CS 3630

Lecture 21:
Structure from Motion and Visual SLAM

Overview

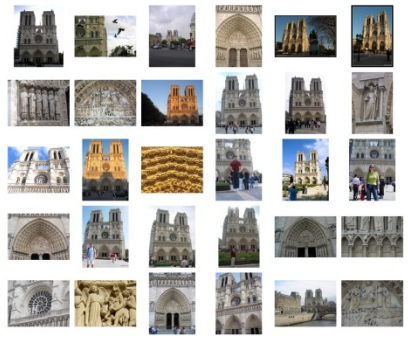
- Structure from Motion
- The Correspondence Problem
- Optimization
- Visual SLAM
- 4D Reconstruction

Multi-view Stereo

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

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.,

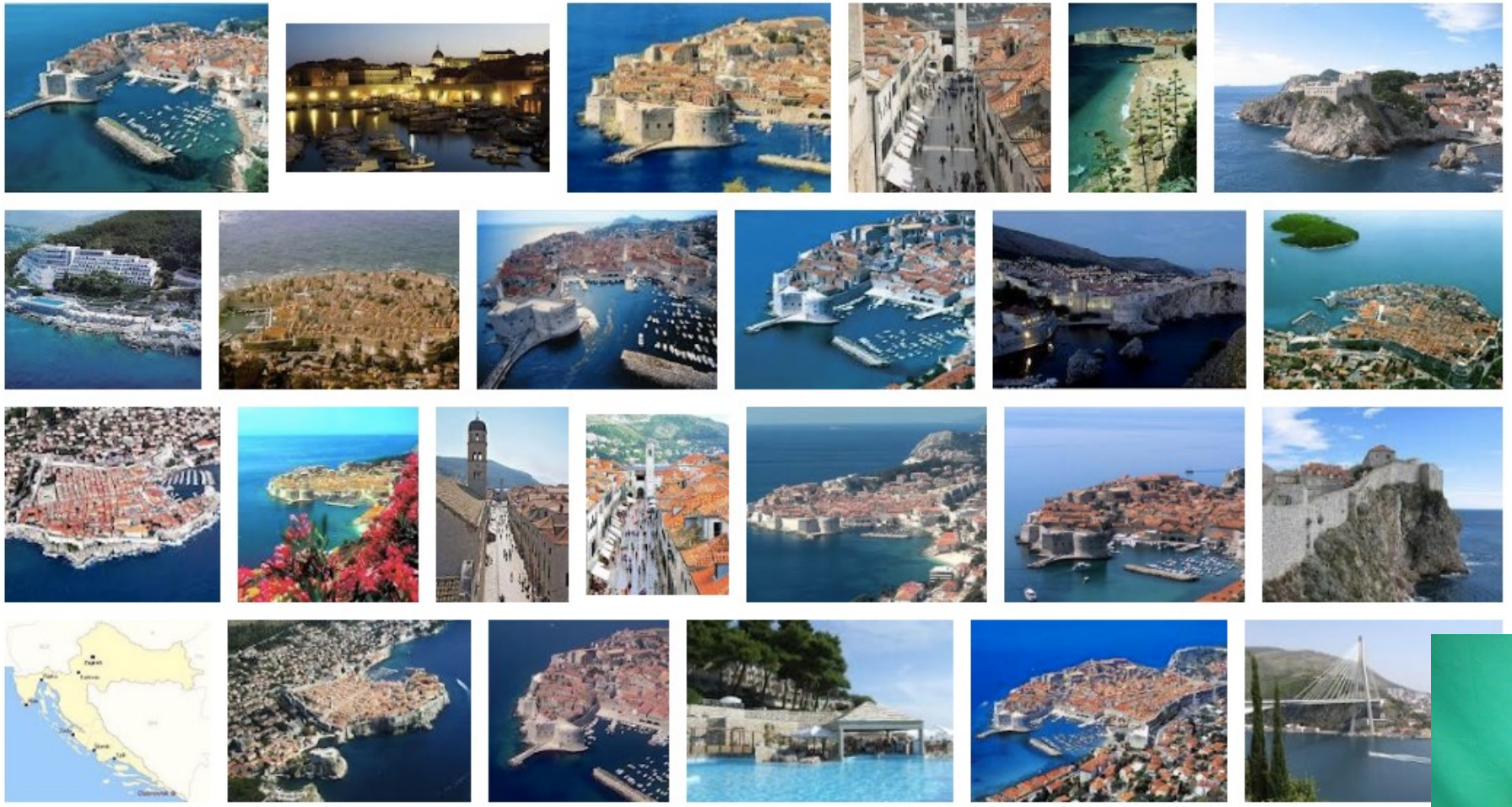


Figure by Aggarwal et al.

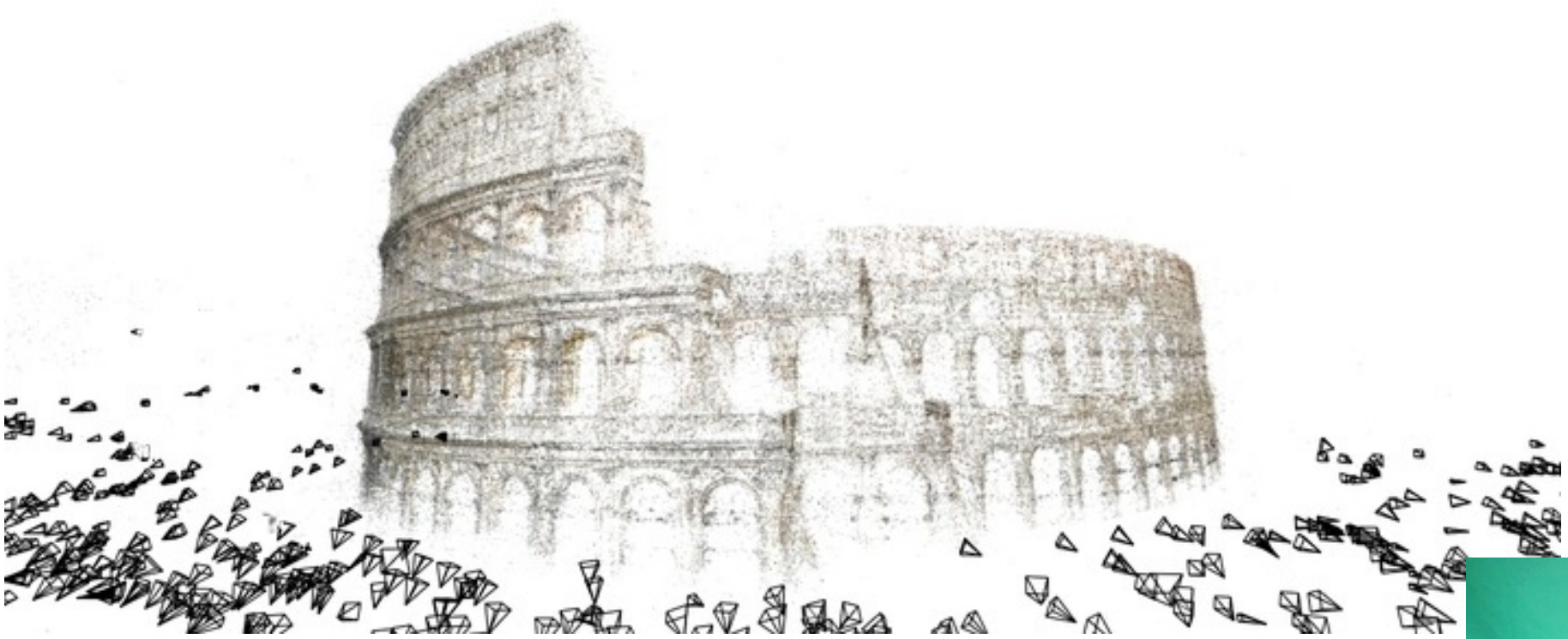
3D Models from Community Databases

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

Movie by Aggarwal et al.

