Image Based Reconstruction I

Qixing Huang

Jan. 31th 2017

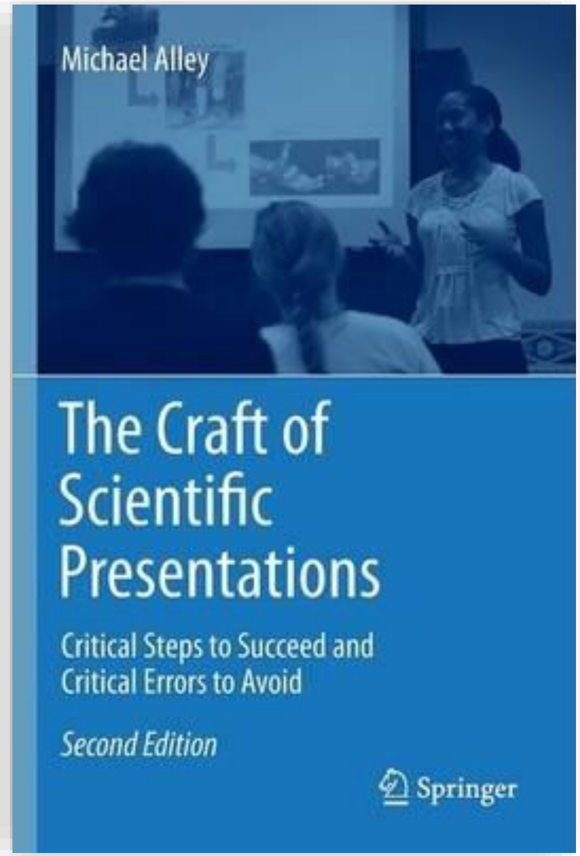
A Couple of Words on Paper Presentations

- Four components:
  - Motivation
  - Technical Merit
  - Results
  - Broader Impact

