CS 3630
Spring 2022

Lecture 28:
Structure from Motion and Visual SLAM

Outline

- Structure from Motion
- Correspondence
- Optimization
- Visual SLAM
- 4D Reconstruction
- Dense Surface Reconstruction

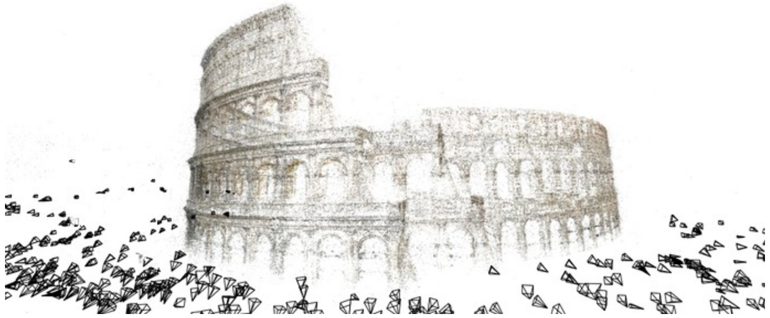
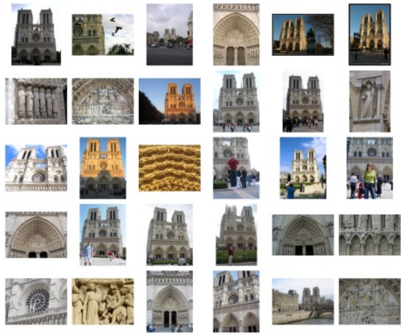


Photo Tourism

Noah Snavely, Steven M. Seitz, Richard Szeliski, Photo tourism: Exploring photo collections in 3D," ACM Transactions on Graphics (SIGGRAPH Proceedings), 25(3), 2006, 835-846.



Input photographs

Scene reconstruction

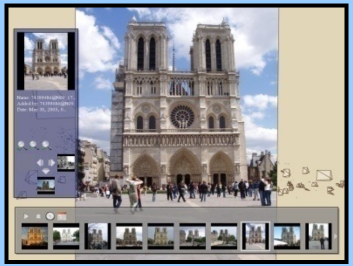
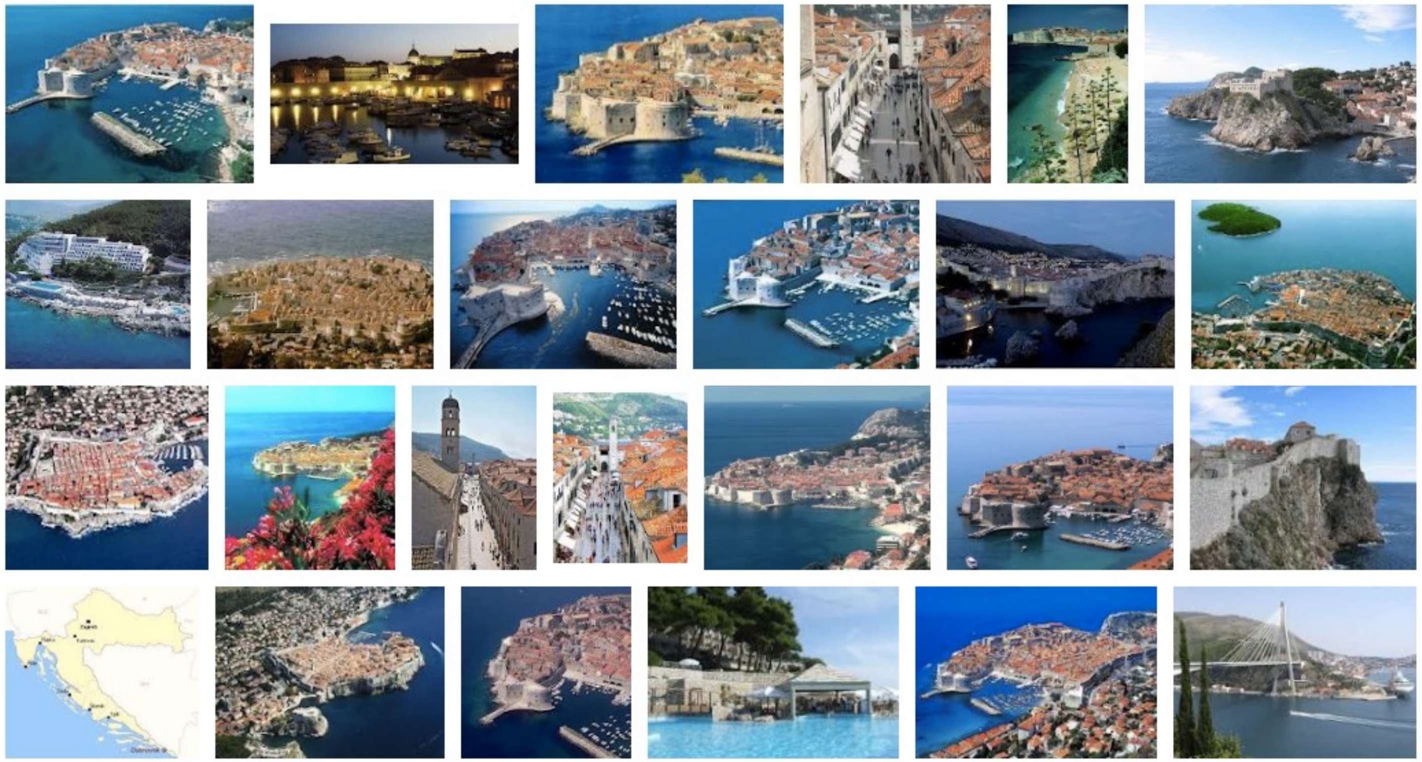


Photo Explorer

http://phototour.cs.washington.edu/

3D Models from Community Databases

E.g., Google image search on "Dubrovnik"



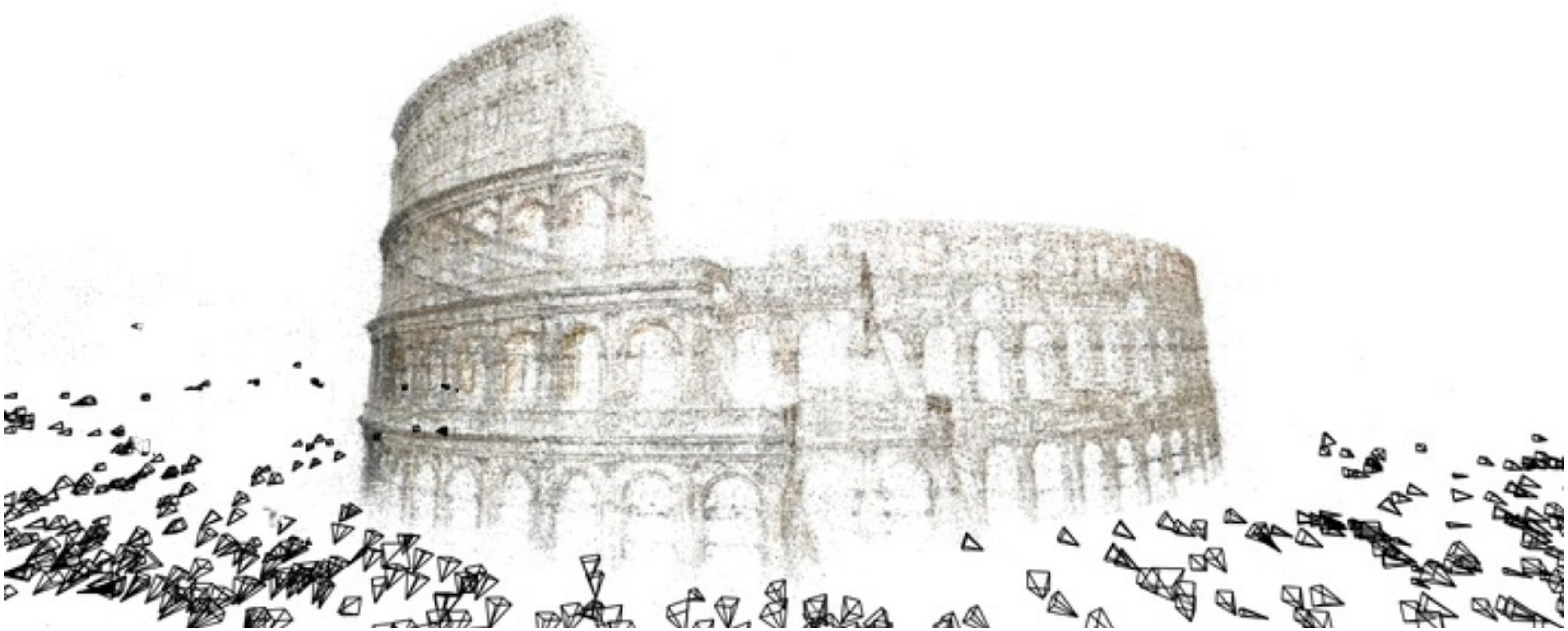
3D Models from Community Databases

5K images, 3.5M points, >10M factors

Movie by Aggarwal et al.

Building Rome in a Day

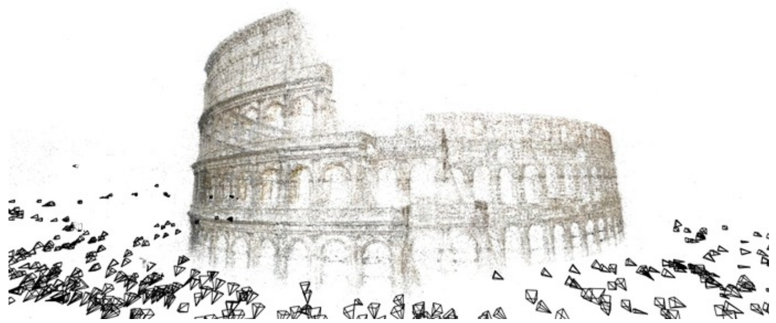
[Building Rome in a Day Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski International Conference on Computer Vision, 2009, Kyoto, Japan.]



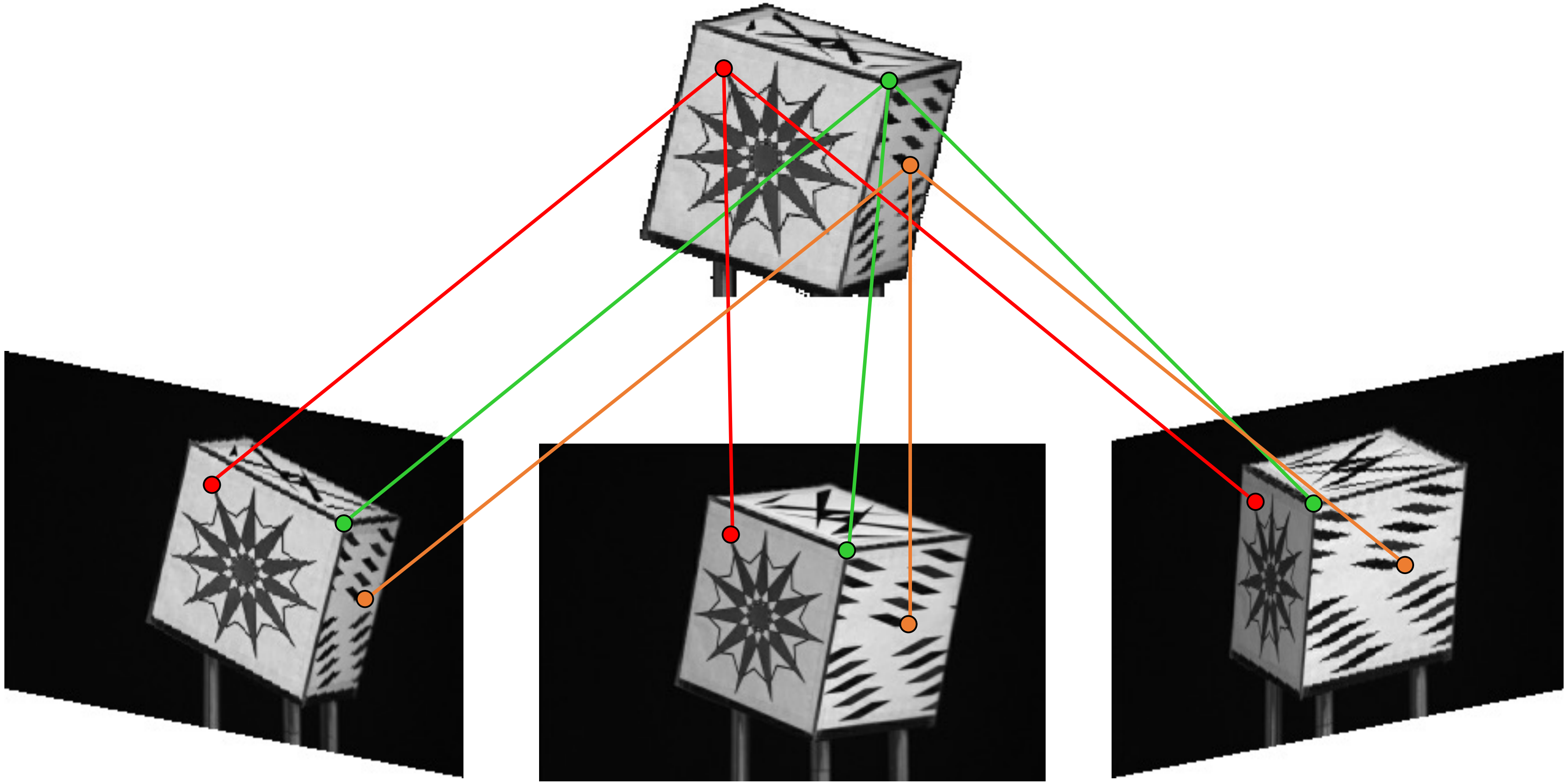
http://grail.cs.washington.edu/rome/

Outline

- Structure from Motion
- Correspondence
- Optimization
- Visual SLAM
- 4D Reconstruction
- Dense Surface Reconstruction