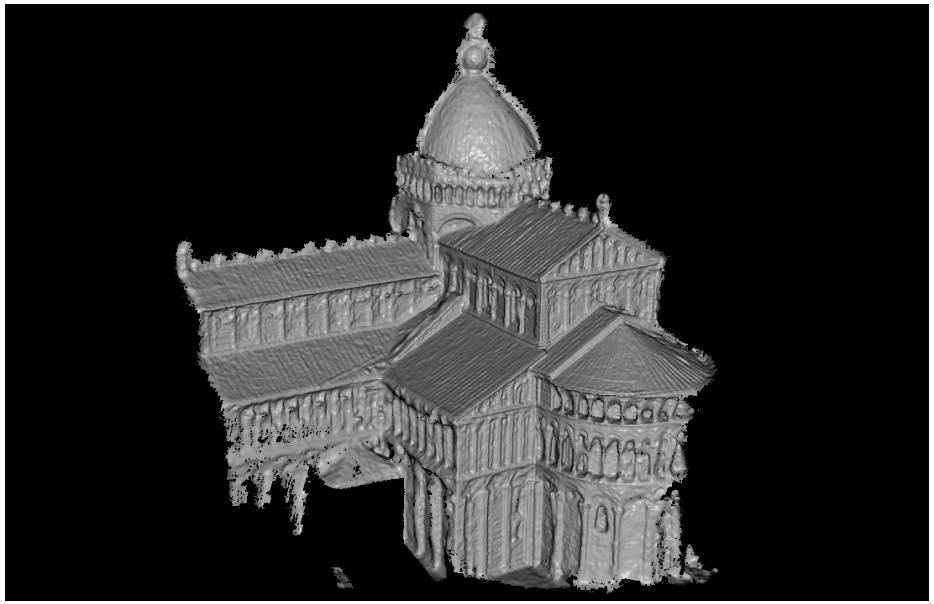
Building Rome in a Day

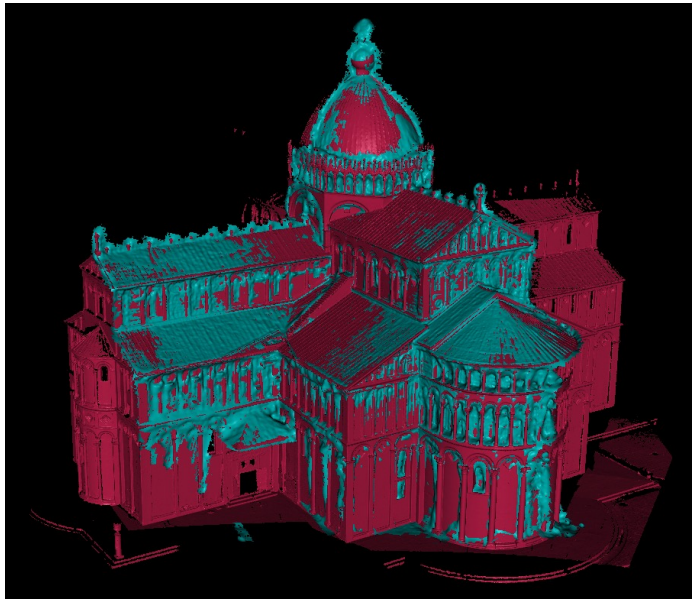
Building Rome in a Day Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski International Conference on Computer Vision, 2009, Kyoto, Japan.



