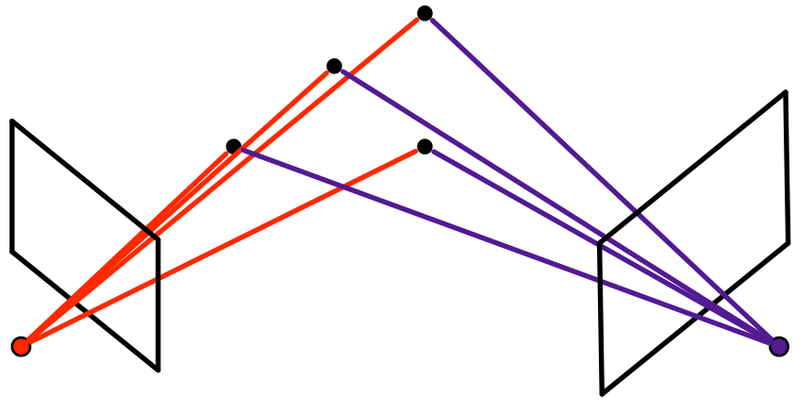
- “Ground Truth” for a 20K data-set.
- 0.25% of total matches performed.
- 90% of the true matches recovered.

Reconstruction

1. Choose two images to
   seed the reconstruction.

Reconstruction

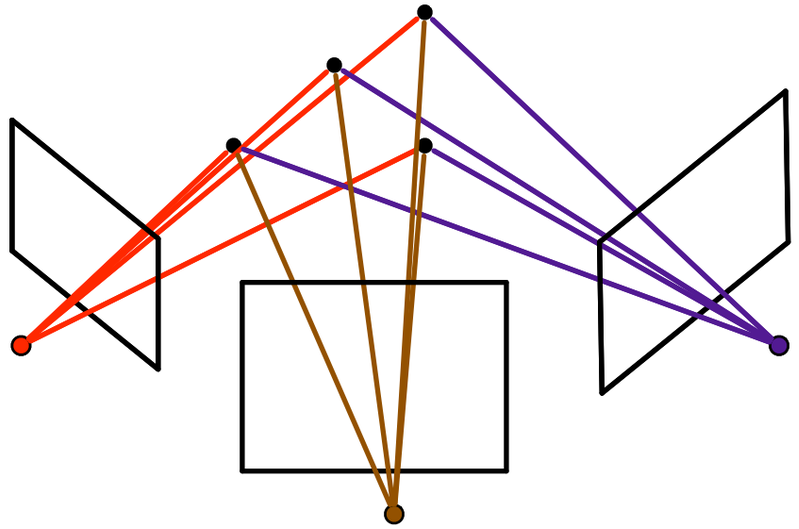
1. Choose two images to seed the reconstruction.



Reconstruction

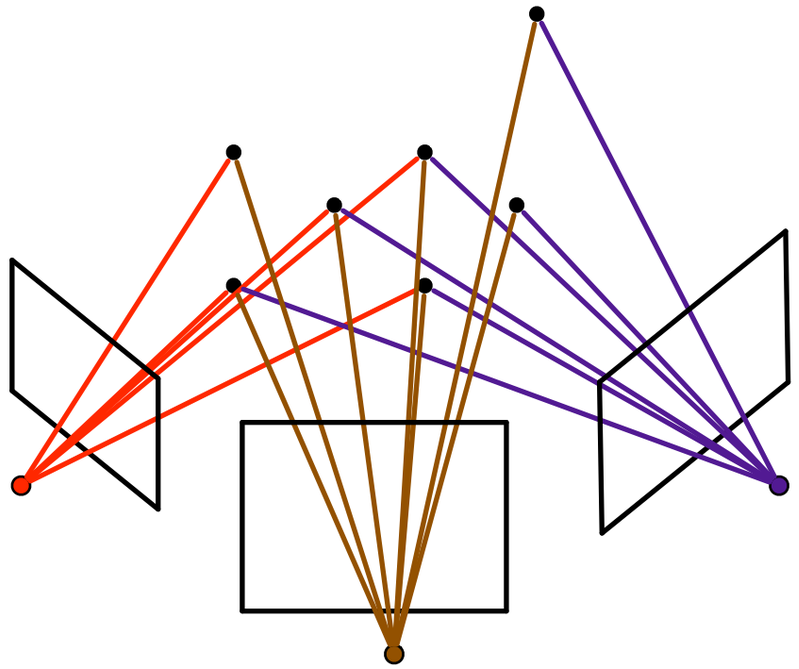
1. Choose two images to seed the reconstruction.