- Paper Strength/Weakness

- Read relevant papers as well

Making Presentations



Tools We Will Utilize

- Robust Norms

- MRF Inference

- Continuous Optimization
  - Newton method

Geometry Reconstruction from Images

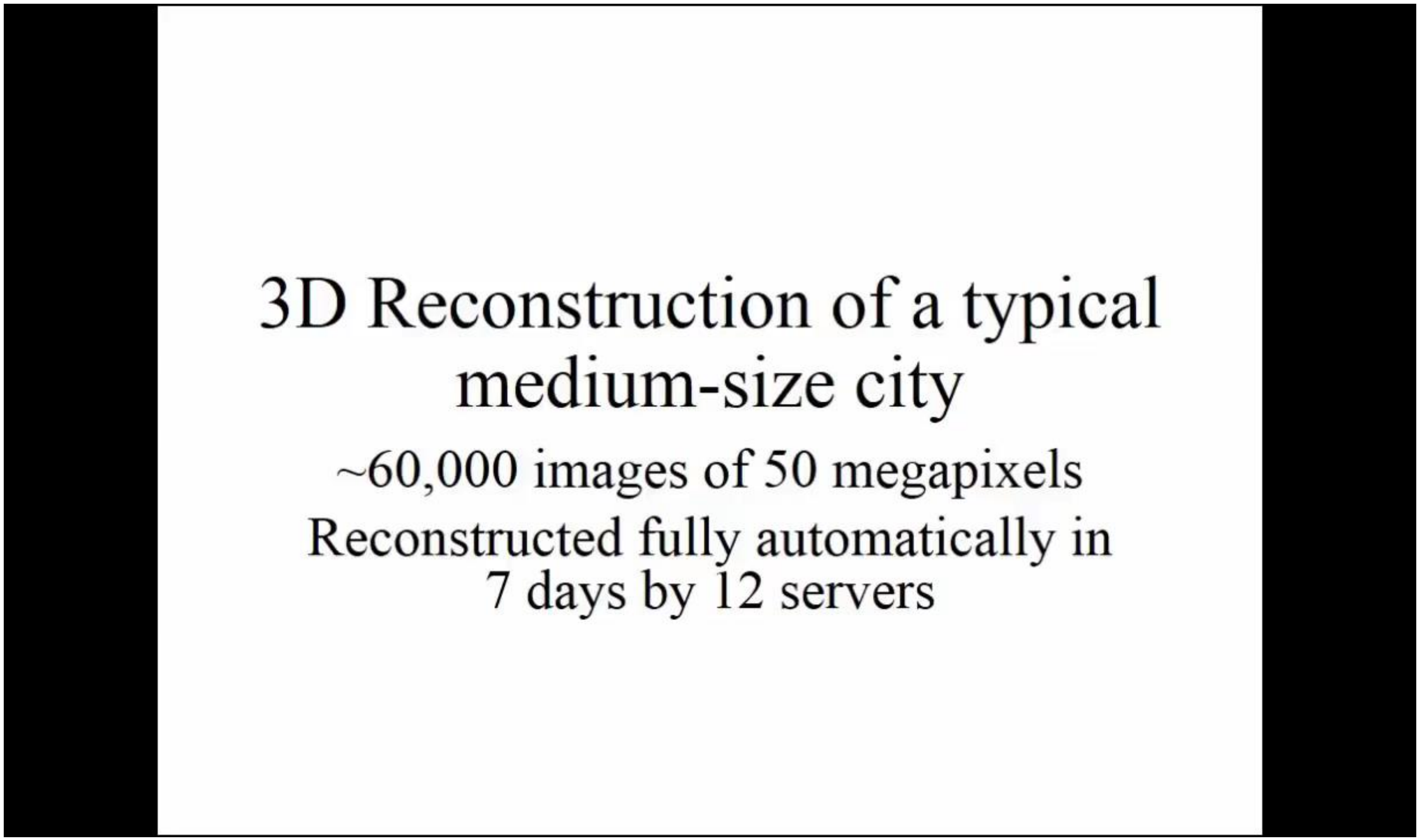
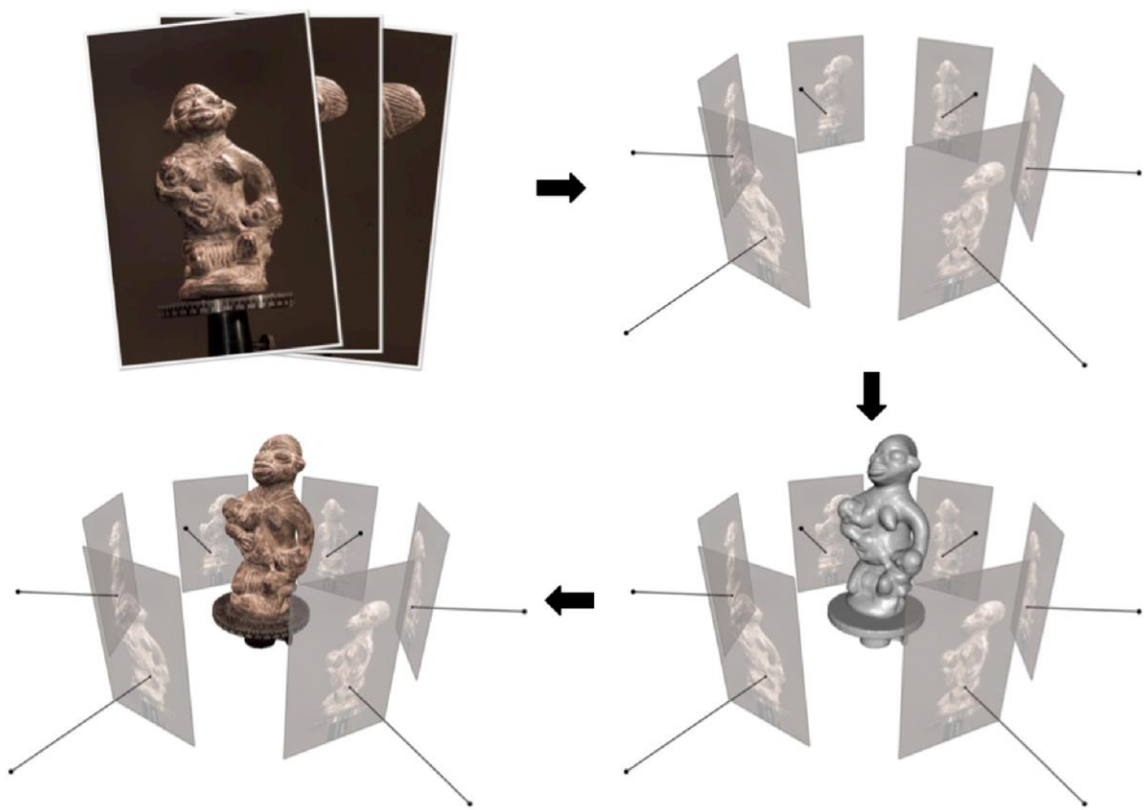
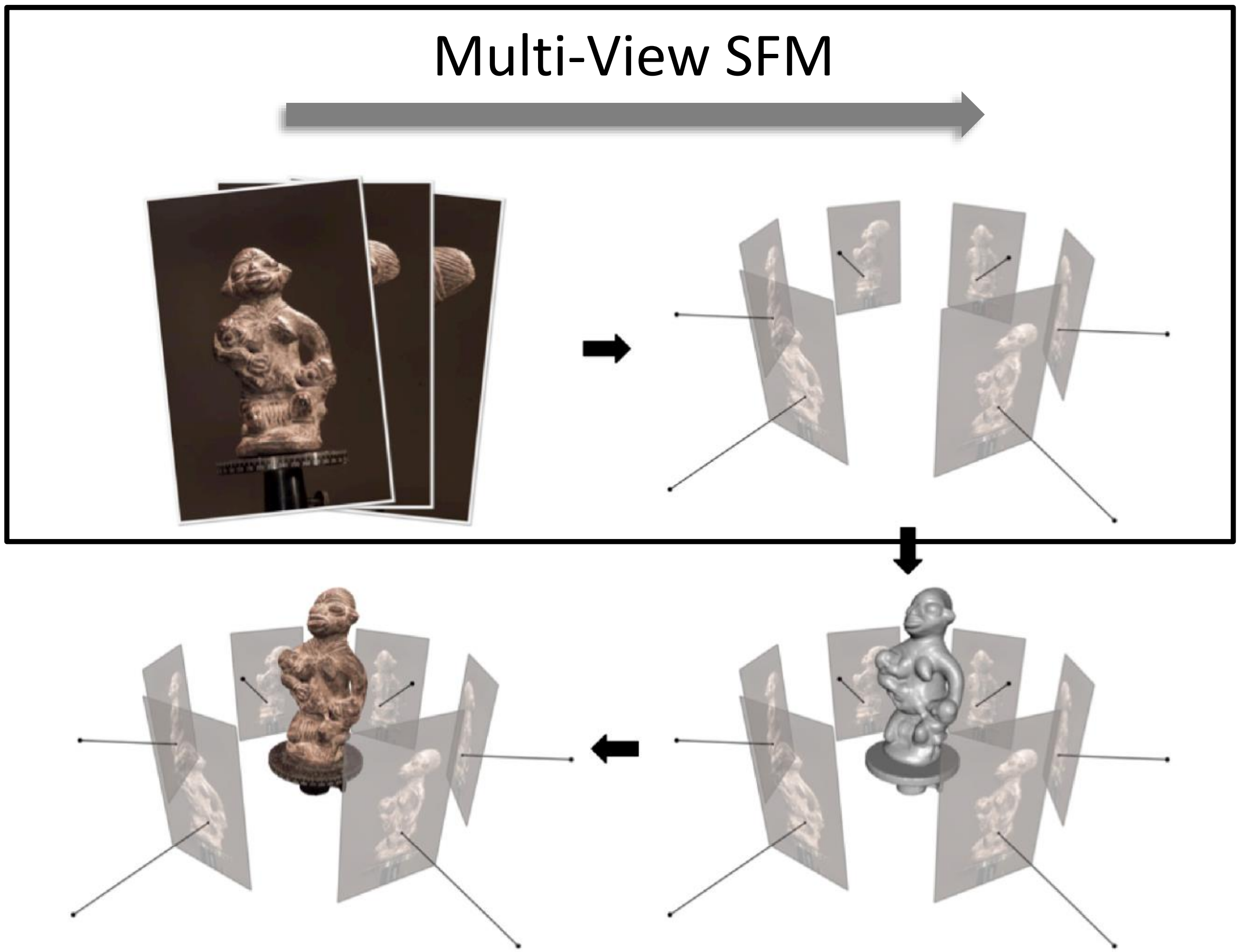