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

Multi-view Stereo





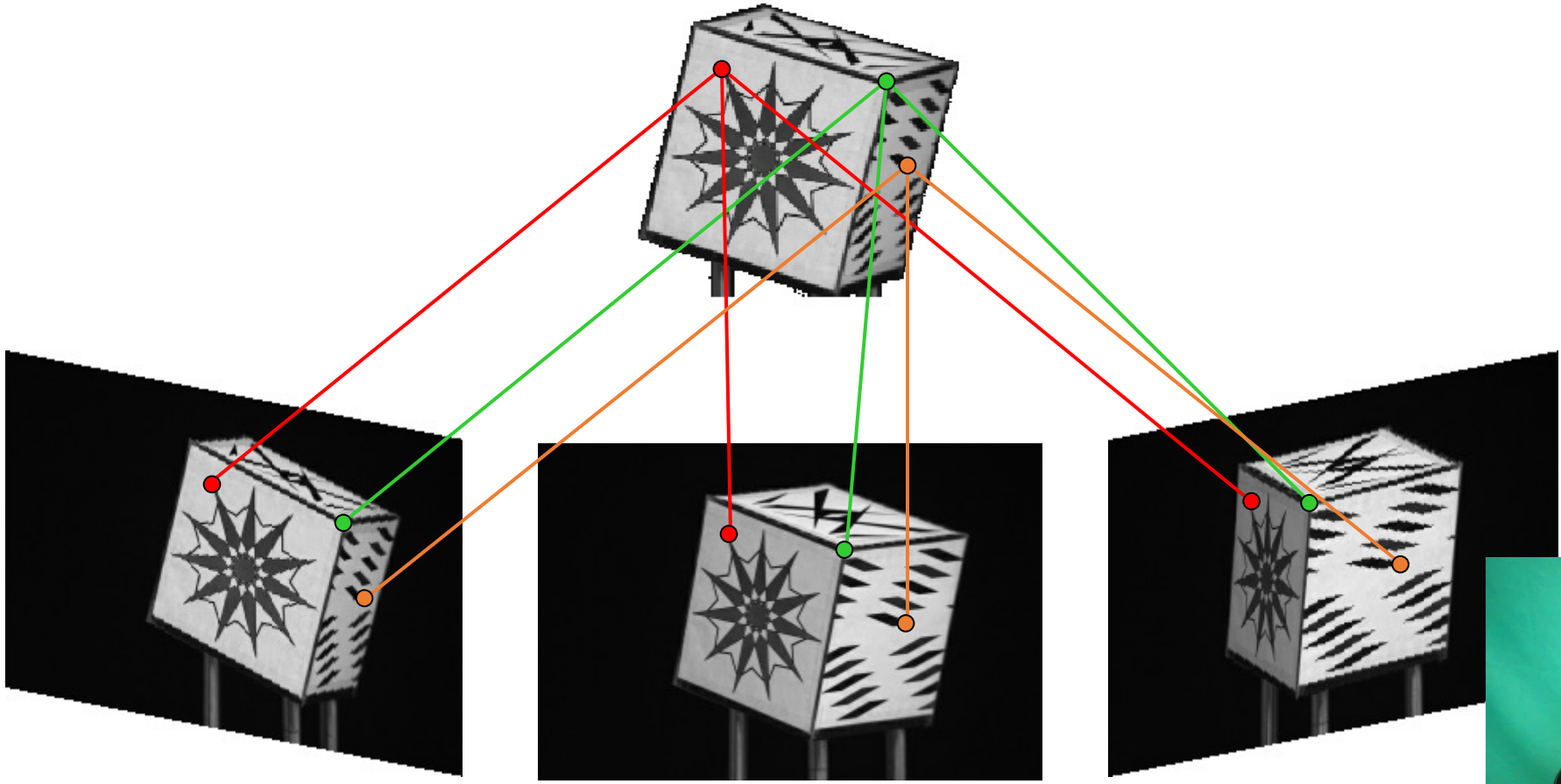
Compared with Laser-Scanner

2 Problems !

Correspondence

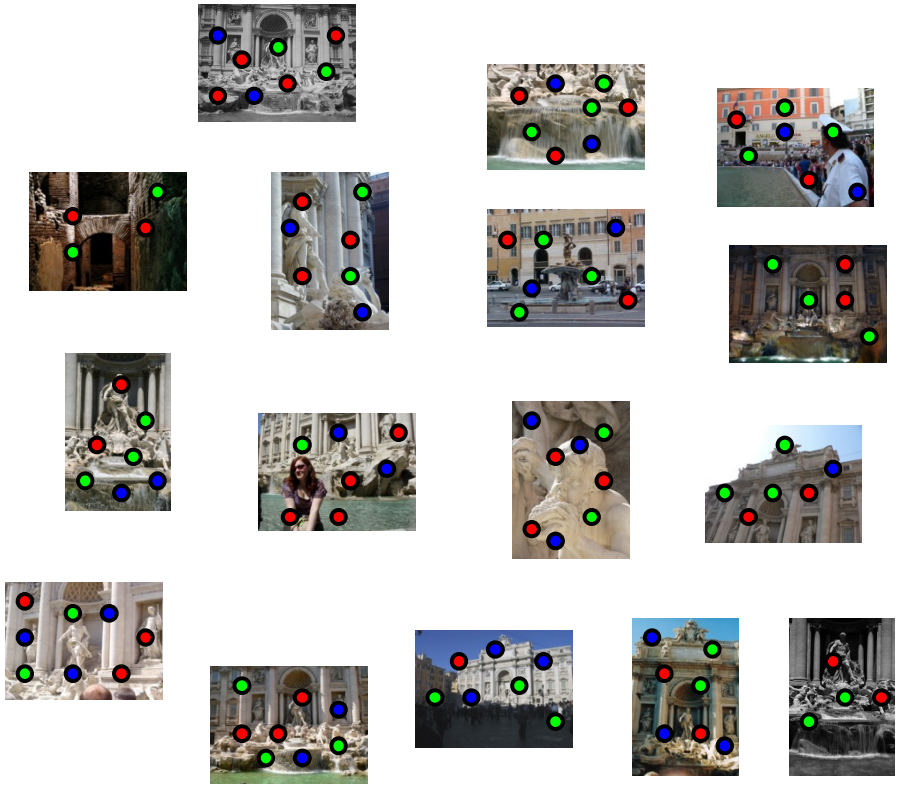
Optimization

A Correspondence Problem



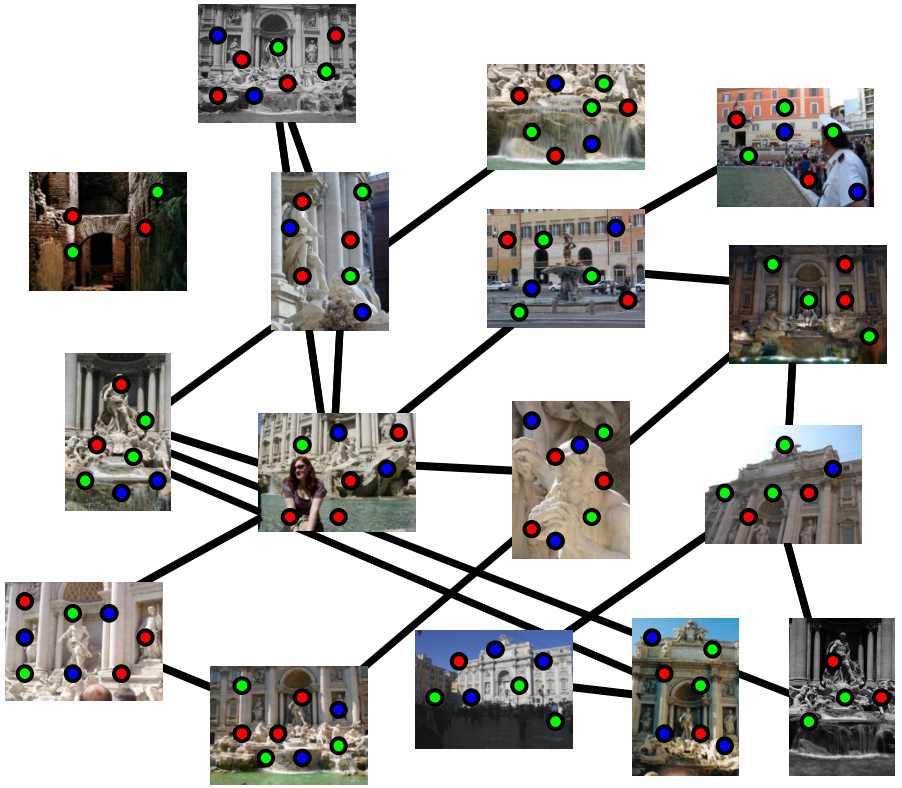
Feature detection

- Detect features using SIFT [Lowe, IJCV 2004]