2. Add cameras using pose estimation.



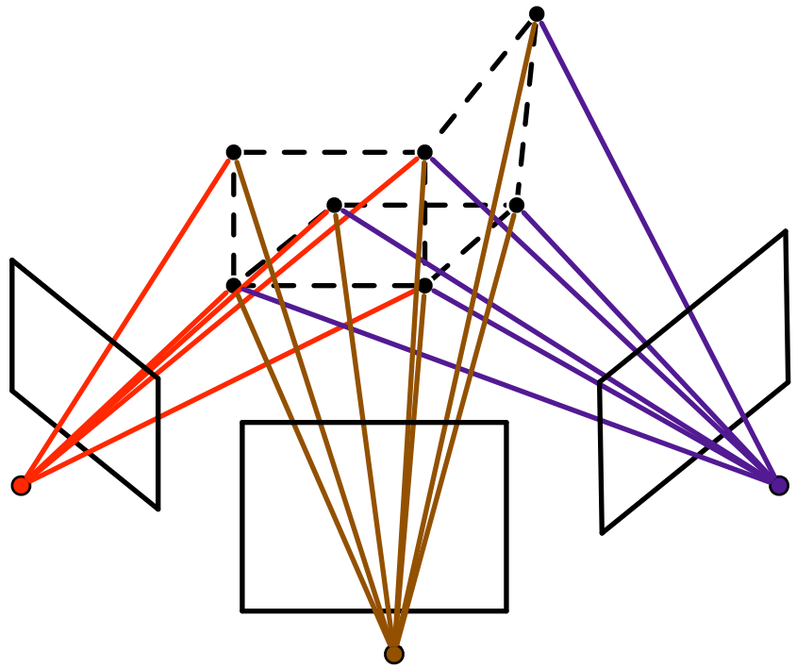
Reconstruction

1. Choose two images to seed the reconstruction.
2. Add cameras using pose estimation.
3. Add 3d points via triangulation.



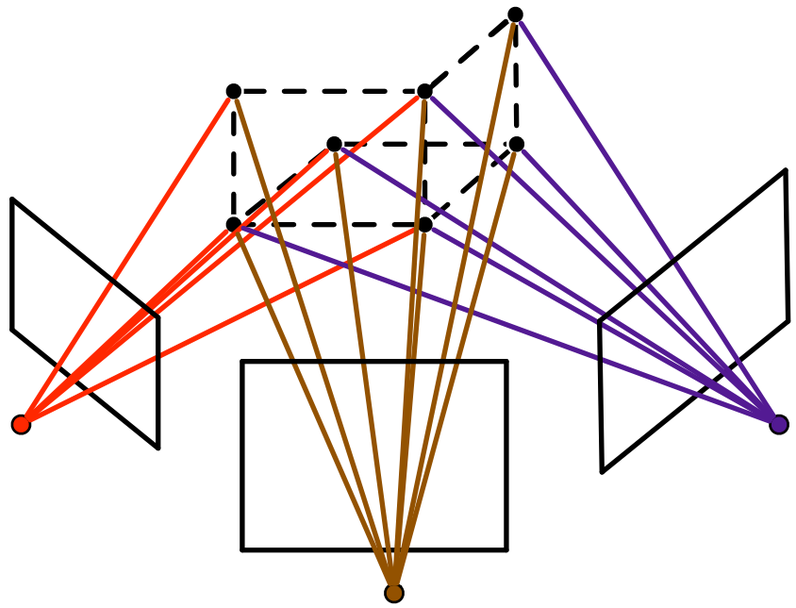
Reconstruction

1. Choose two images to seed the reconstruction.
2. Add cameras using pose estimation.
3. Add 3d points via triangulation.