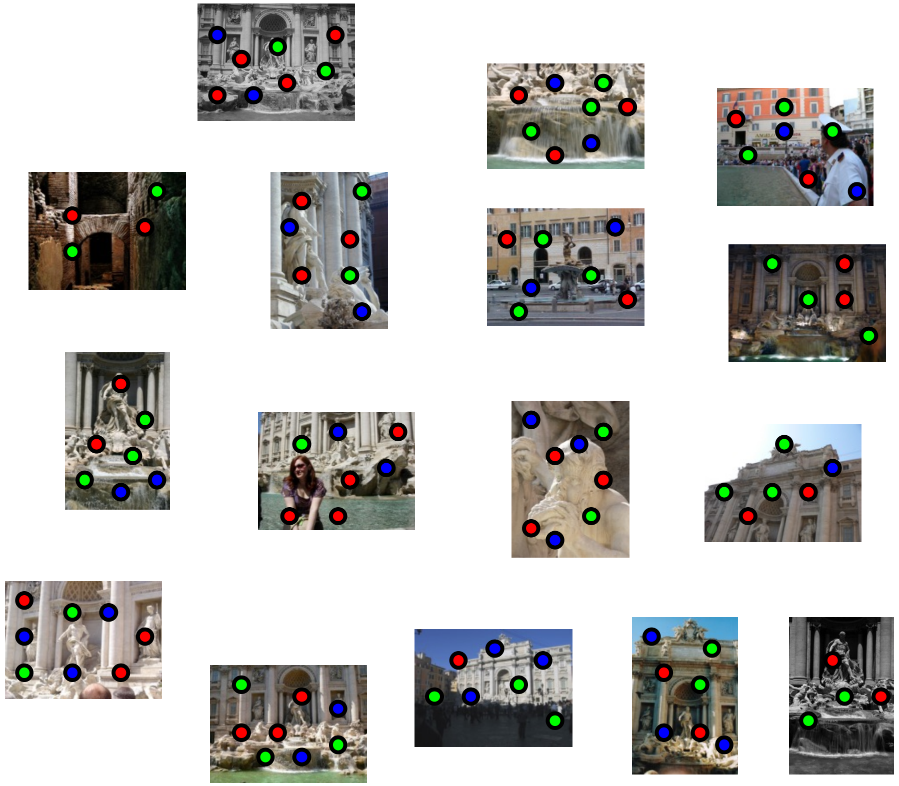


A Correspondence Problem



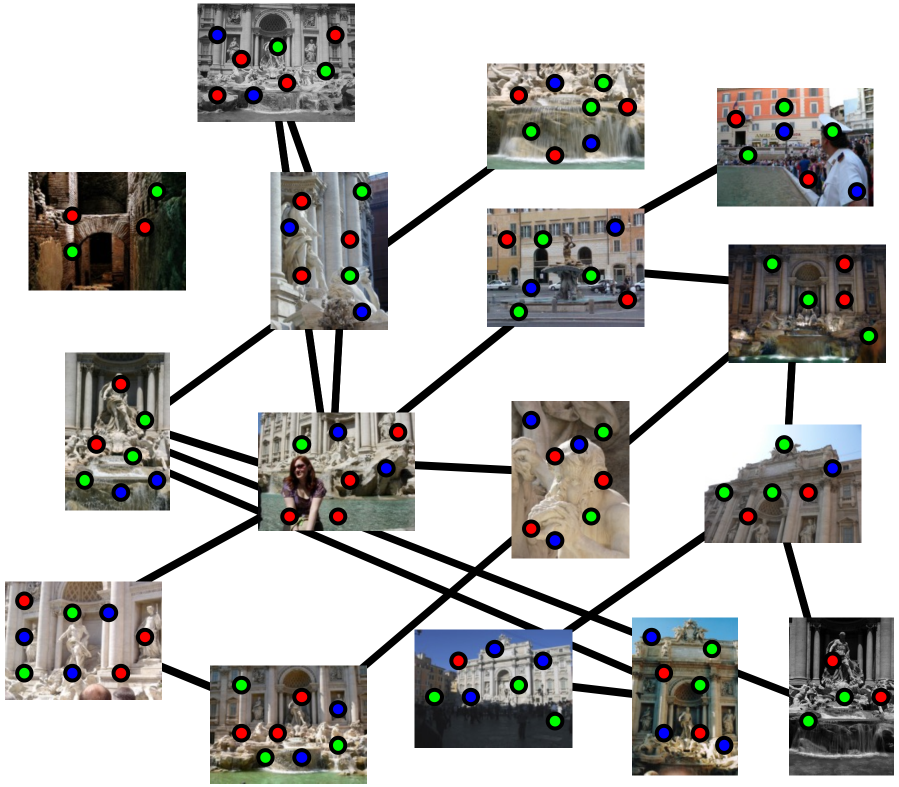
Feature detection

- Detect features using SIFT [Lowe, IJCV 2004]

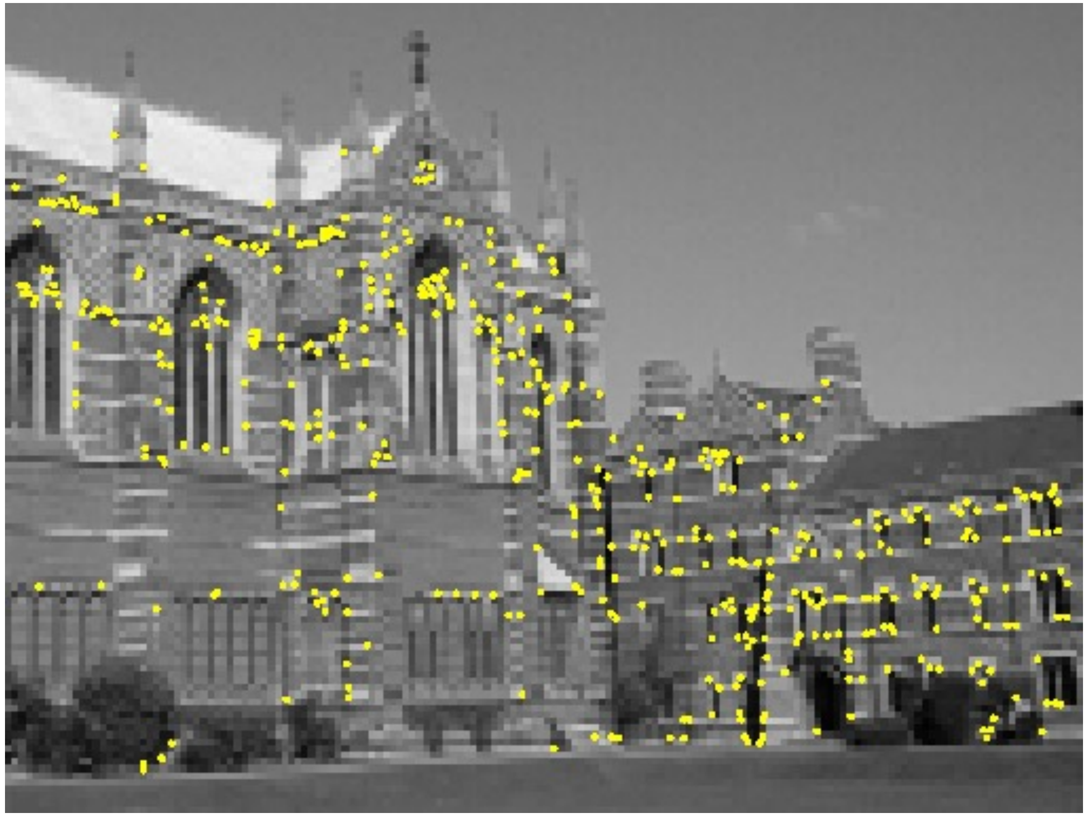


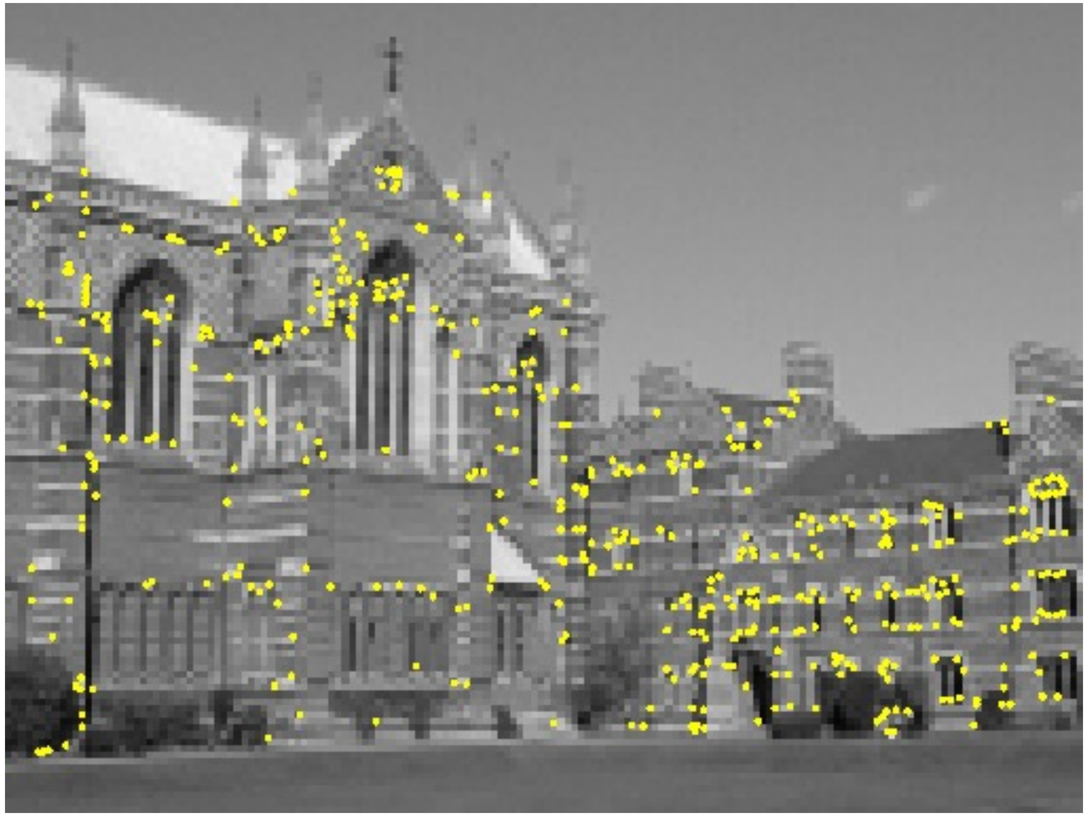
Feature matching

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



Feature Matching !





Real World Challenges

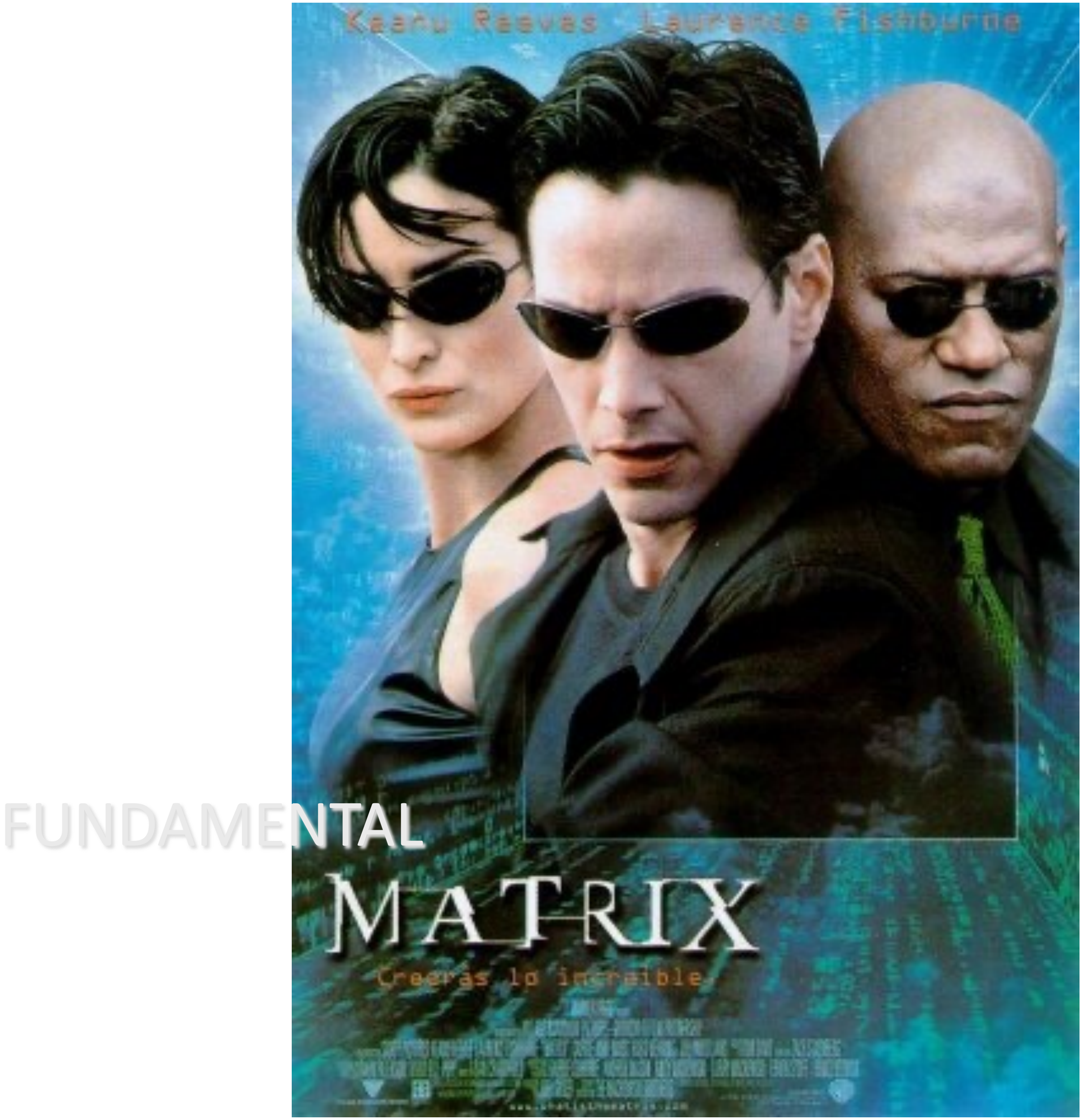
Bad News: Good correspondences are hard to find