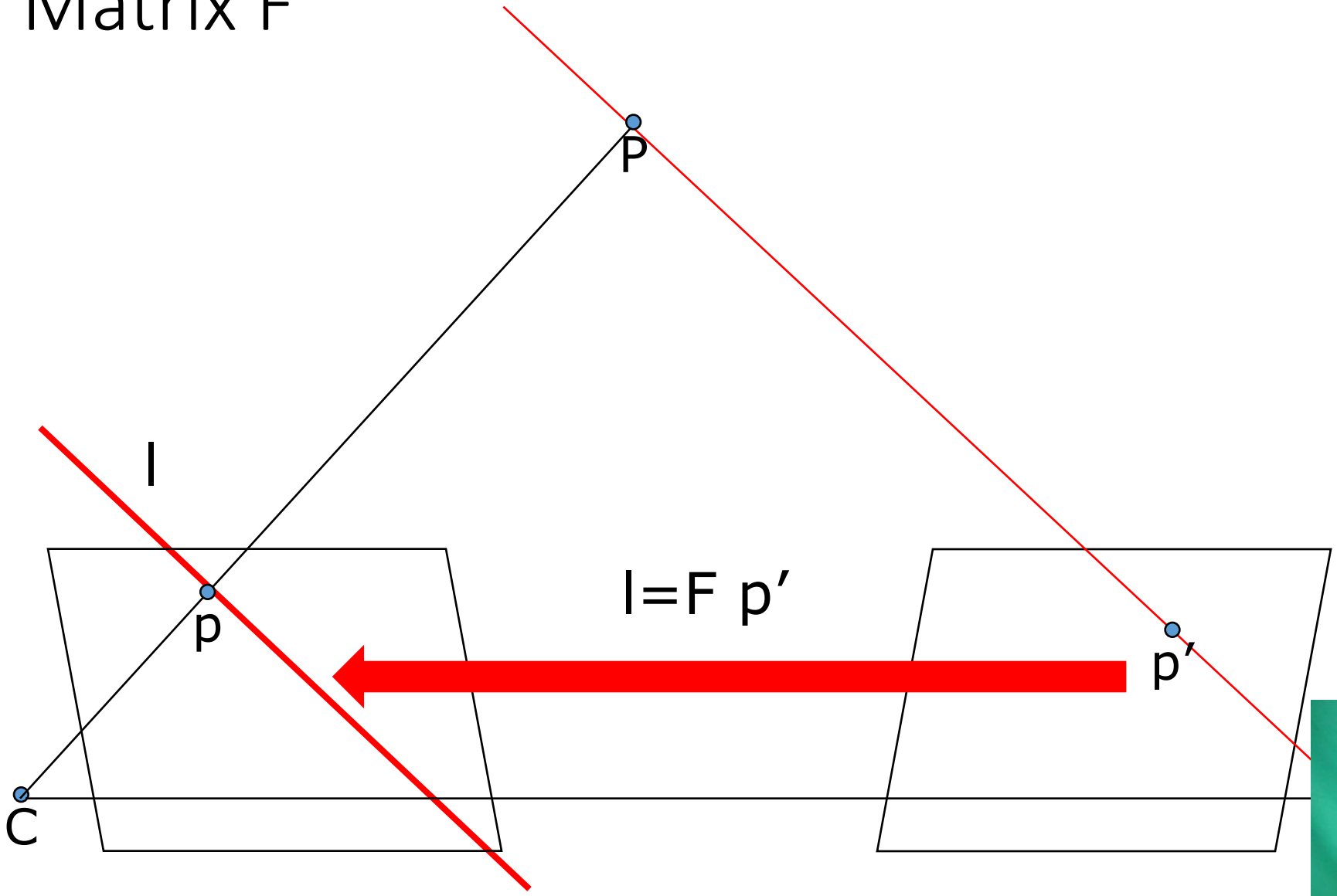


Feature matching

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



Recap: Two views and Fundamental Matrix F



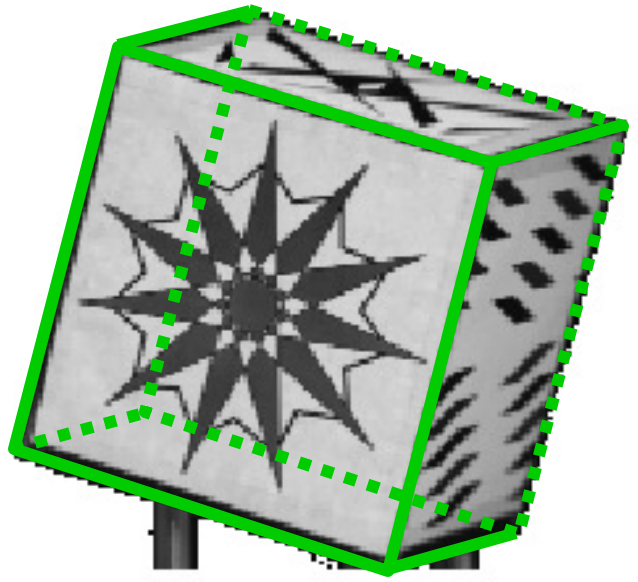
2 Problems !

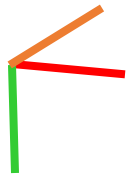
Correspondence

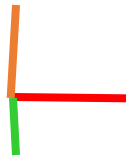
Optimization

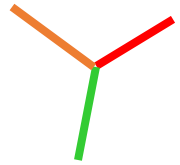
An Optimization Problem

- Find the most likely structure and motion Θ





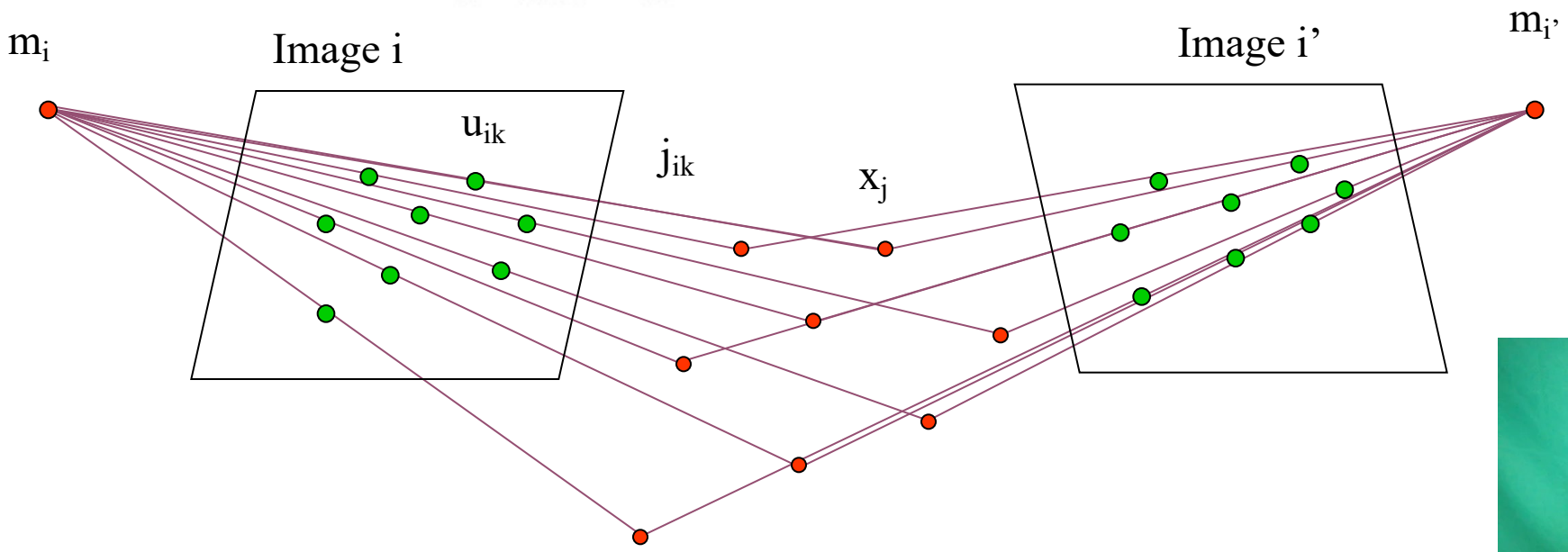




Optimization

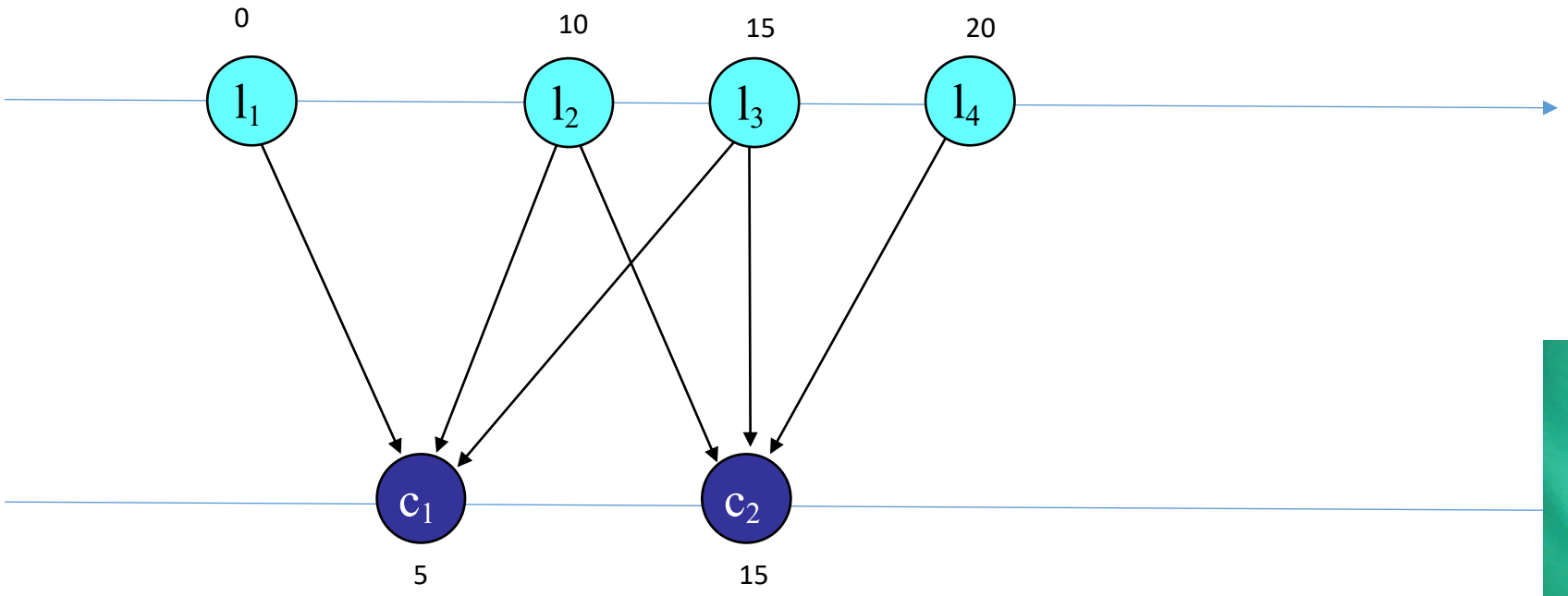
=Non-linear Least-Squares !