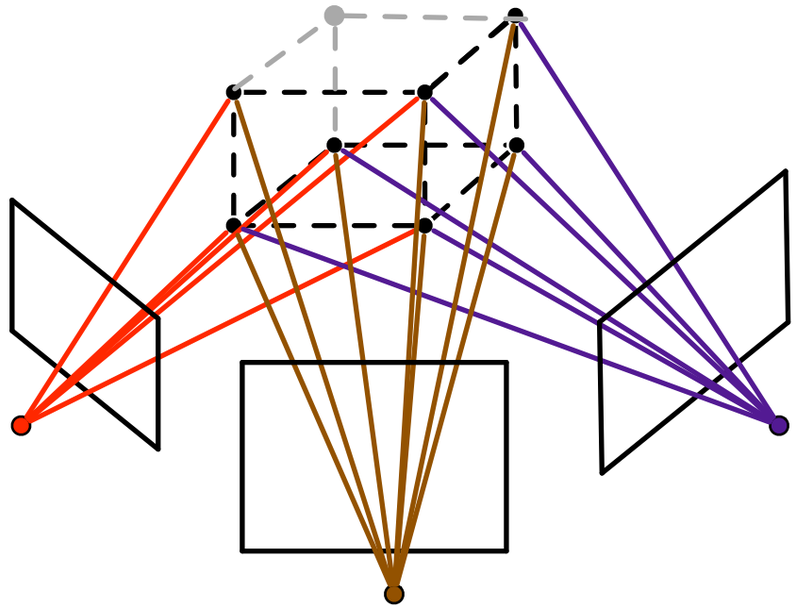
Reconstruction

1. Choose two images to seed the reconstruction.
2. Add cameras using pose estimation.
3. Add 3d points via triangulation.
4. Non-linear refinement.



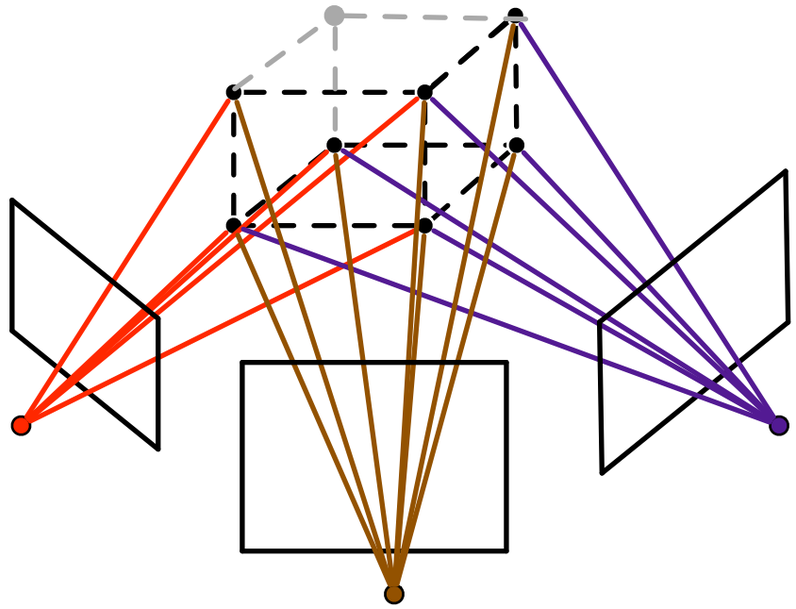
Reconstruction

1. Choose two images to seed the reconstruction.
2. Add cameras using pose estimation.
3. Add 3d points via triangulation.
4. Non-linear refinement.
5. Goto step 2.