- Good news: Geometry constrains possible correspondences.
  - 4 DOF between x and x'; only 3 DOF in X.
  - Constraint is manifest in the **Fundamental matrix**

$$x'^T F x = 0.$$

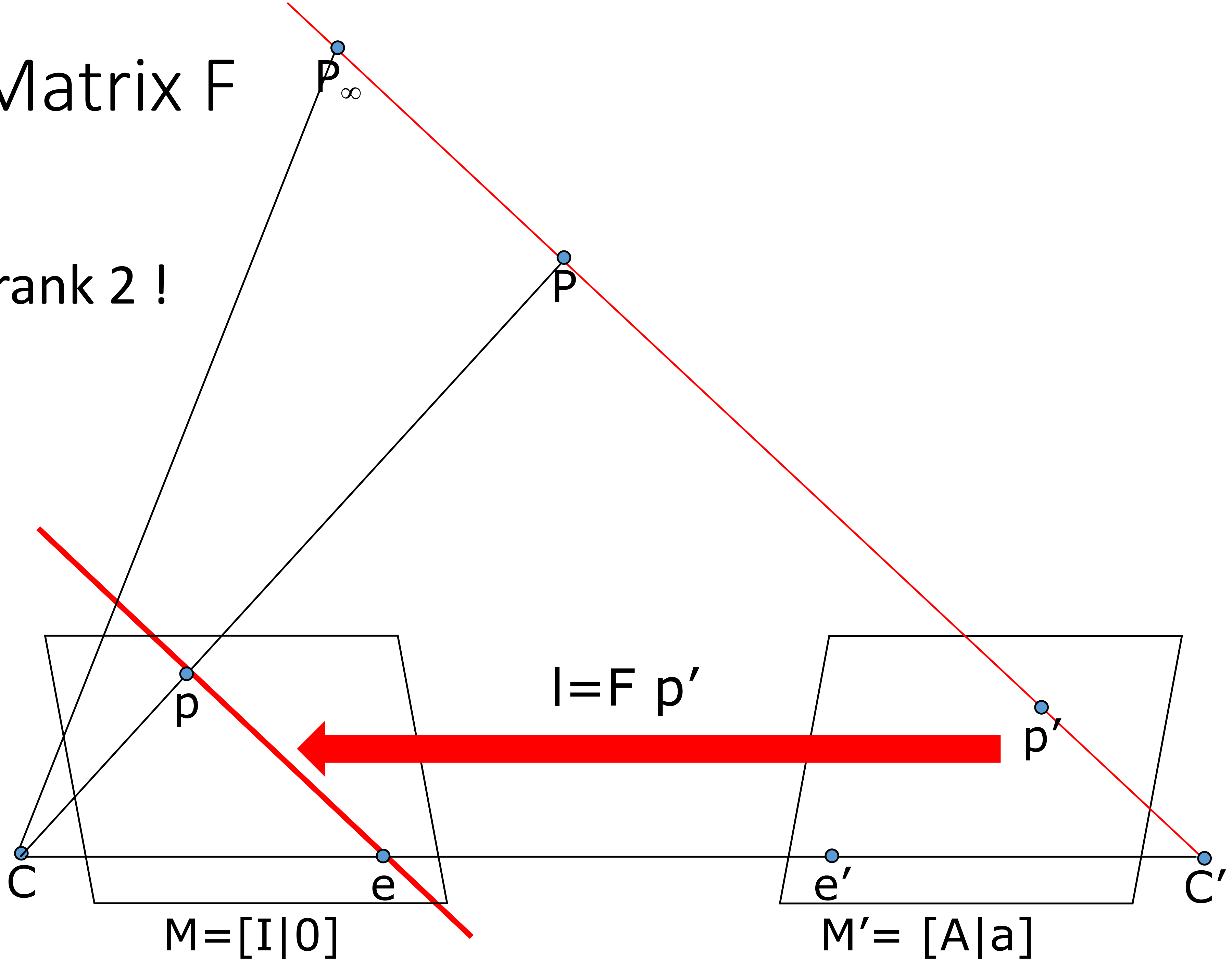
  - F can be calculated either from camera matrices or a set of good correspondences.

Fundamental Matrix



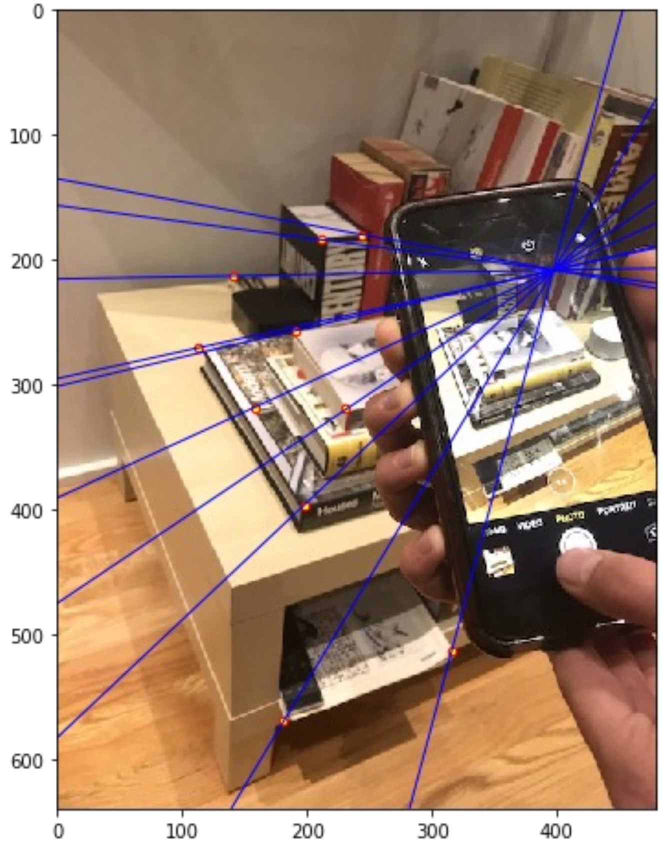
Fundamental Matrix F

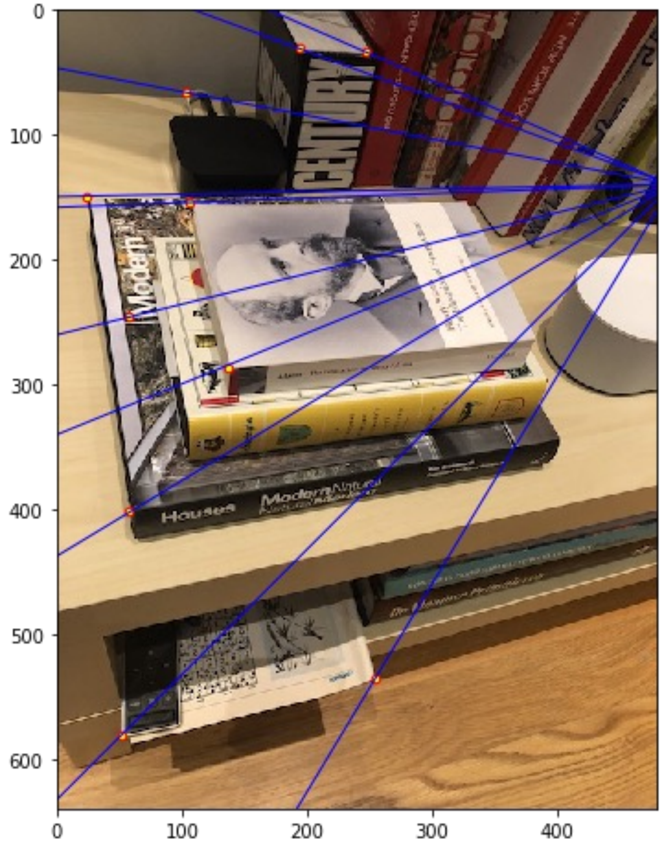
- 3x3 = 9 DOF
- However, scale, rank 2 !
- => 7 DOF



Epipoles inside the image: zoom-like setup.

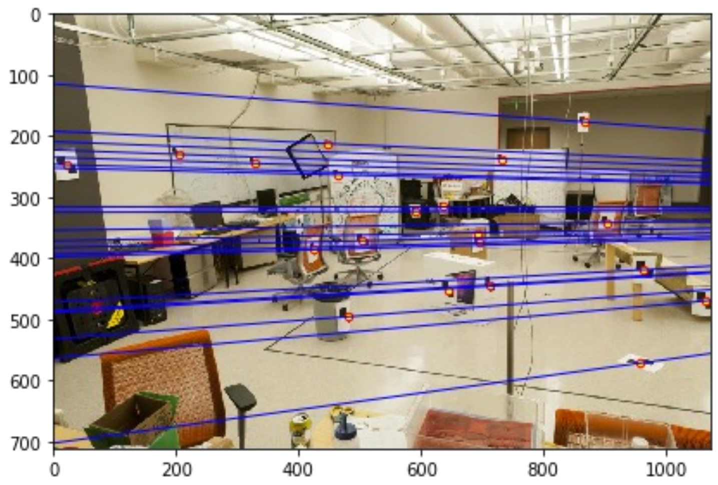
Epipoles are where the other camera is!

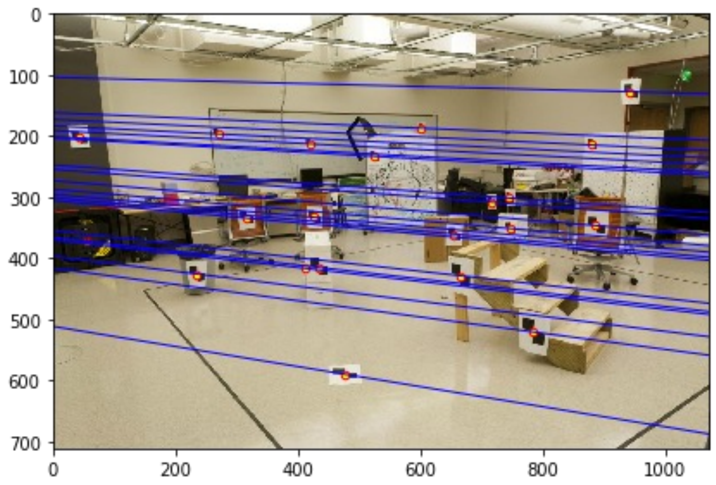




Epipoles in near-stereo config.

Epipoles are where the other camera is!