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



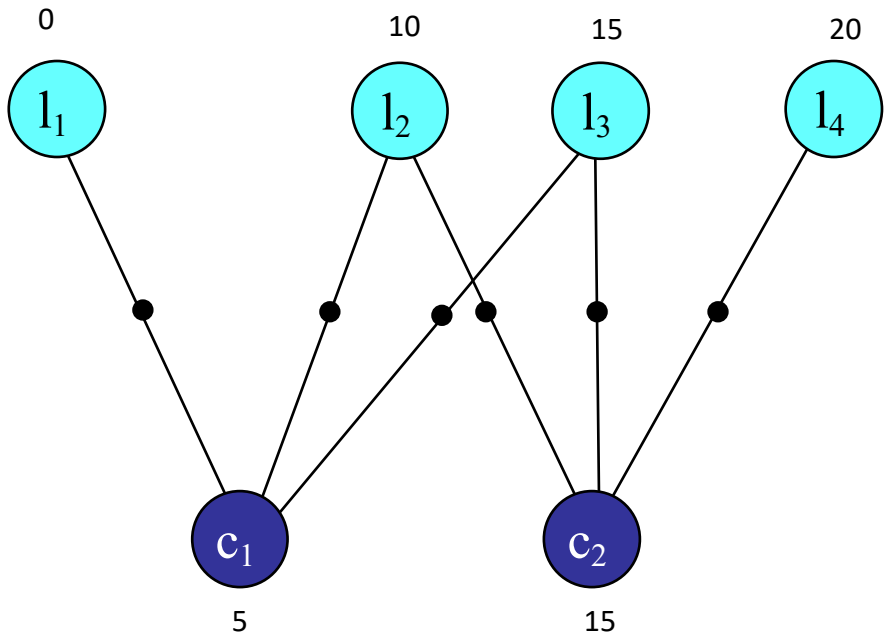
Sparse nonlinear least squares

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

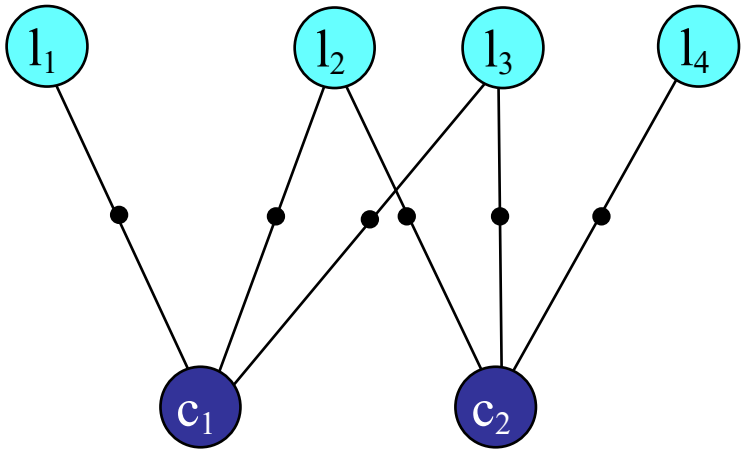


Model with Factor Graphs

- 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
```

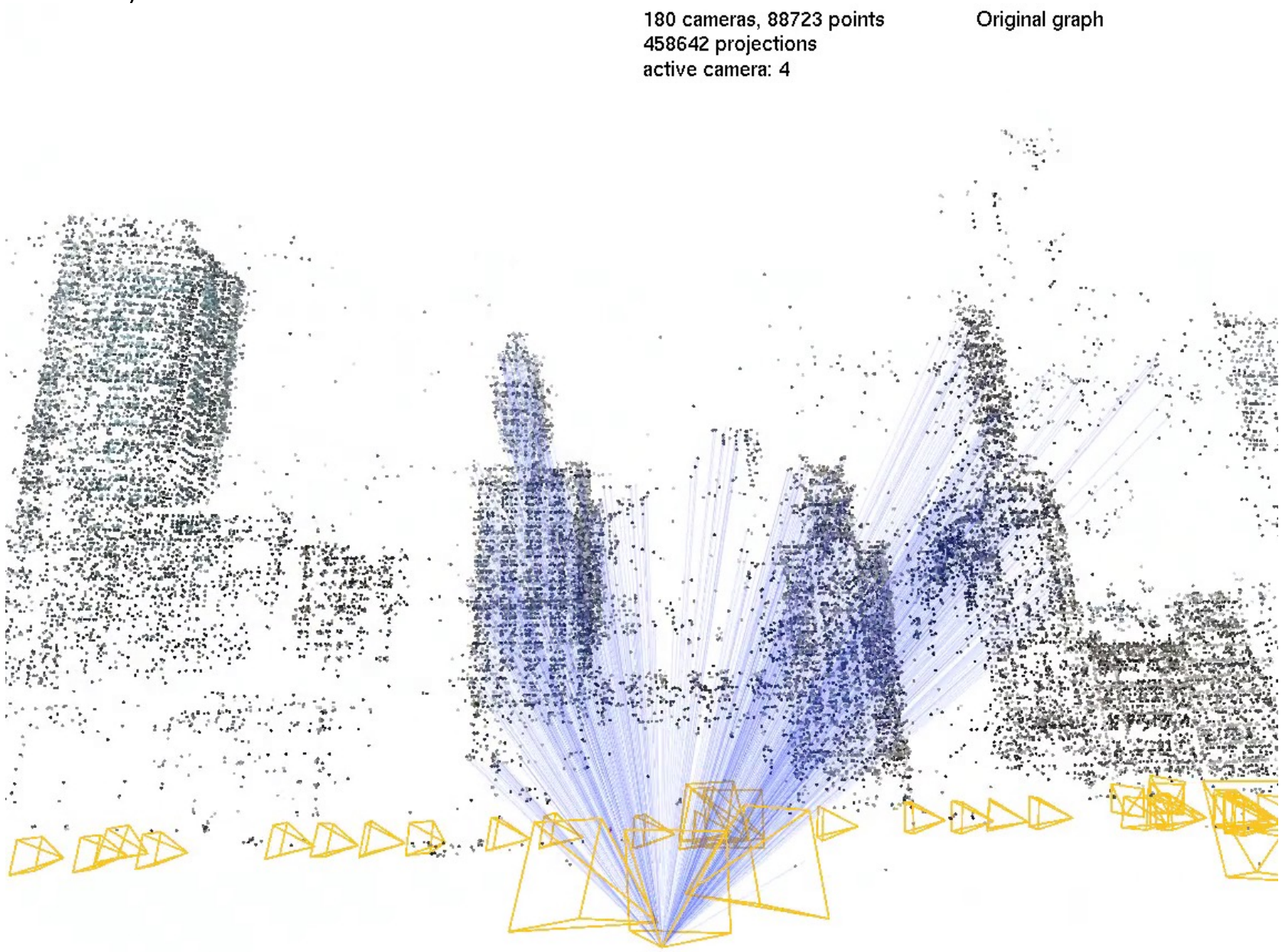
A'*A

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 =

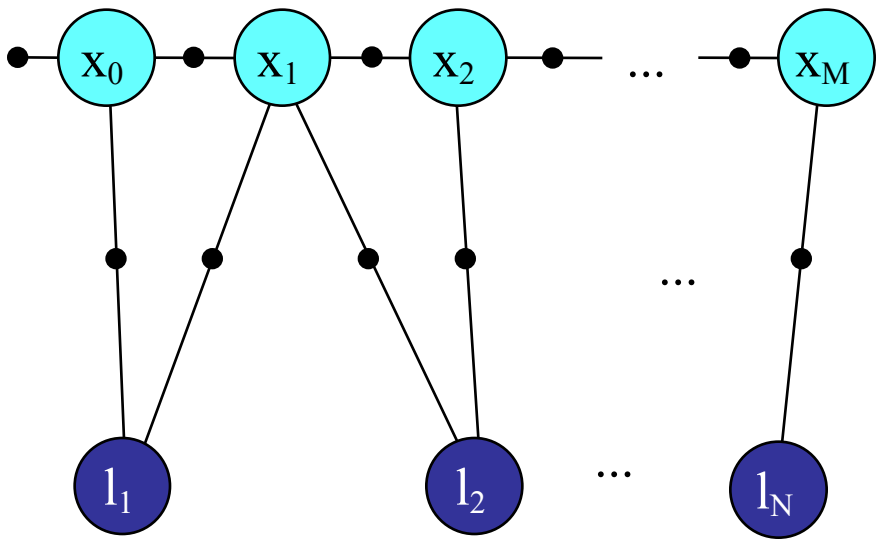