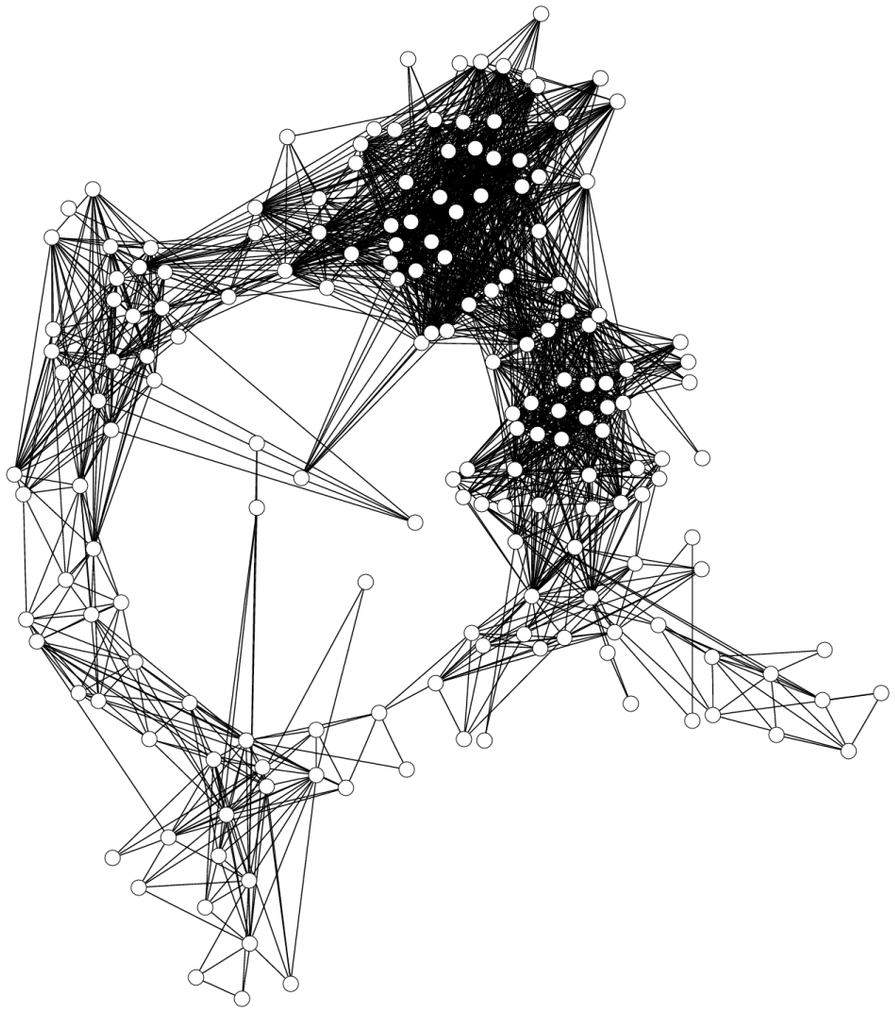


Reconstruction

1. Choose two images to seed the reconstruction.
2. Add cameras using pose estimation.
3. Add 3d points via triangulation.
4. Non-linear refinement.
5. Goto step 2.