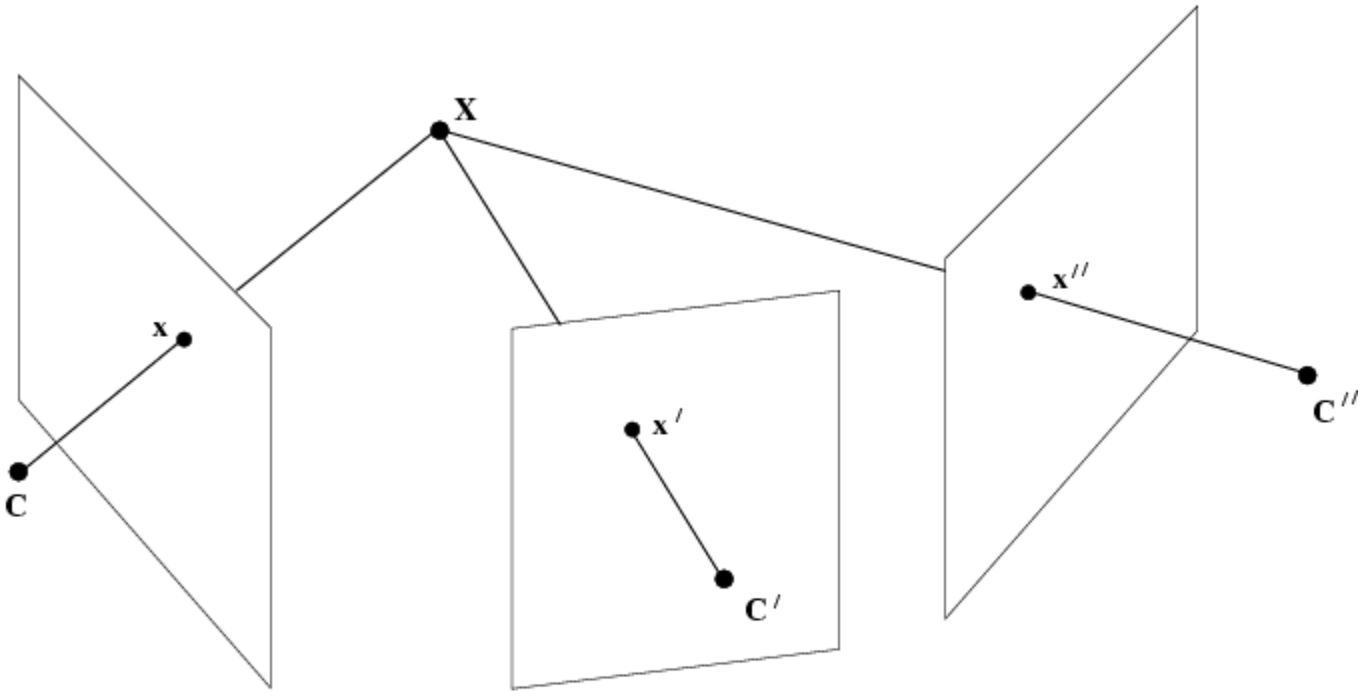
Trinocular Camera rig

https://www.skydio.com/

Navigation Camera System

CONFIGURATION	6x cameras in trinocular configuration top and bottom
SENSOR TYPE	Sony 1/3" 4K color CMOS
LENS APERTURE	f/1.8
FIELD-OF-VIEW	200°
ENVIRONMENT COVERAGE	True 360°

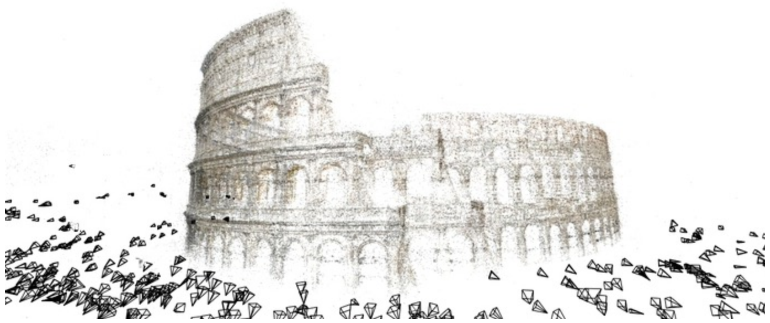
Trifocal Geometry



$$[\mathbf{x}']_\times (\sum_i x^i \mathtt{T}_i)[\mathbf{x}'']_\times = 0_{3\times 3}$$

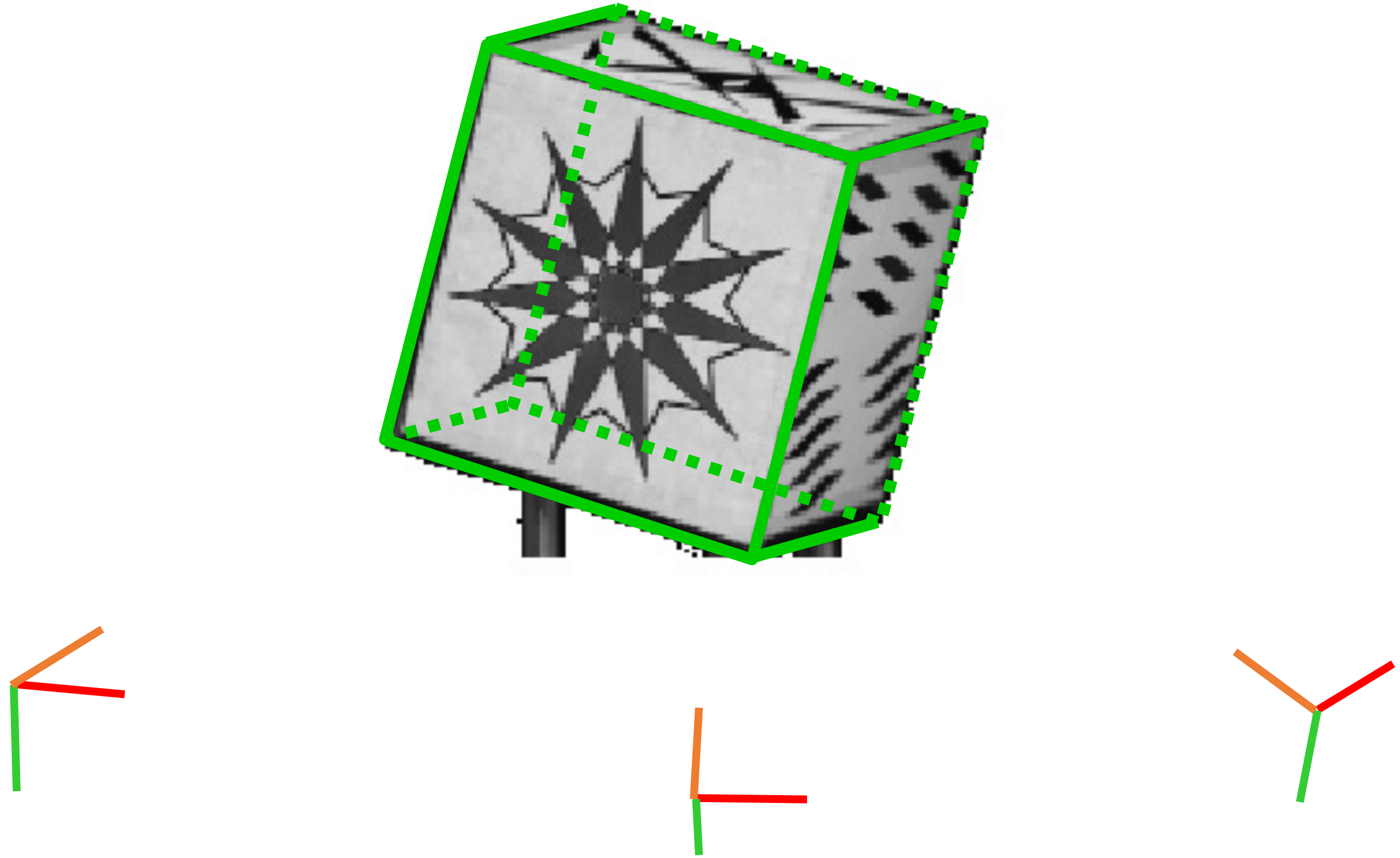
Outline

- Structure from Motion
- Correspondence
- Optimization
- Visual SLAM
- 4D Reconstruction
- Dense Surface Reconstruction



An Optimization Problem

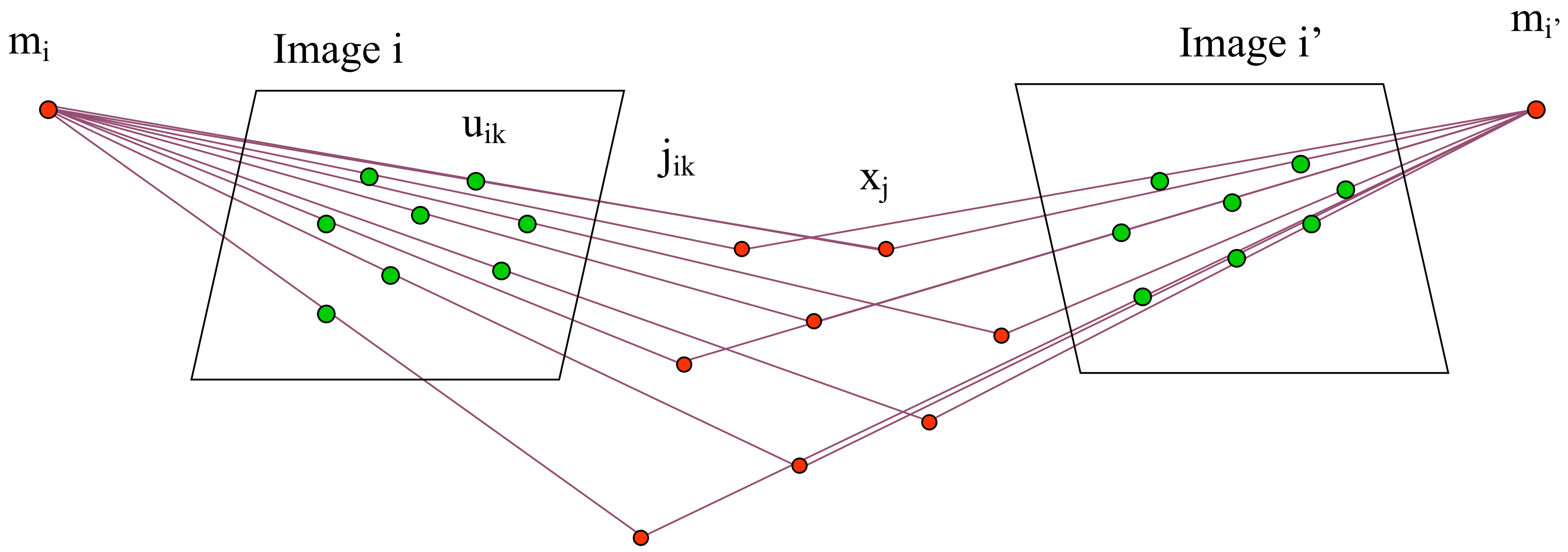
- Find the most likely structure and motion Θ



Optimization

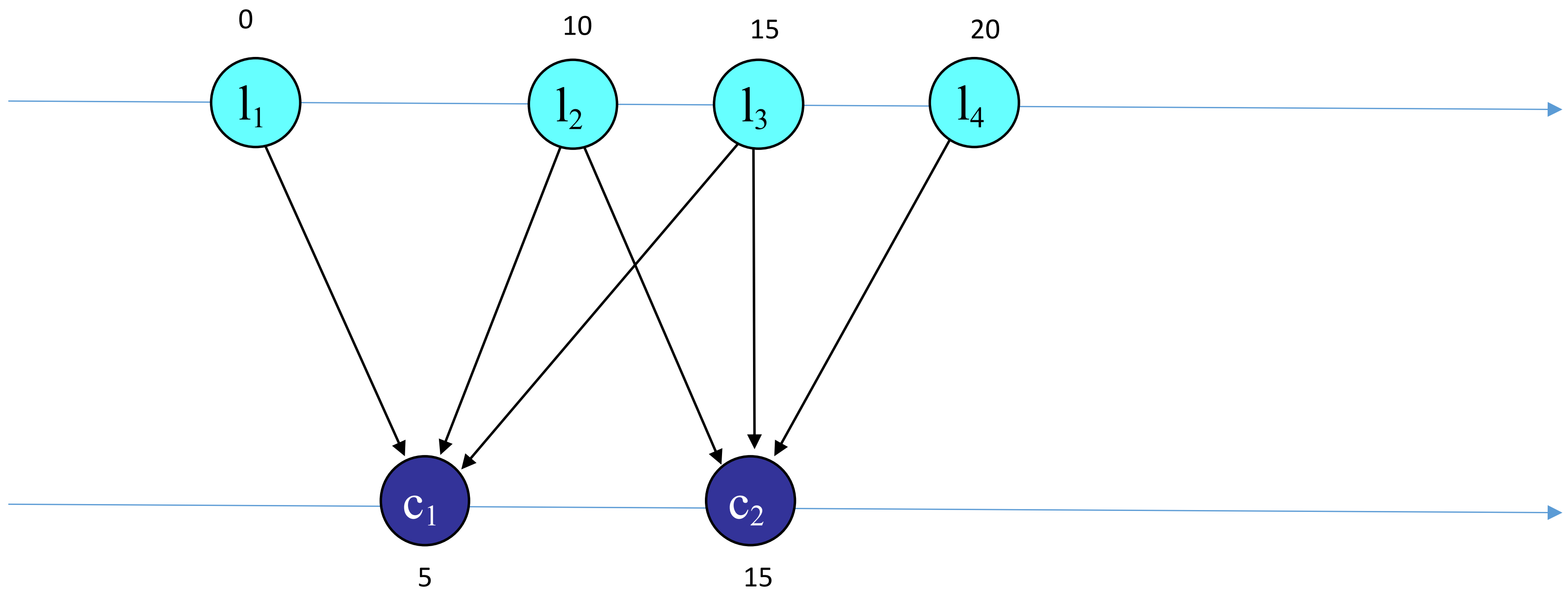
=Non-linear Least-Squares !

$$\sum_{i=1}^{m} \sum_{k=1}^{K_i} \|\mathbf{u}_{ik} - \mathbf{h}(\mathbf{m}_i, \mathbf{x}_{\mathbf{j}_{ik}})\|^2$$