```
    5.0000
   15.0000
    0.0000
   10.0000
   15.0000
   20.0000
```

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



Visual SLAM: SfM for Robots

Visual SLAM Factor Graph



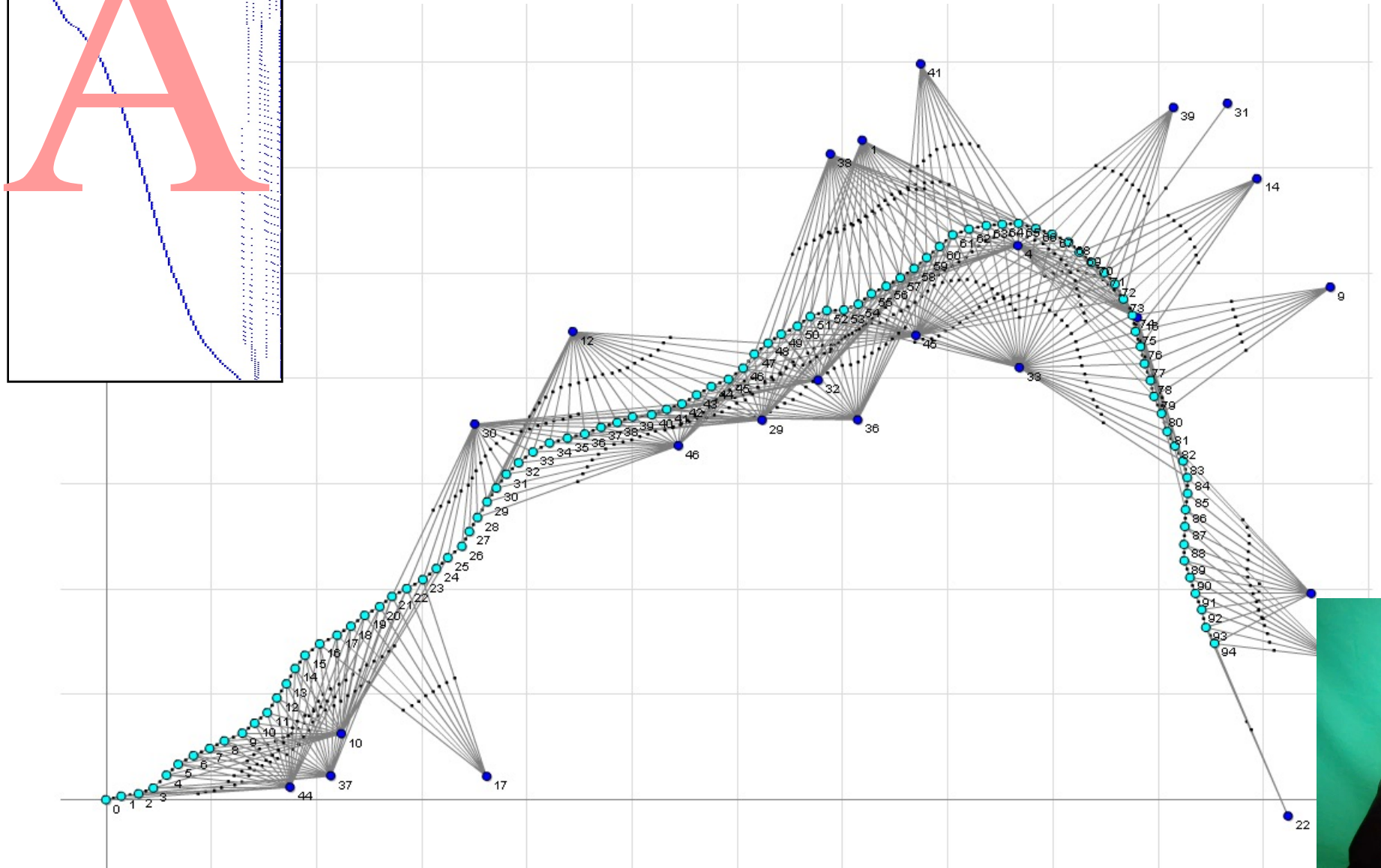
- Trajectory of Robot

- Landmark Measurements

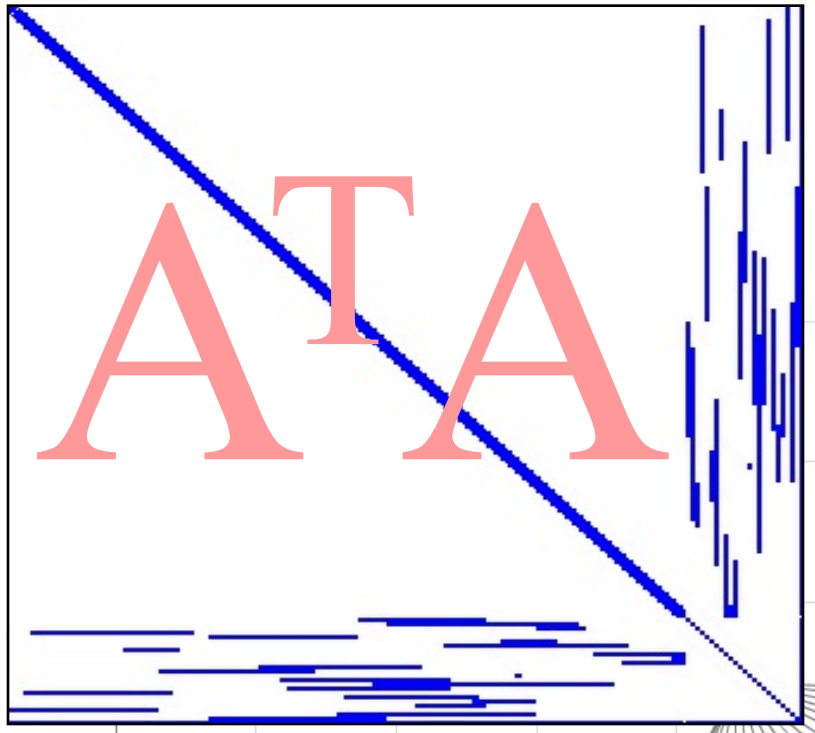
- "Landmarks"

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

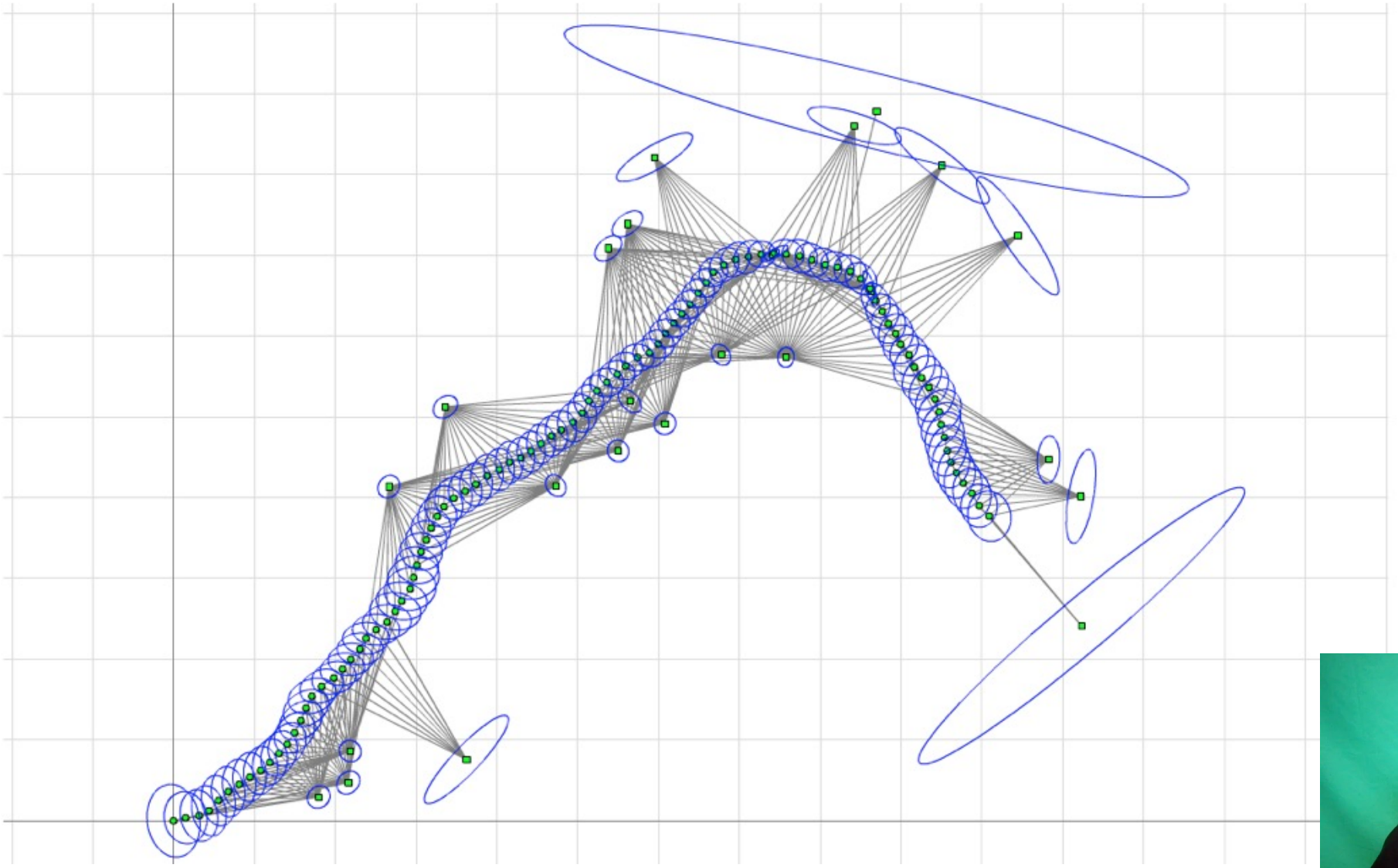