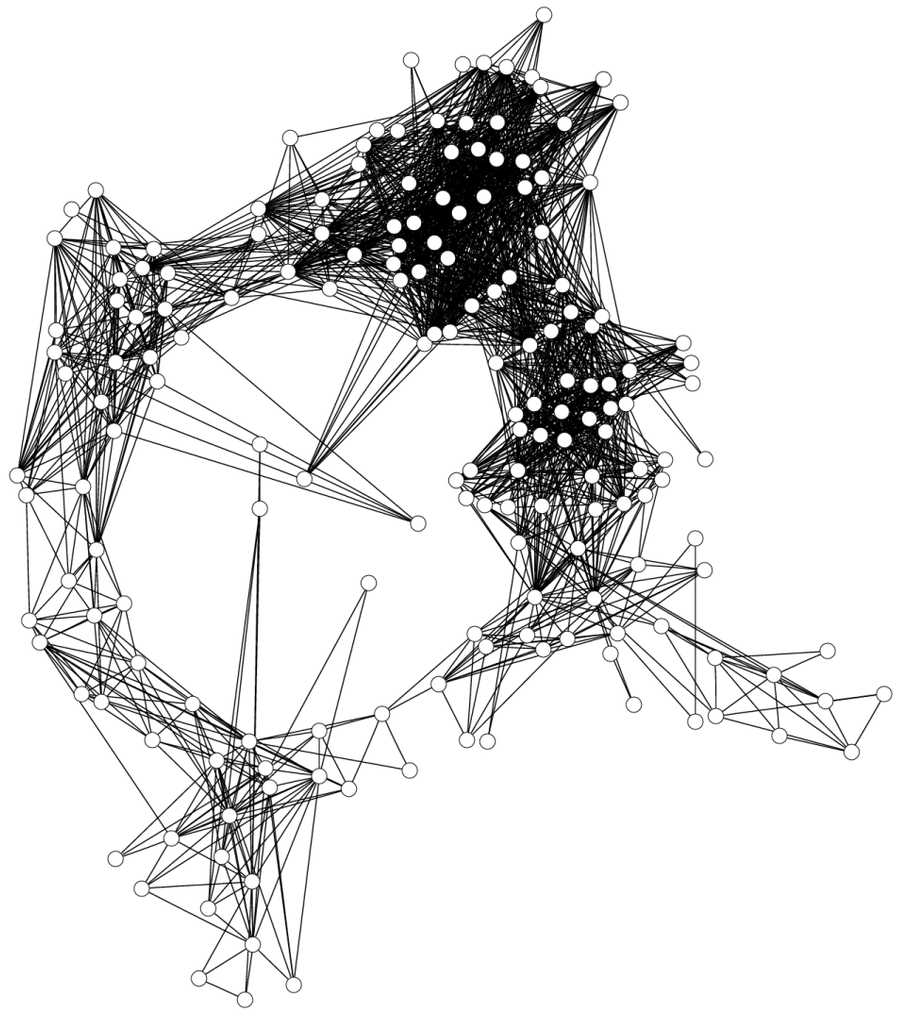


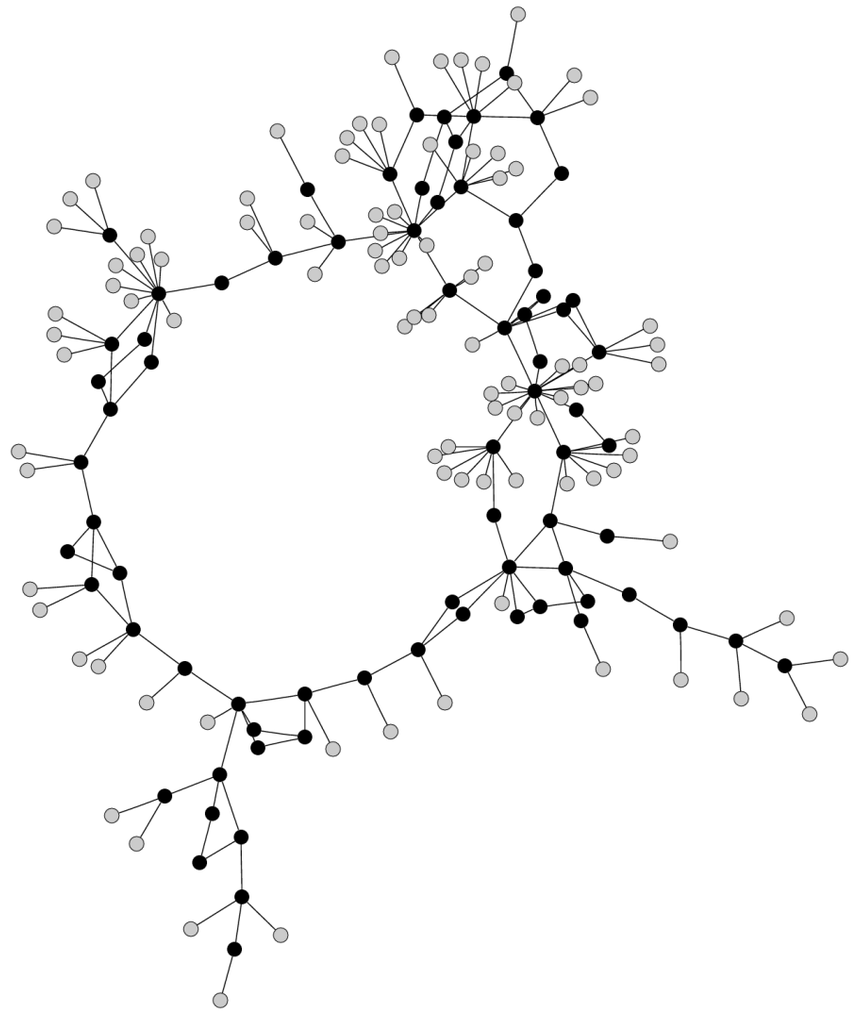
Effective but very slow



Stonehenge

Internet collections represent very non-uniform samplings of viewpoints -- all images are not created equal.





Stonehenge

Skeletal Set

Given an image graph G, select a small set S of important images to reconstruct, bounding the loss in quality of the reconstruction.

Snavely et al. CVPR 07

Large Scale Reconstruction

1. Identify the Skeletal Set
2. Reconstruct the Skeletal Set incrementally
3. Add the remaining images using pose estimation.
4. Triangulate new points.
5. Non-linear refinement (Bundle Adjustment)

Results

Statistics

Datasets	Size	Cores	Largest component	Registered	Matching pairs	Match Time	Reconstruction Time
Dubrovnik	57,845	352	4,619	11,868	498,982	5 hrs	8 hrs
Rome	150,000	496	2,106	36,658	2,712,301	13 hrs	2 hrs
Venice	250,000	496	14,079	47,925	6,119,207	27 hrs	49 hrs