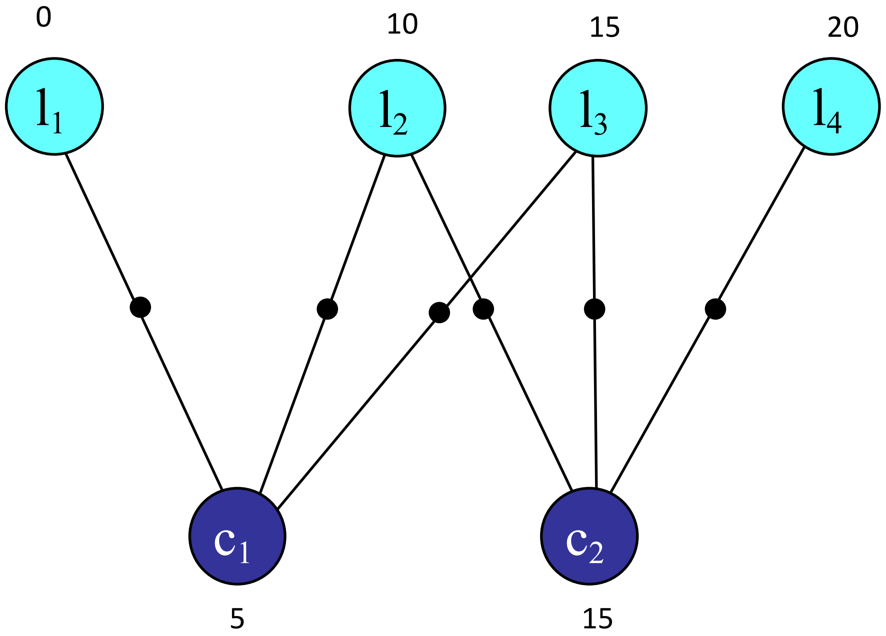
Sparse nonlinear least squares

- Simple 1-Dimensional Example
- p = 2 cameras and 4 points: {c1 c2 l1 l2 l3 l4}
- $f(u_{ik};p)$ = difference in x position = $l_{j(ik)} - c_i$

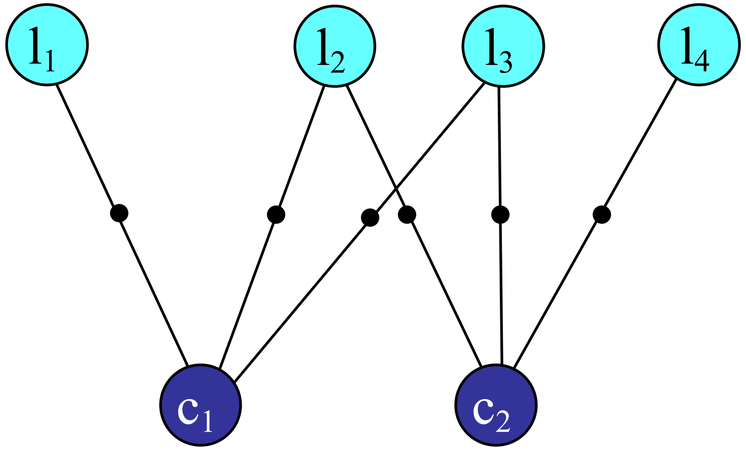


Model with Factor Graphs

- Connectivity = sparsity!
- Factor is function of small set.



Sparse Jacobian and Hessian

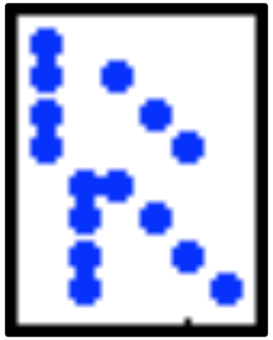


A =

c1	c2	l1	l2	l3	l4
1	0	0	0	0	0
-1	0	1	0	0	0
-1	0	0	1	0	0
-1	0	0	0	1	0
0	-1	1	0	0	0
0	-1	0	1	0	0
0	-1	0	0	1	0
0	-1	0	0	0	1

b =

5
-5
5
10
-15
-5
0
5



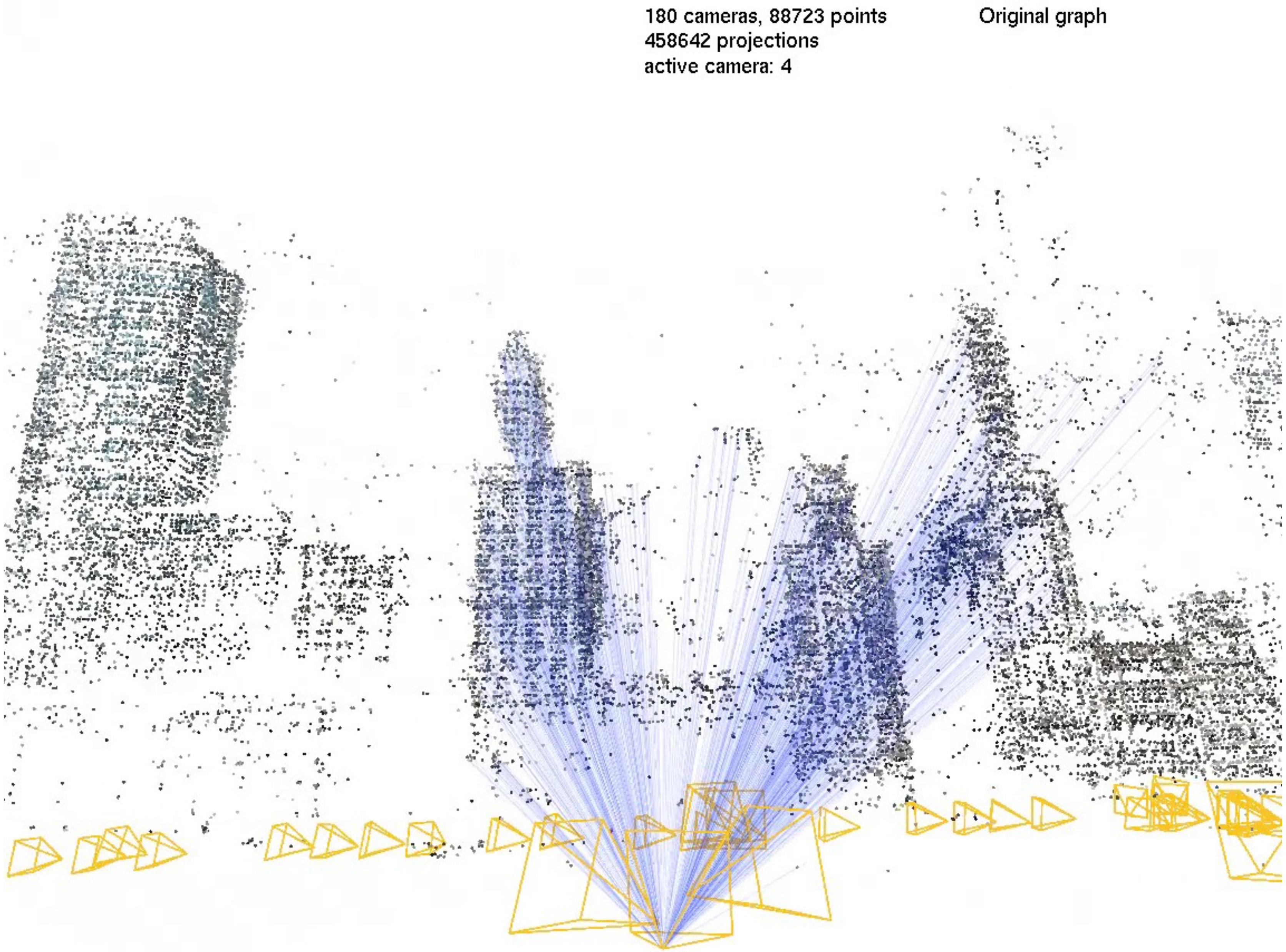
A'*A

c1	c2	l1	l2	l3	l4
4	0	-1	-1	-1	0
0	4	-1	-1	-1	-1
-1	-1	2	0	0	0
-1	-1	0	2	0	0
-1	-1	0	0	2	0
0	-1	0	0	0	1

(A'*A)\A'*b =

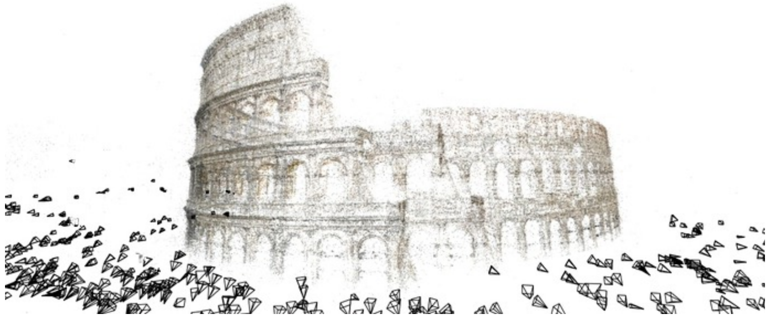
5.0000
15.0000
0.0000
10.0000
15.0000
20.0000

Structure from Motion (Chicago, movie by Yong Dian Jian)

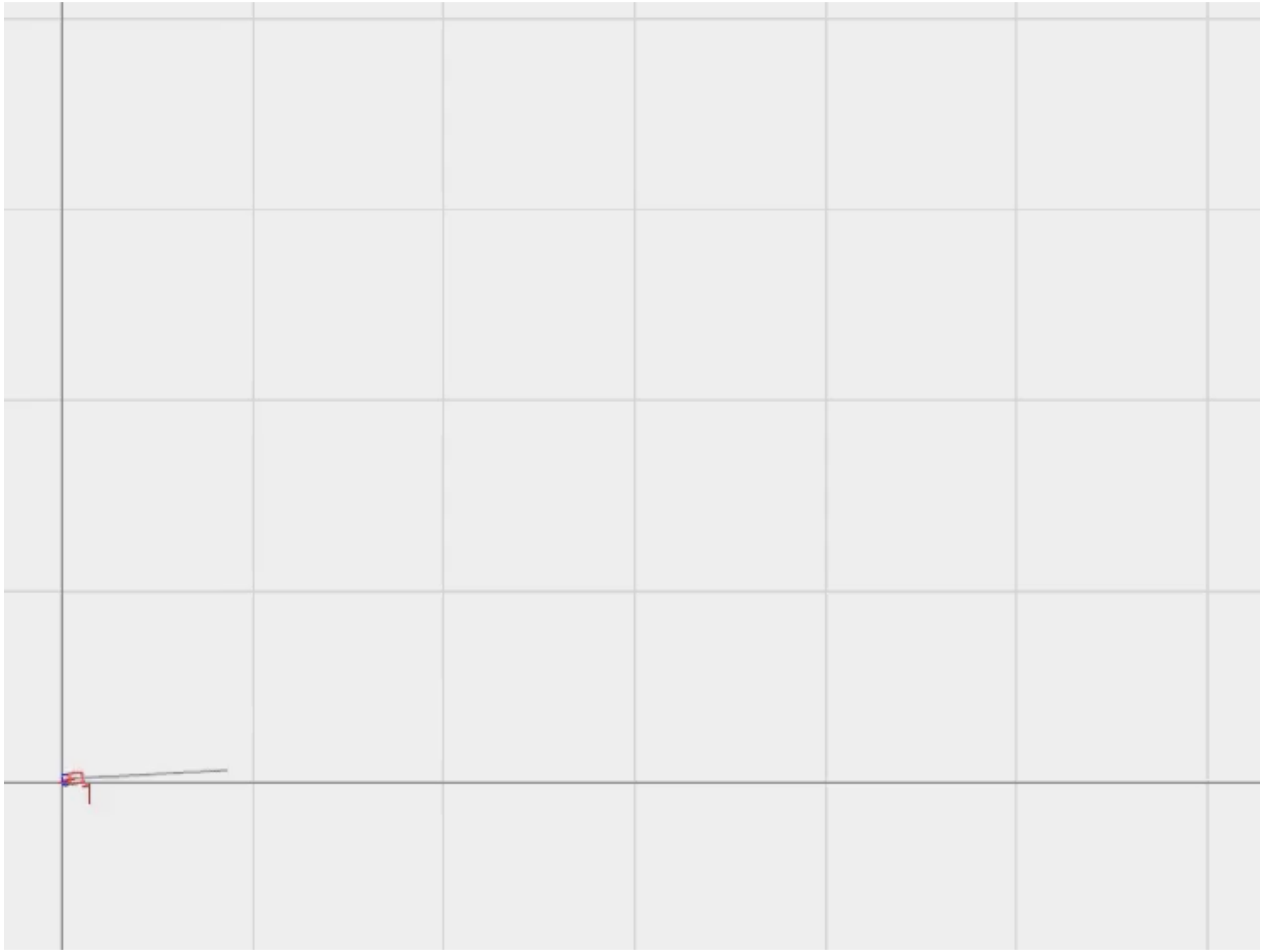


Outline

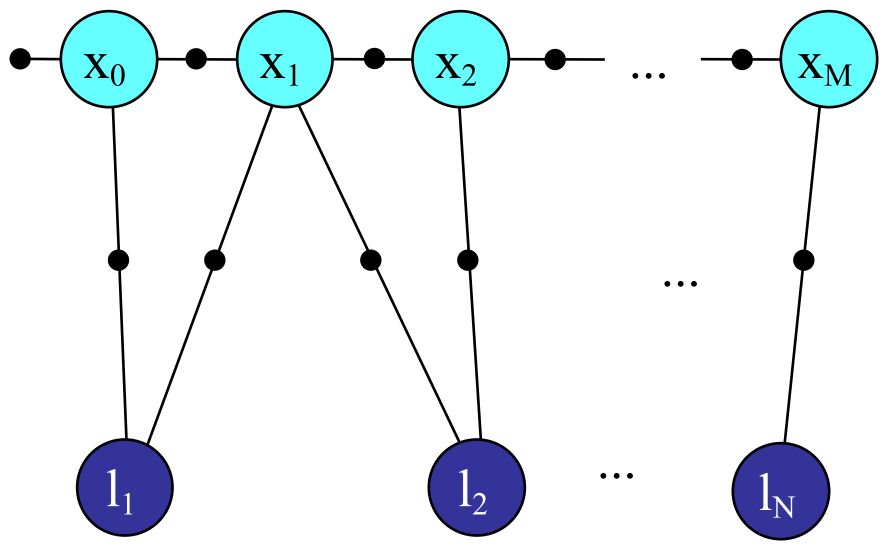
- Structure from Motion
- Correspondence
- Optimization
- Visual SLAM
- 4D Reconstruction
- Dense Surface Reconstruction