Visual SLAM Factor Graph



Hessian

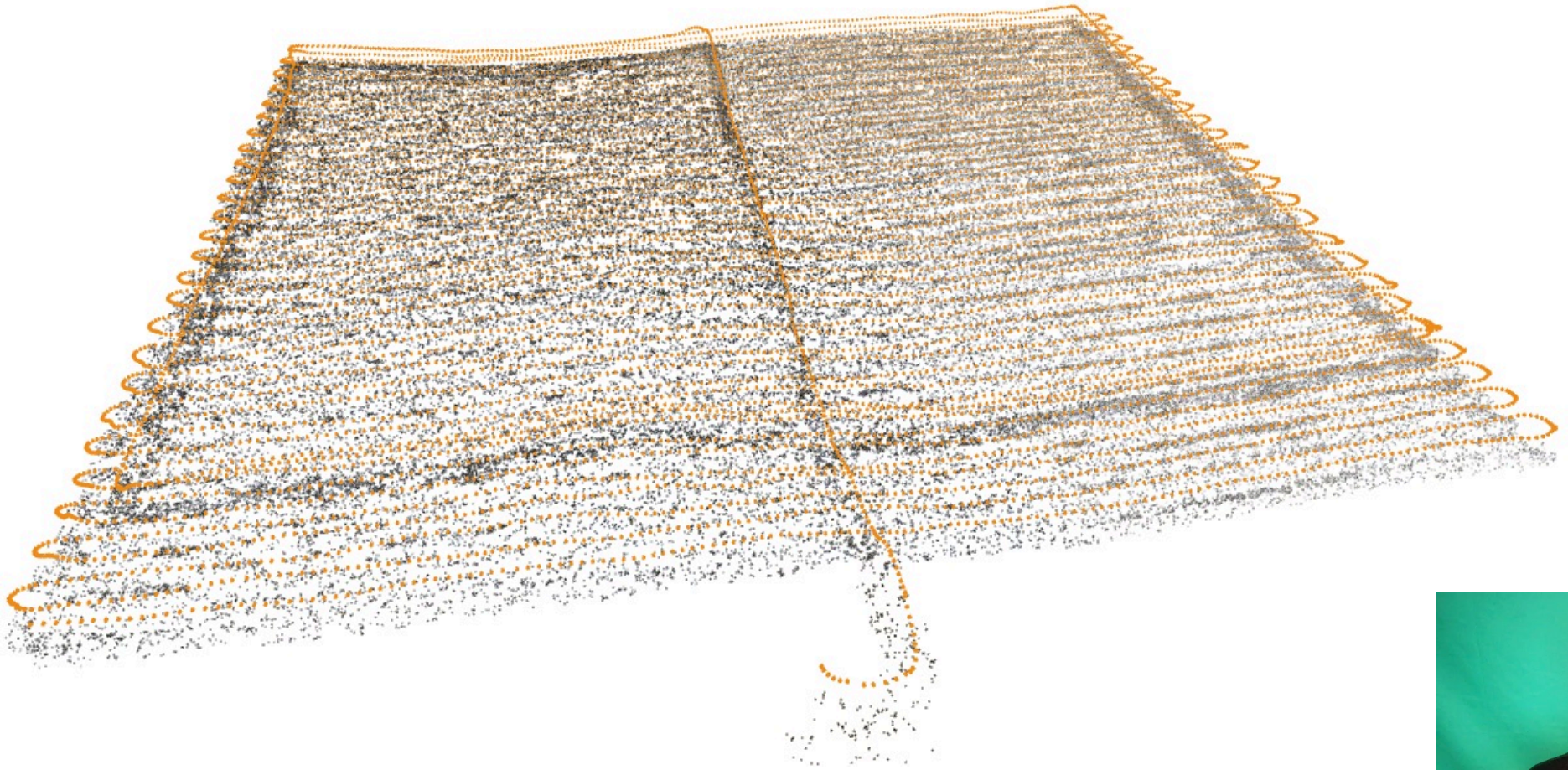


End result: Solution + Uncertainty

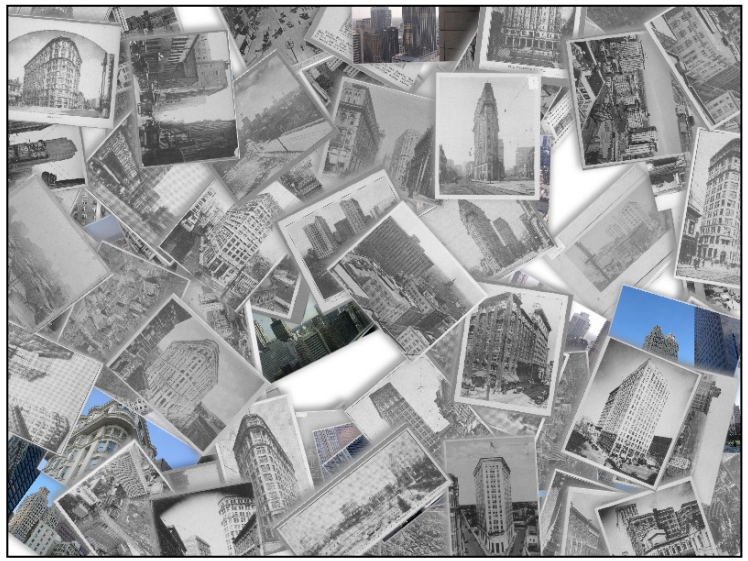


Example: Underwater SLAM

9831 camera poses, 185261 landmarks, and 350988 factors

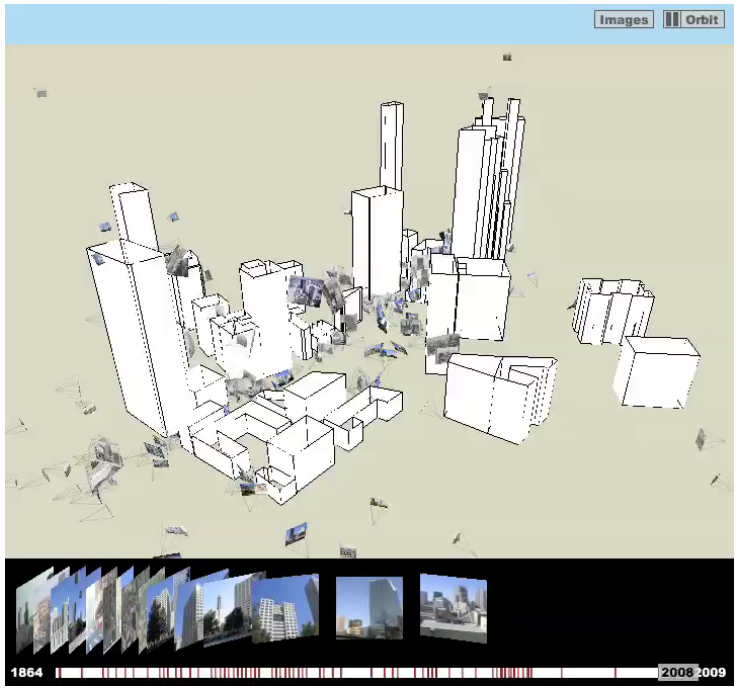


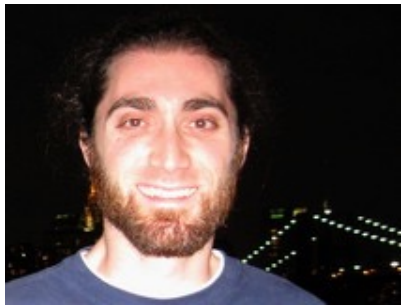
Spatiotemporal Reconstruction



Historical Image Collection

4D Cities: 3D + Time





Grant Schindler

Supported by NSF CAREER, Microsoft