Rome: Colosseum

Rome: St. Peters Cathedral

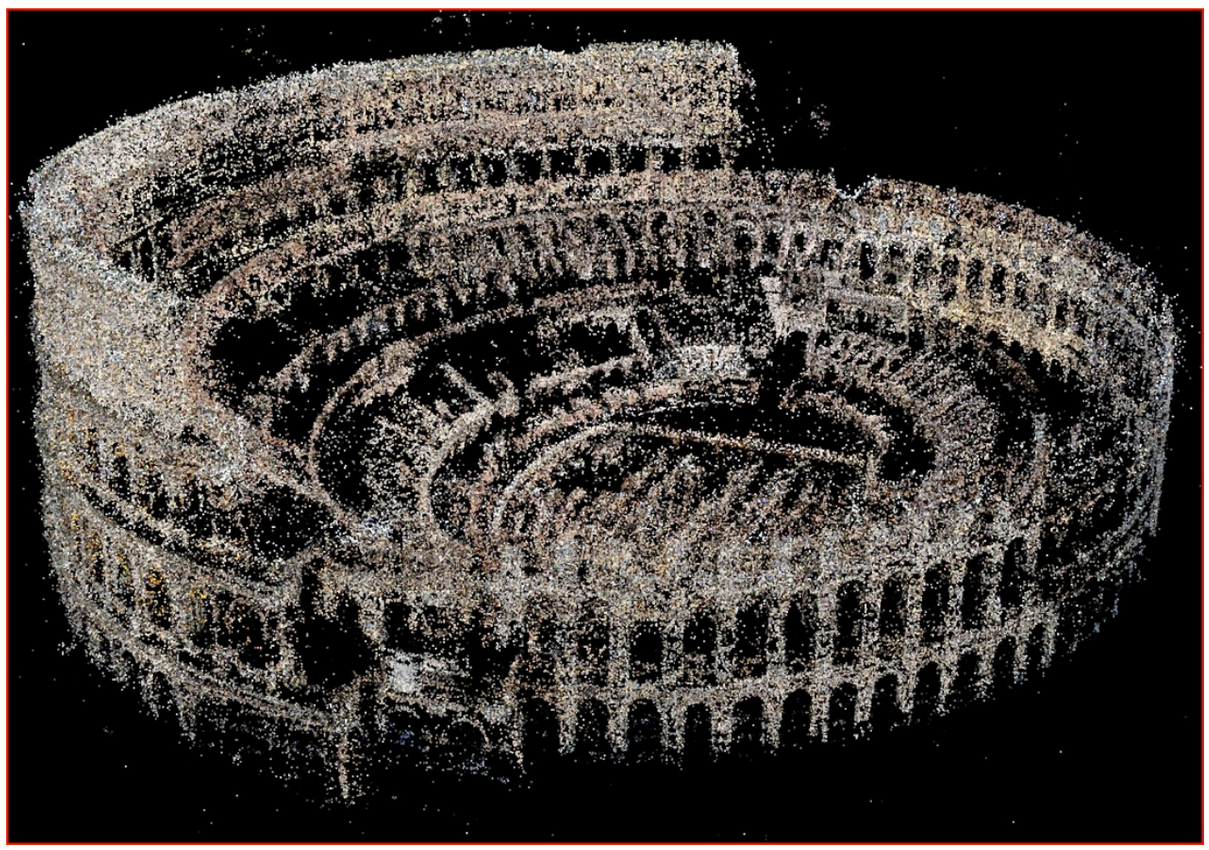
Venice: The Canal

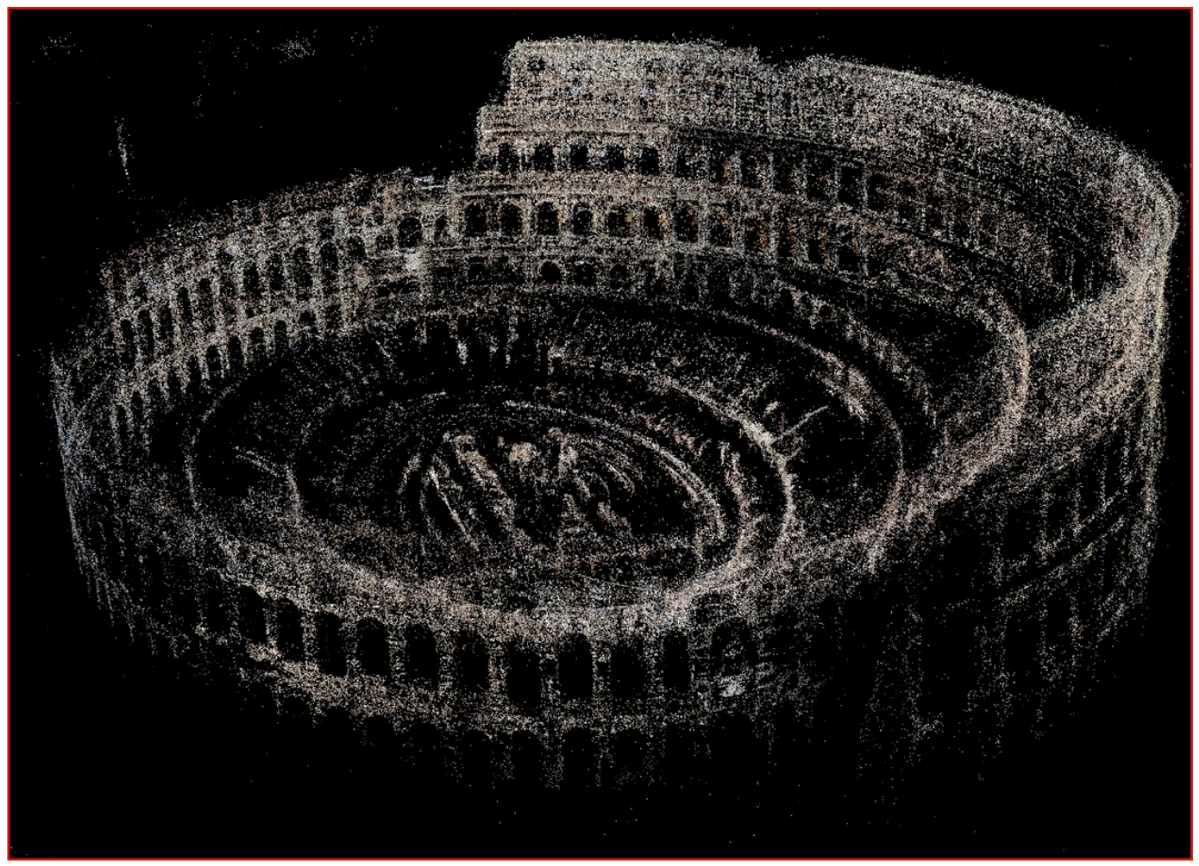
Venice: San Marco

Dubrovnik

The road ahead...

A live, fully textured, ever growing digital model of the world.