Visual SLAM: SfM for Robots



Visual SLAM Factor Graph

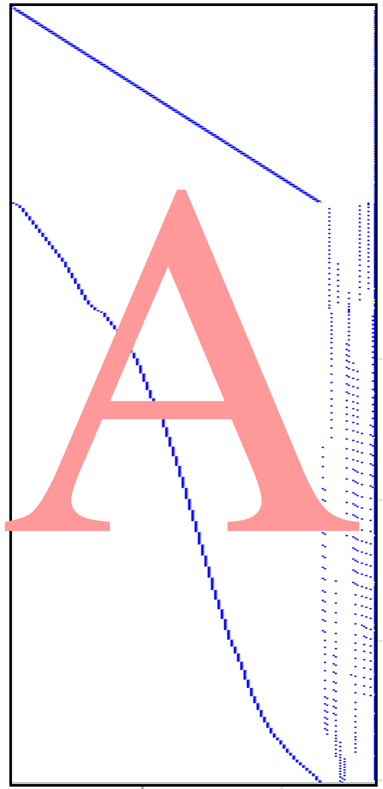


- Trajectory of Robot

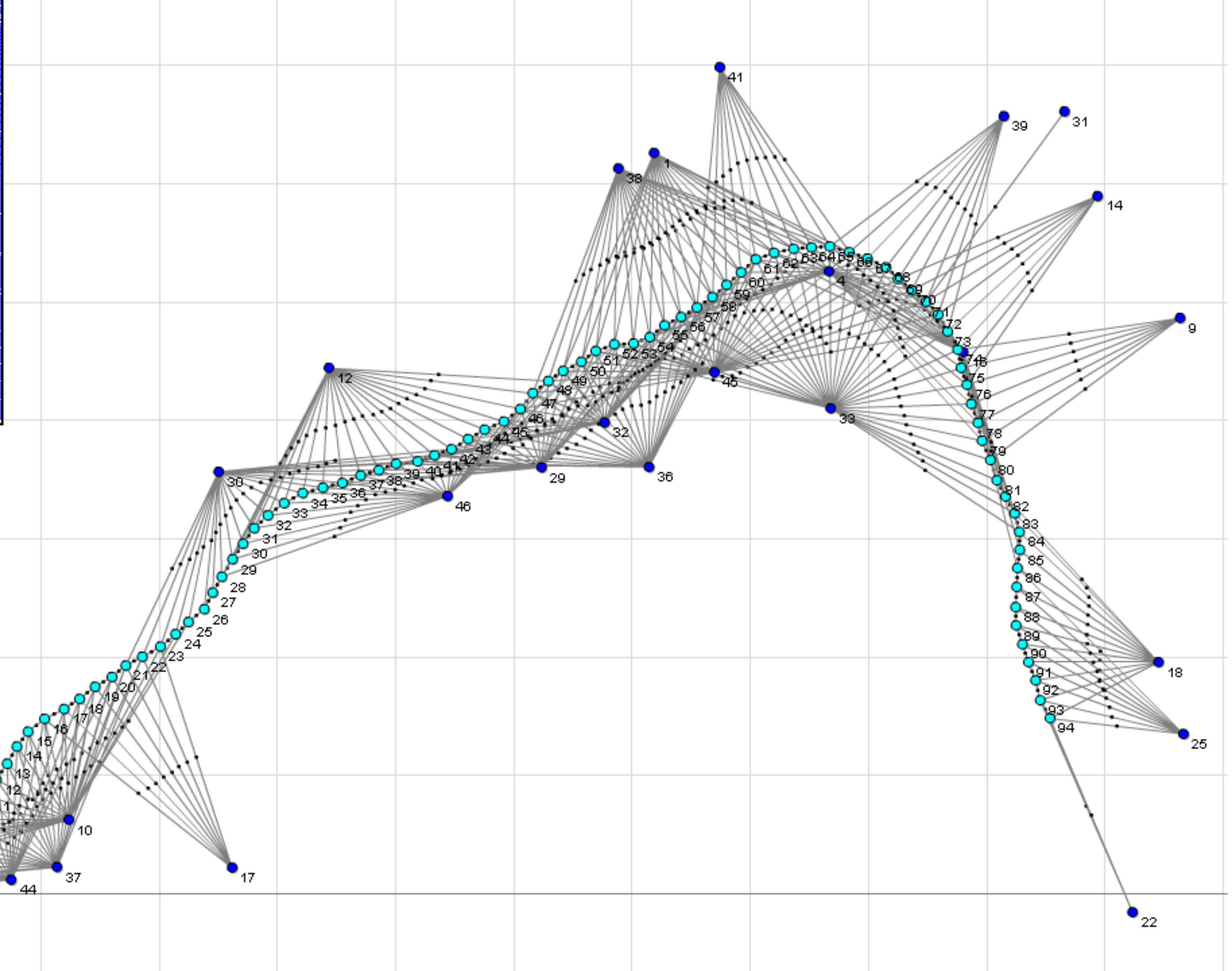
- Landmark Measurements

- “Landmarks”

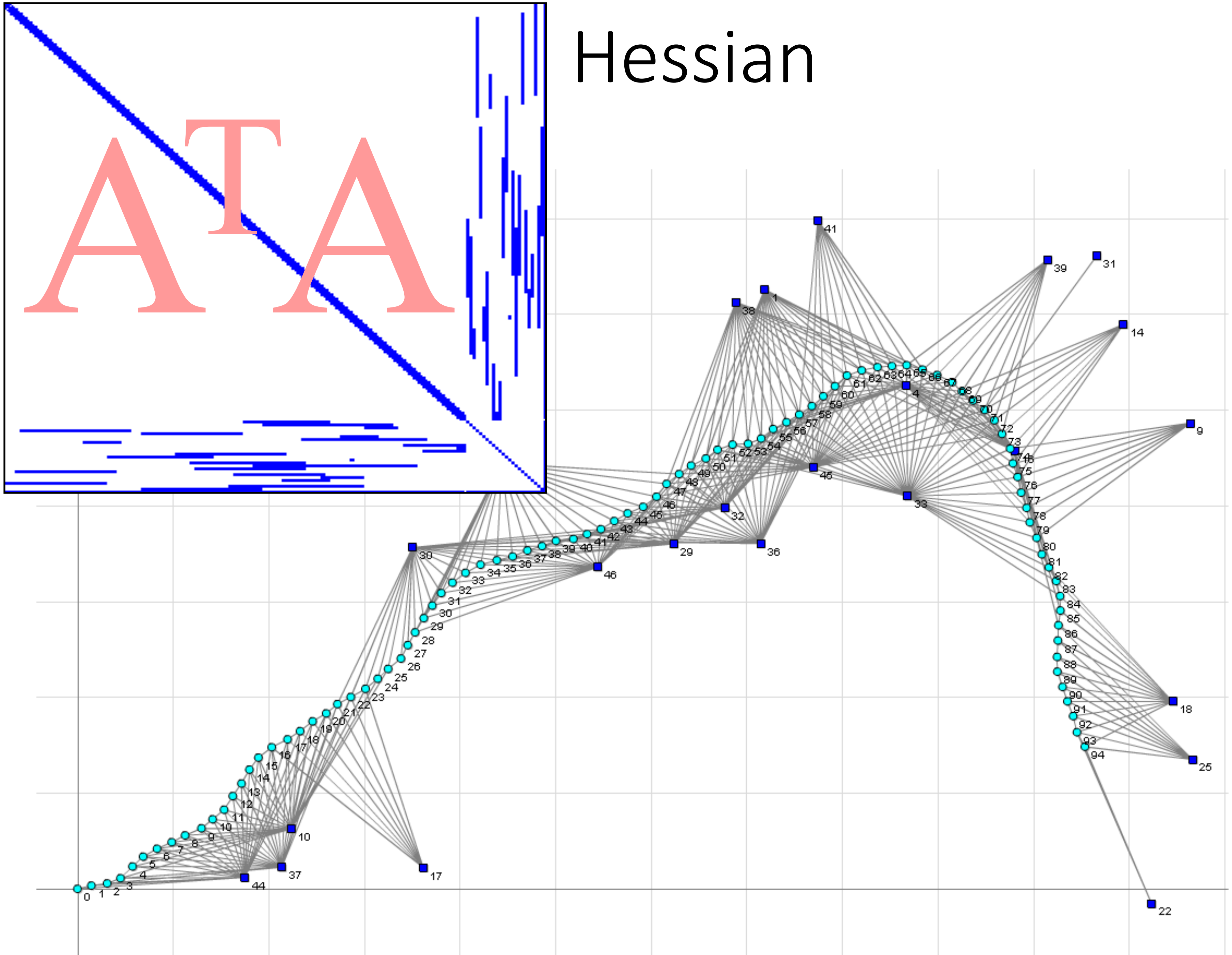
$$P(X,M) = k^* \, P(x_0) \prod_{i=1}^{M} P(x_i | x_{i-1}, u_i) \times \prod_{k=1}^{K} P(z_k | x_{i_k}, l_{j_k})$$