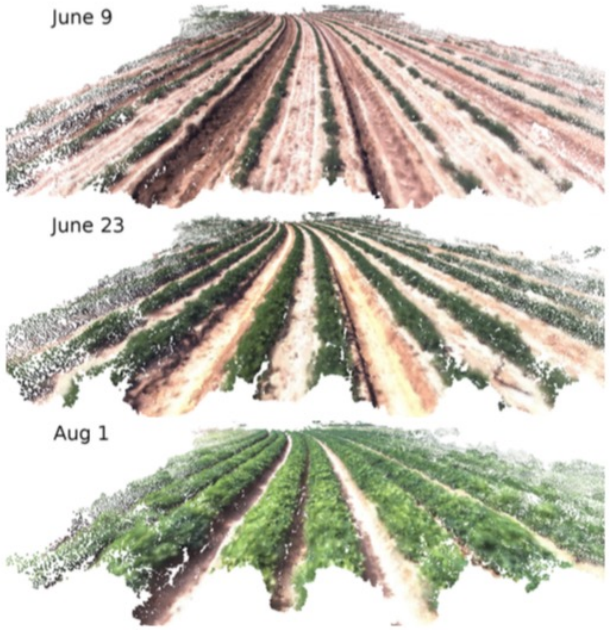
Recent revival: NSF NRI award on 4D crops for precision agriculture…

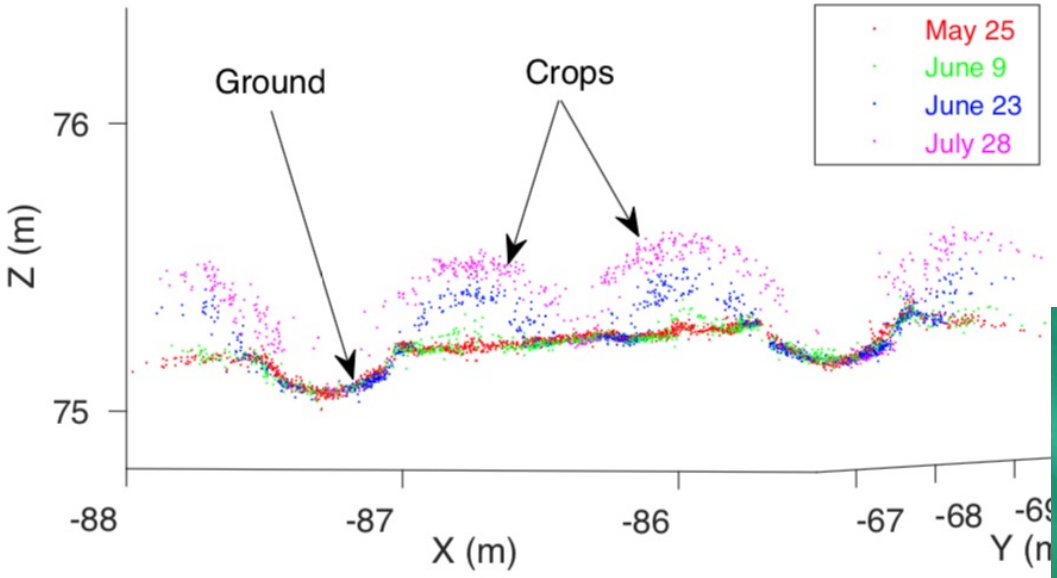
4D Reconstruction of Lower Manhattan

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

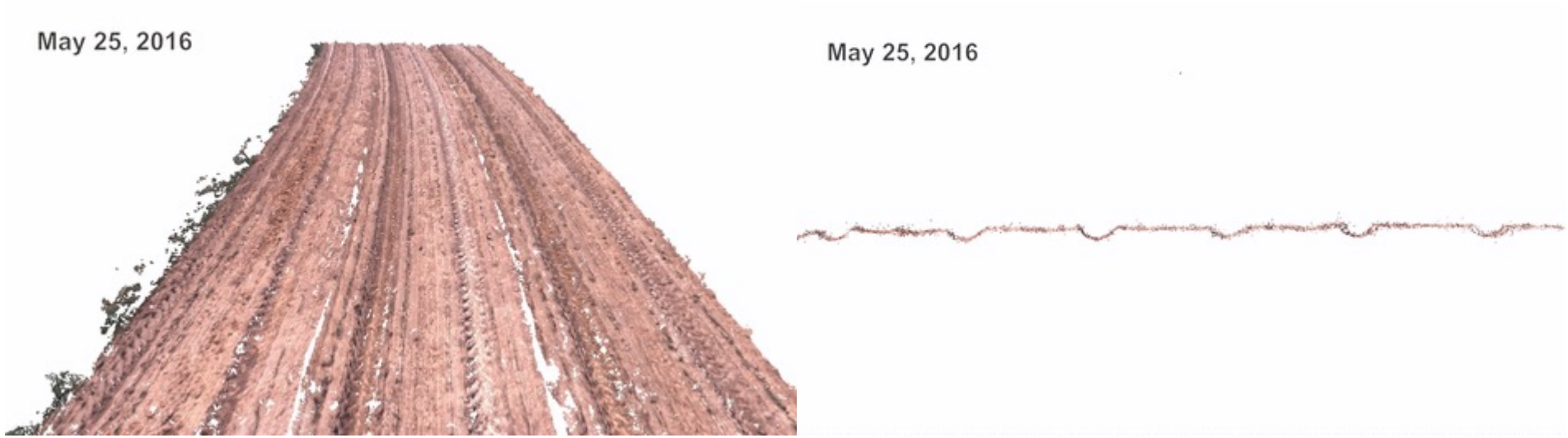
4D Structure over Time

4D crop monitoring (Jing Dong)





Results: video (by Jing Dong)



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