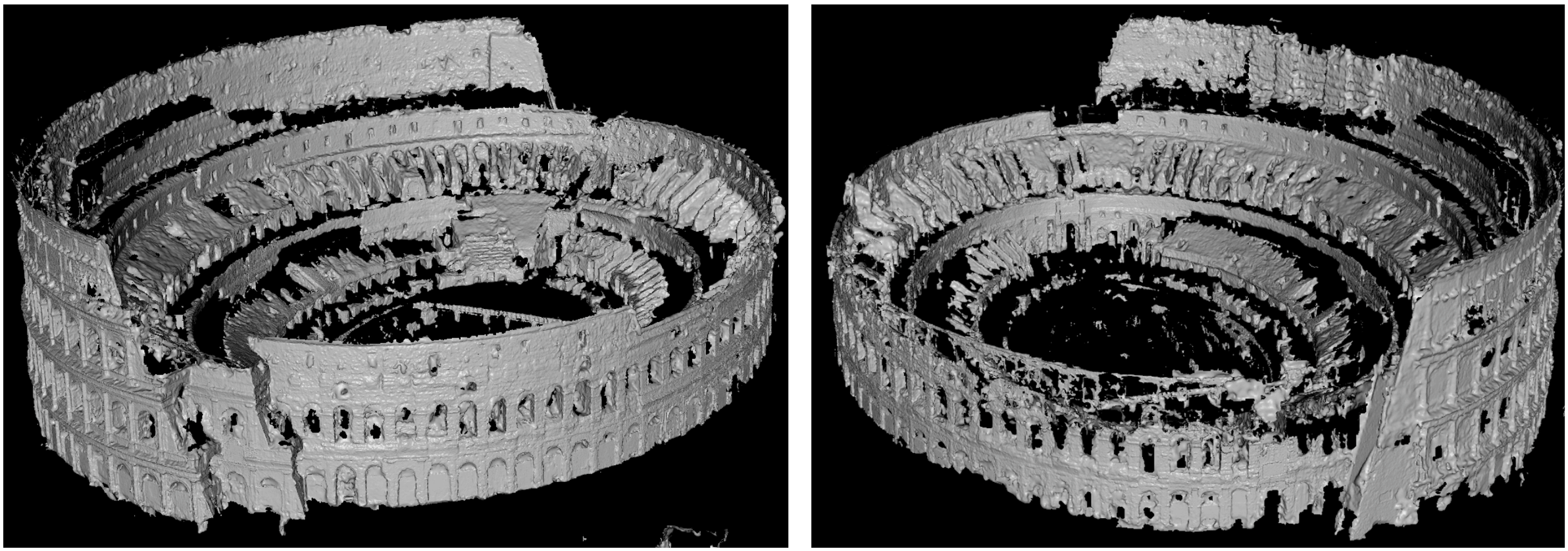




Sparse output from the SfM system

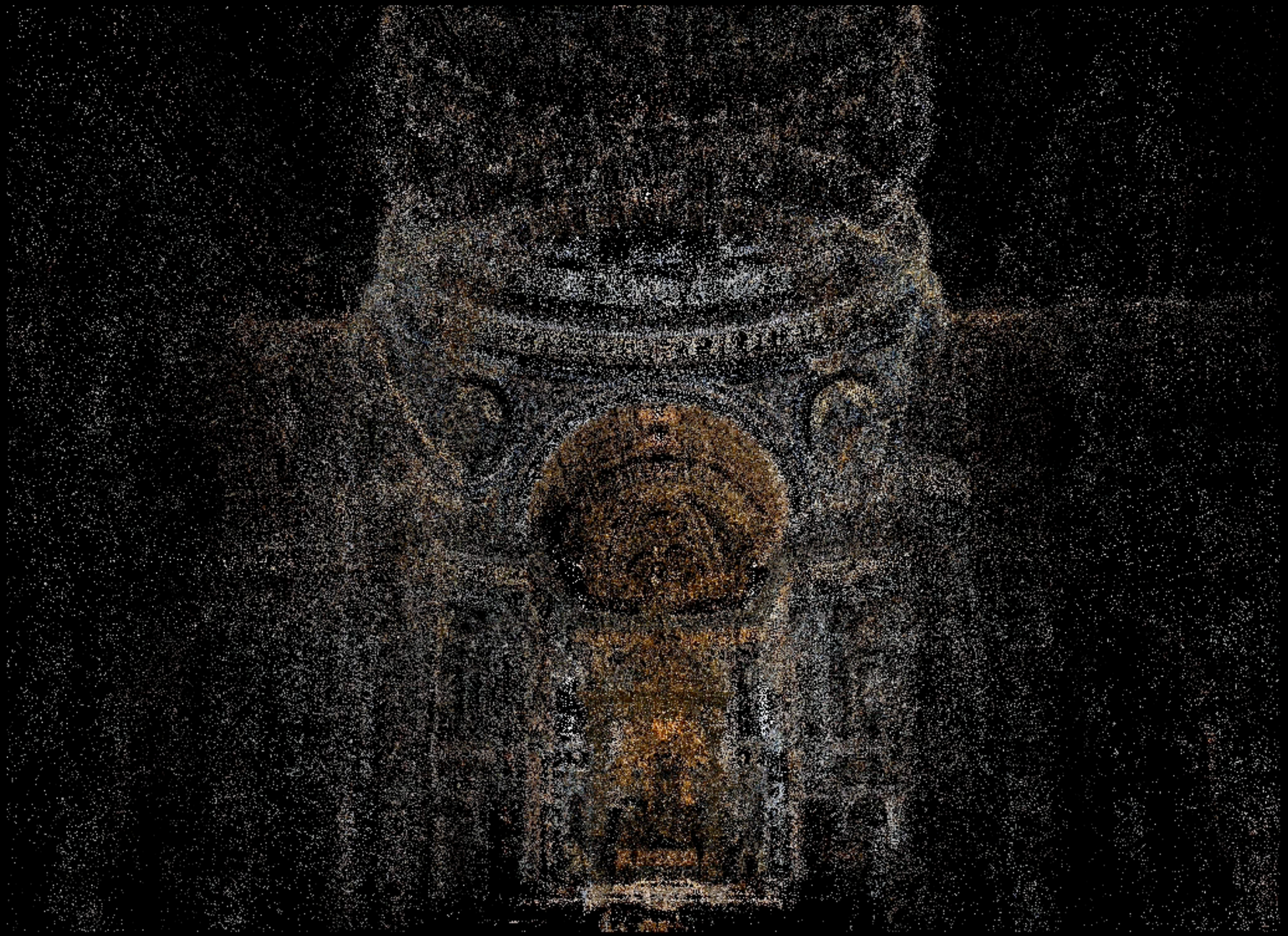
Dense patches from multi-view stereo

Courtesy Yasutaka Furukawa



Mesh output from the multi-view stereo

Courtesy Yasutaka Furukawa



Furukawa et al., CVPR 2010

References

1. **Rebuilding Rome in a Day** - S. Agarwal, N. Snavely, I. Simon, R. Szeliski and S. Seitz, *ICCV 09*.

2. **Accurate, Dense and Robust Multiview Stereopsis**, Y. Furukawa and J. Ponce, *CVPR 07*.

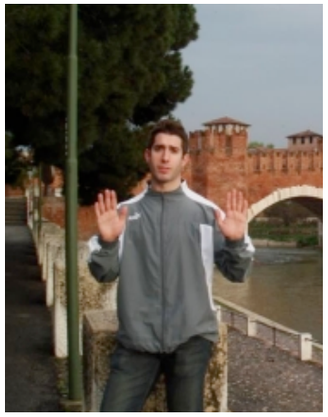
3. **Skeletal Graphs for Efficient Structure from Motion** - N. Snavely, S. Seitz and R. Szeliski, *CVPR 07*.

4. **Photo Tourism: Exploring Photo Collections in 3D** - N. Snavely, S. Seitz and R. Szeliski, *SIGGRAPH 06*.

Collaborators

Noah Snavely
Cornell University

Yasutaka Furukawa



Ian Simon

University of Washington

Steve Seitz

Rick Szeliski

Microsoft Research

Thank you