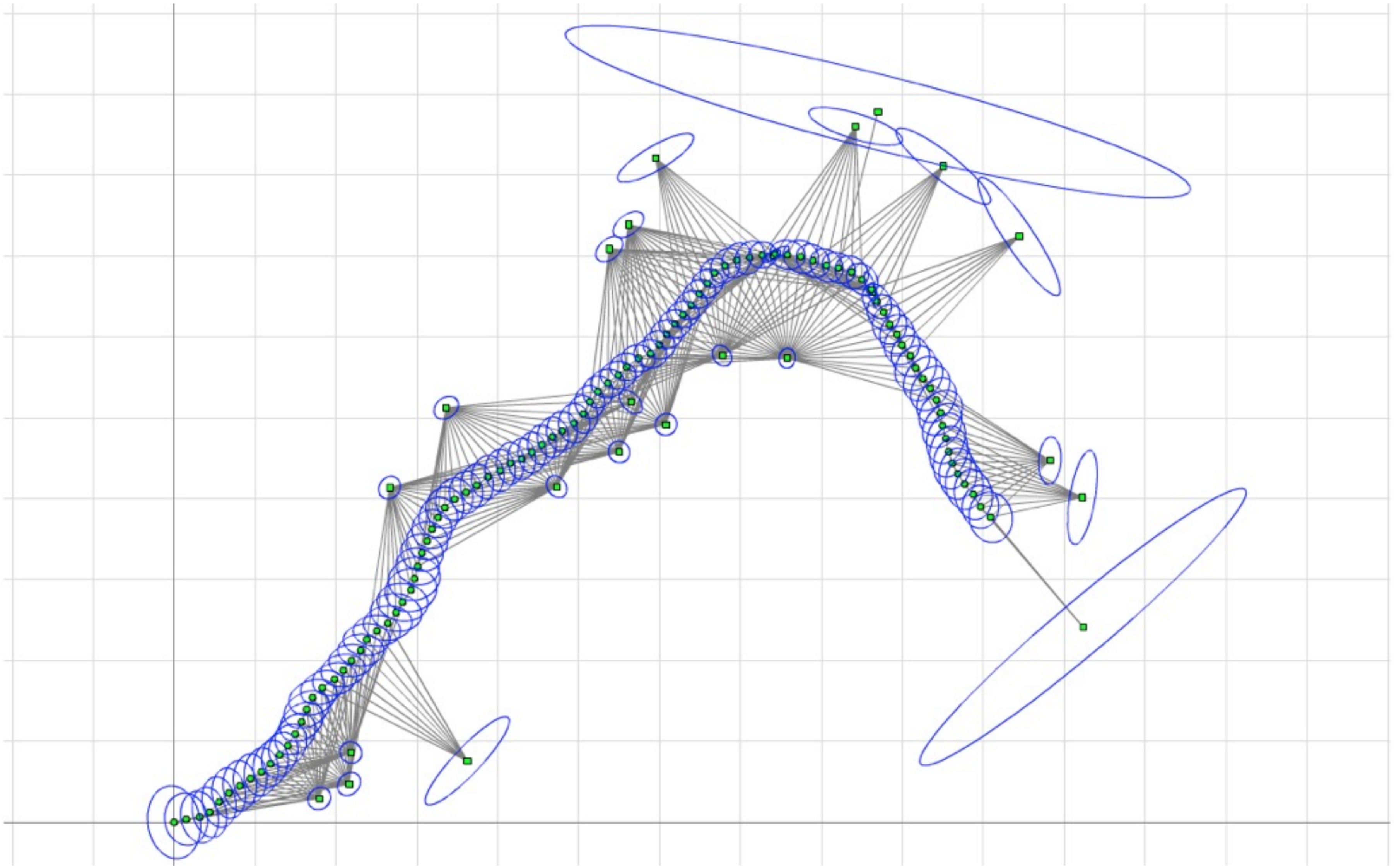
Visual SLAM Factor Graph



Hessian

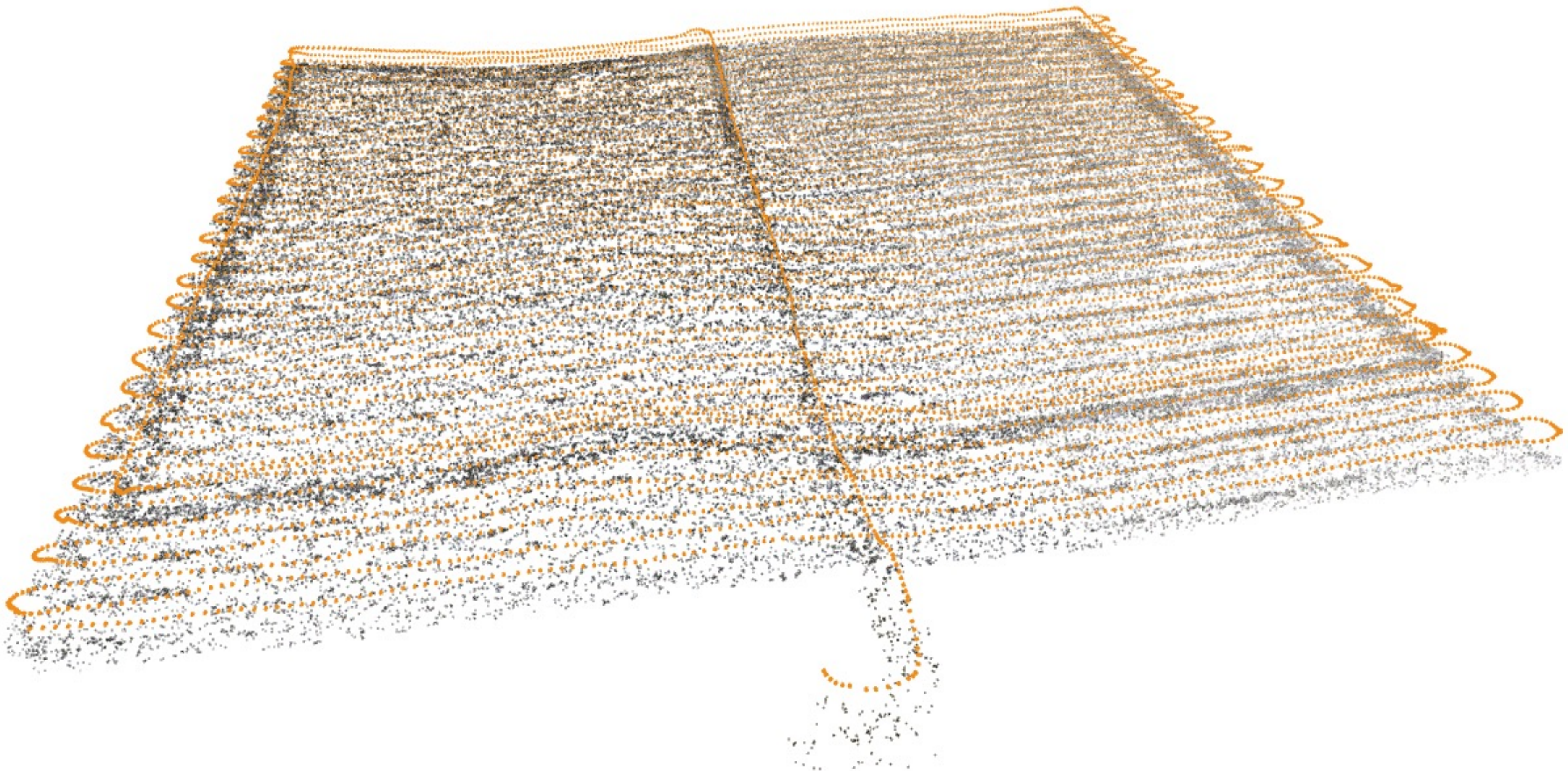


End result: Solution + Uncertainty



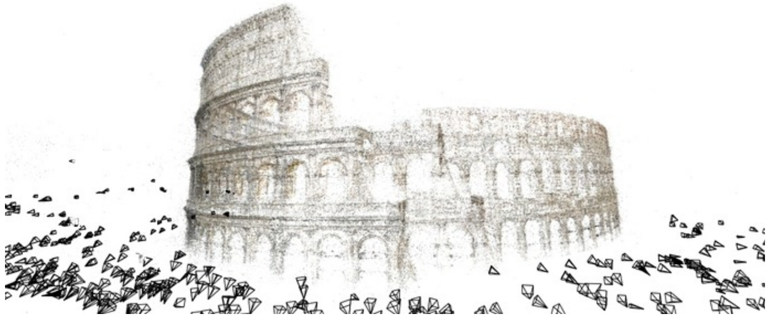
Example: Underwater SLAM

9831 camera poses, 185261 landmarks, and 350988 factors



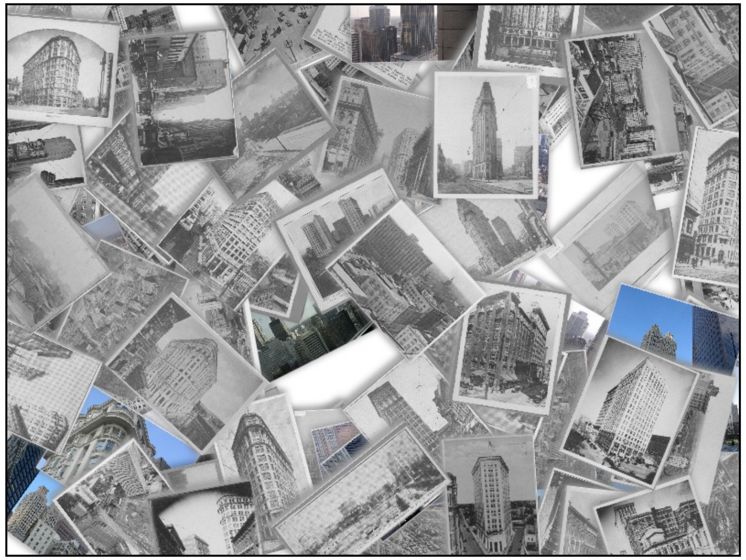
Outline

- Structure from Motion
- Correspondence
- Optimization
- Visual SLAM
- 4D Reconstruction
- Dense Surface Reconstruction

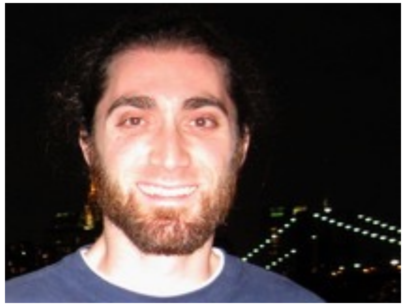


Spatiotemporal Reconstruction

4D Cities: 3D + Time



Historical Image Collection



Grant Schindler

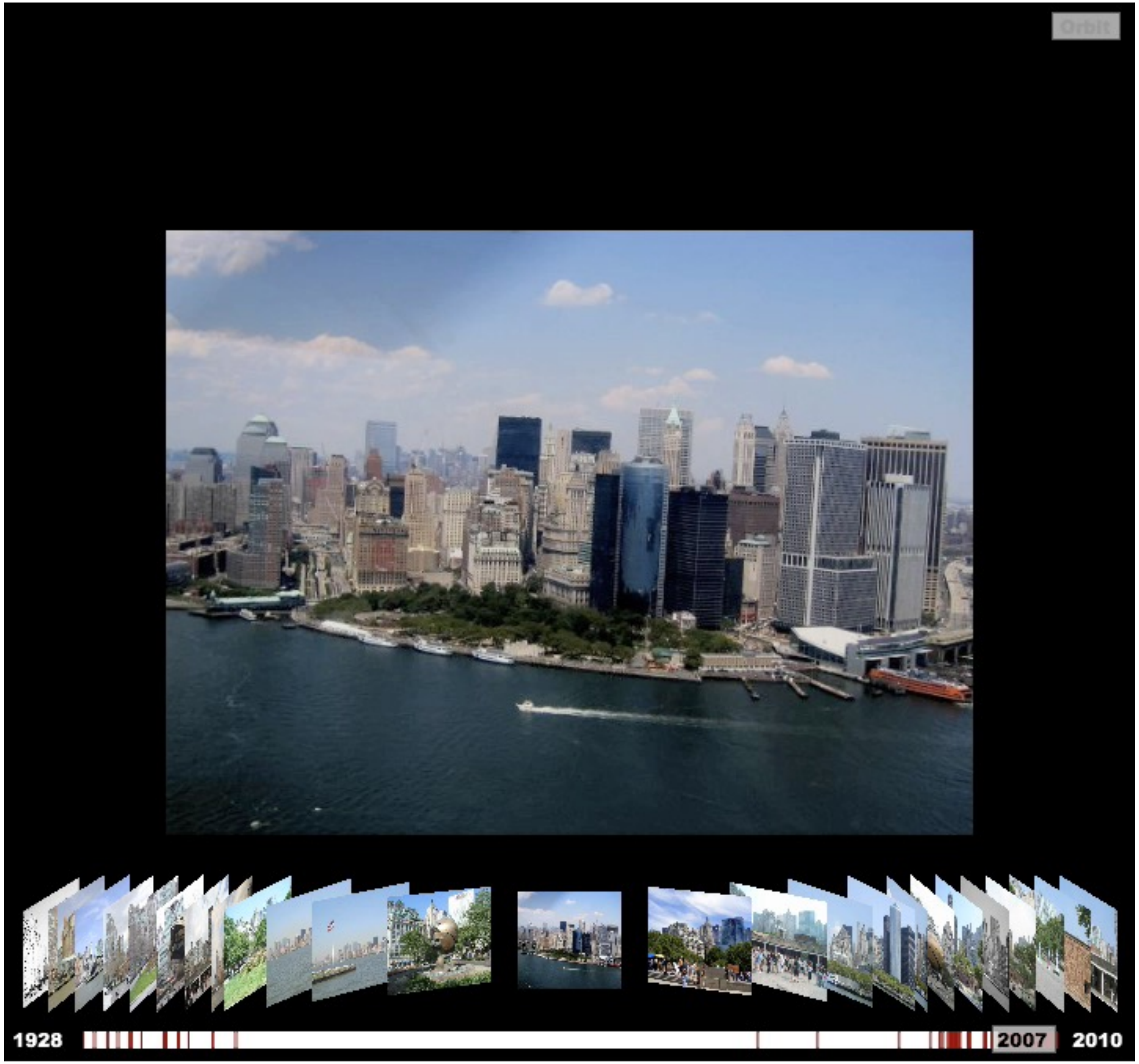
Supported by NSF CAREER, Microsoft

Recent revival: NSF NRI award on 4D crops for precision agriculture…

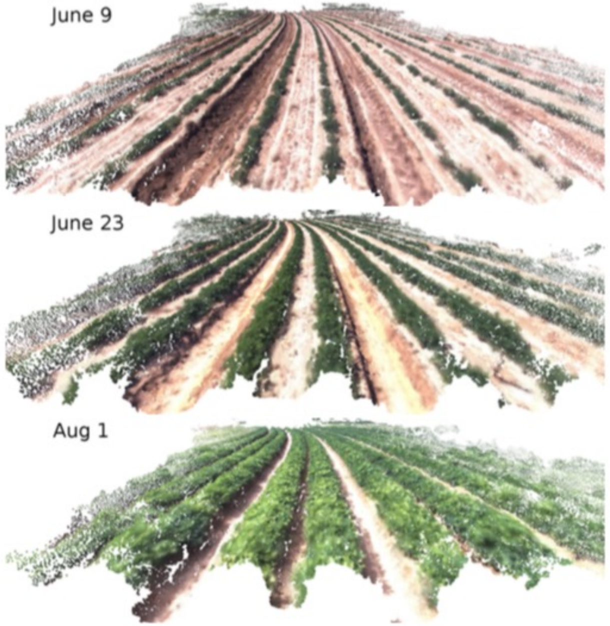
4D Reconstruction of Lower Manhattan

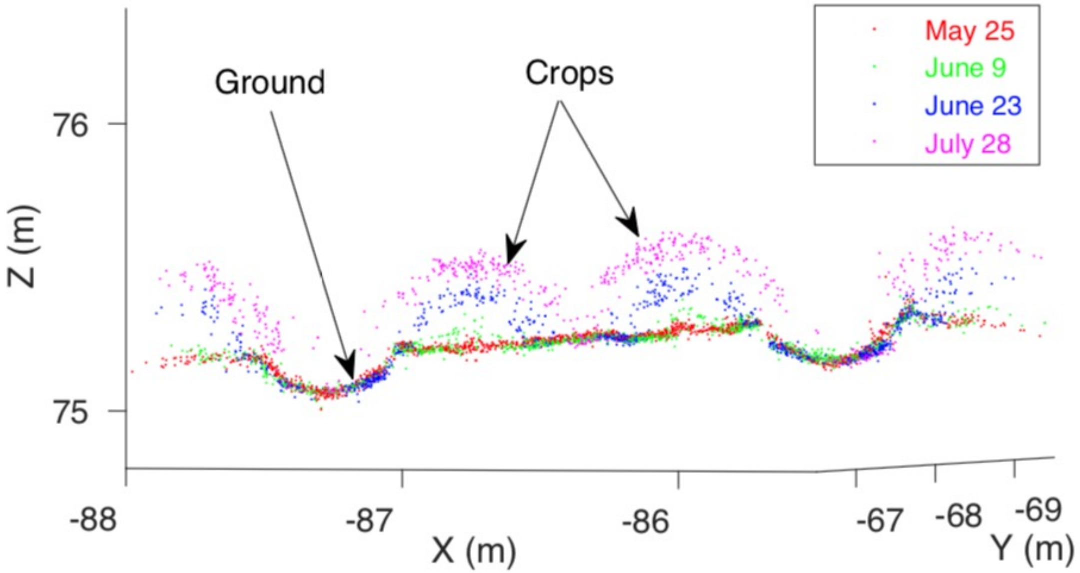
Probabilistic Temporal Inference on Reconstructed 3D Scenes, G. Schindler and F. Dellaert, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

4D Structure over Time

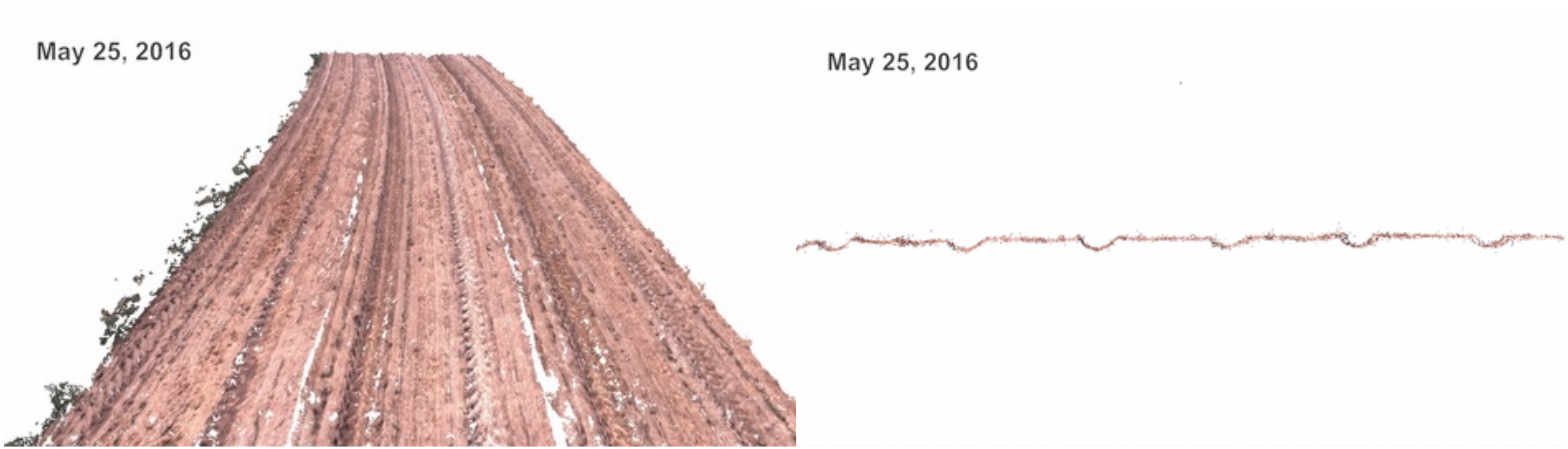


4D crop monitoring (Jing Dong)





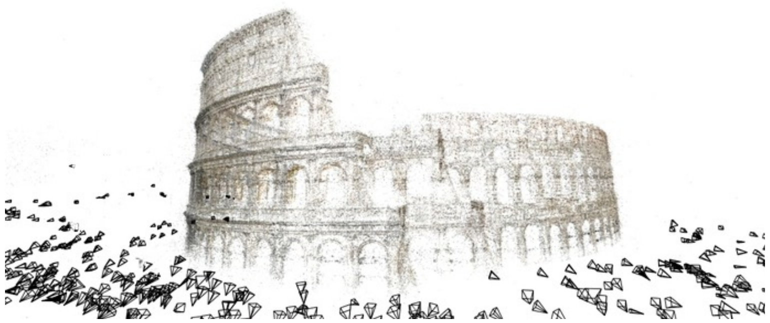
Results: video (by Jing Dong)



4D reconstruction results (by PMVS)
and its cross section

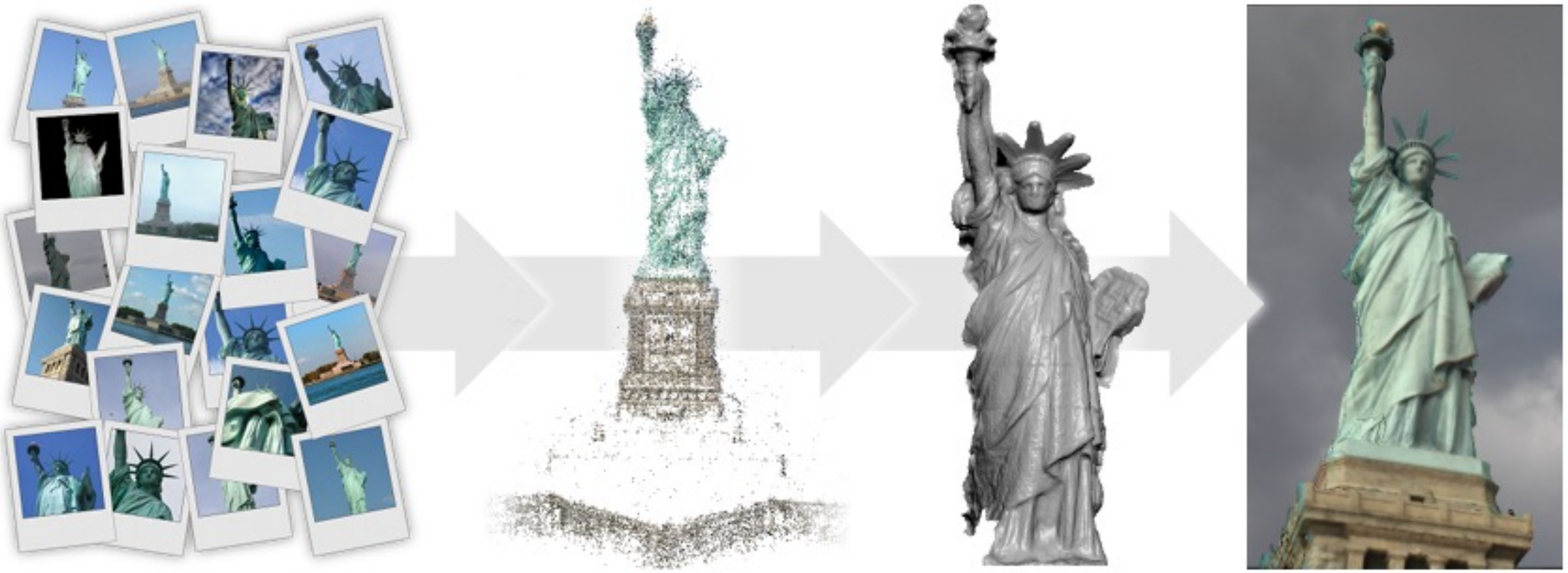
Outline

- Structure from Motion
- Correspondence
- Optimization
- Visual SLAM
- 4D Reconstruction
- Dense Surface Reconstruction



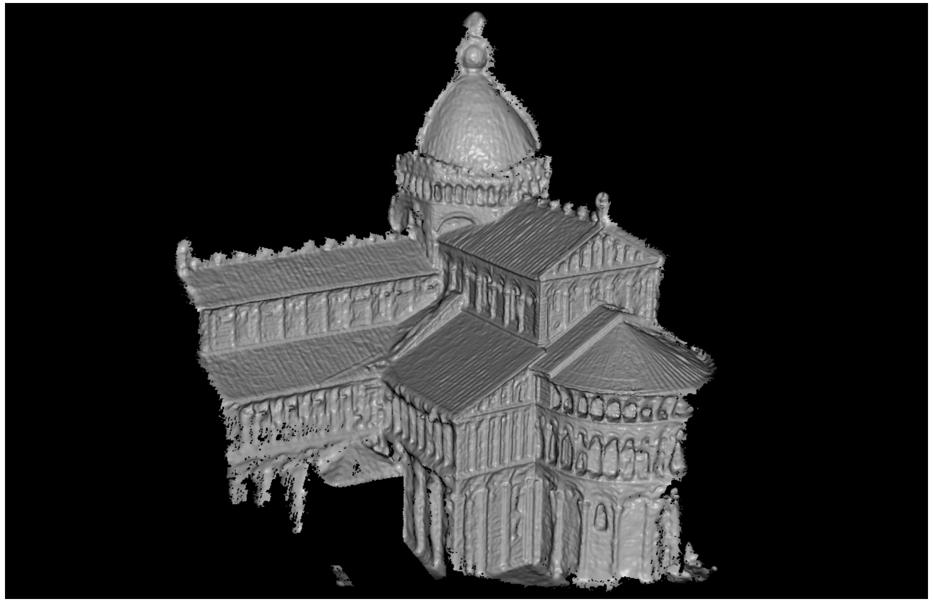
Multi-view Stereo

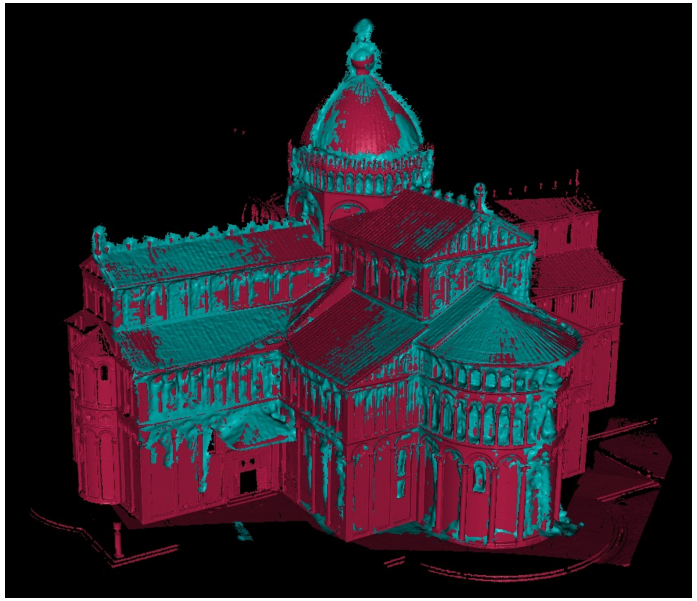
Multi-View Stereo for Community Photo Collections
Michael Goesele, Noah Snavely, Brian Curless, Hugues Hoppe, and Steven M. Seitz
ICCV 2007



Multi-view Stereo

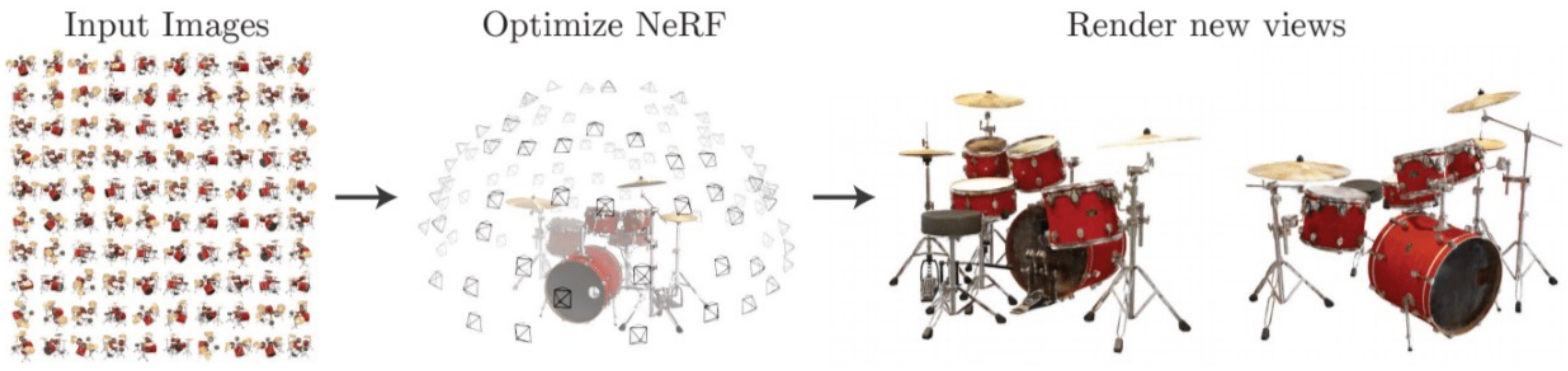
- Poisson Surface Reconstruction





Compared with Laser-Scanner

Neural Radiance Fields (NeRF)



- Original NeRF paper:
  - https://www.matthewtancik.com/nerf
- See two blog posts:
  - https://dellaert.github.io/NeRF/
  - https://dellaert.github.io/NeRF21/