Image-Based Geometry Reconstruction Pipeline



This Lecture: Multi-View SFM

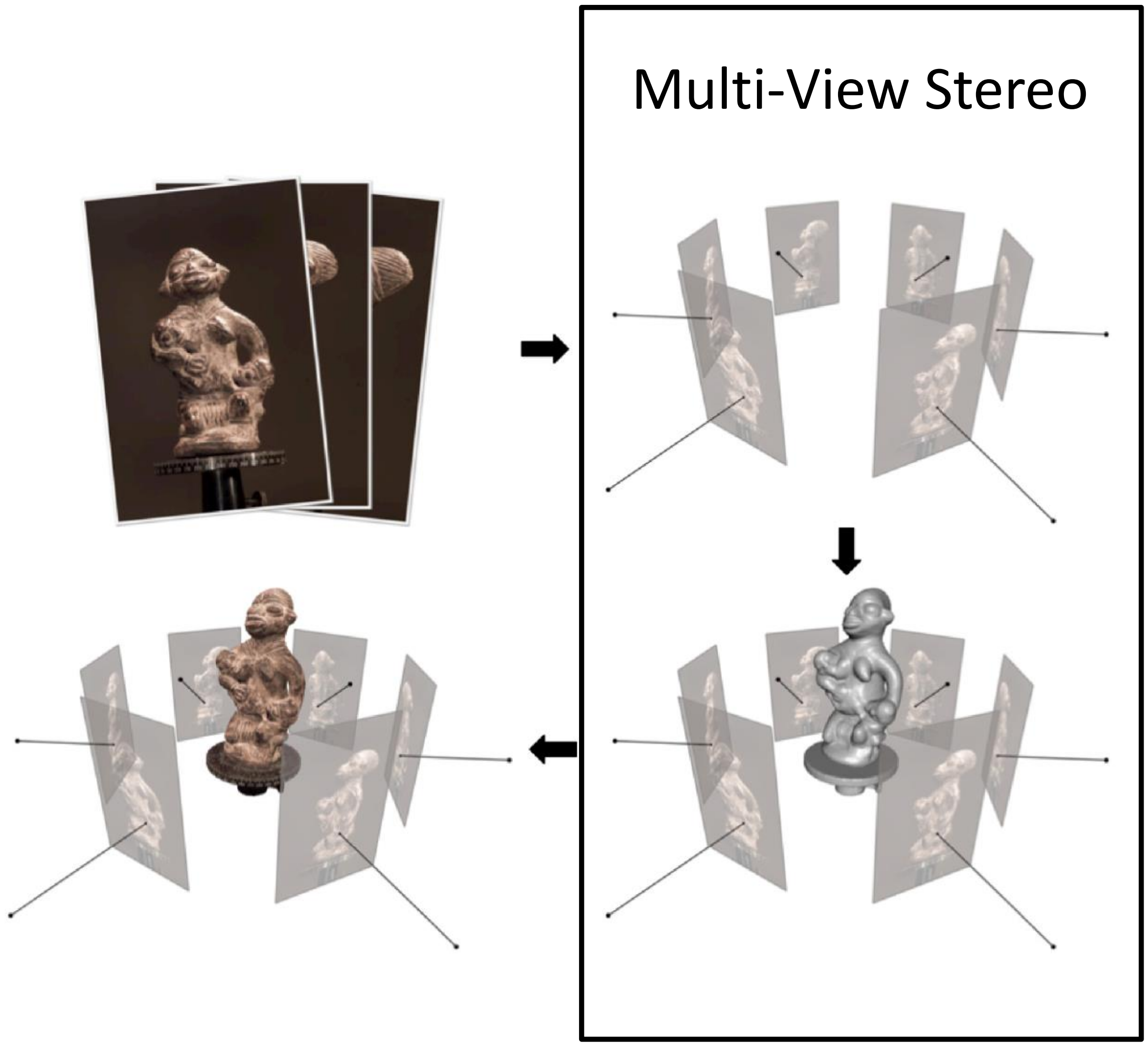


SFM Outputs Cameras + (Sparse) Point clouds

[Crandall et al. 13]

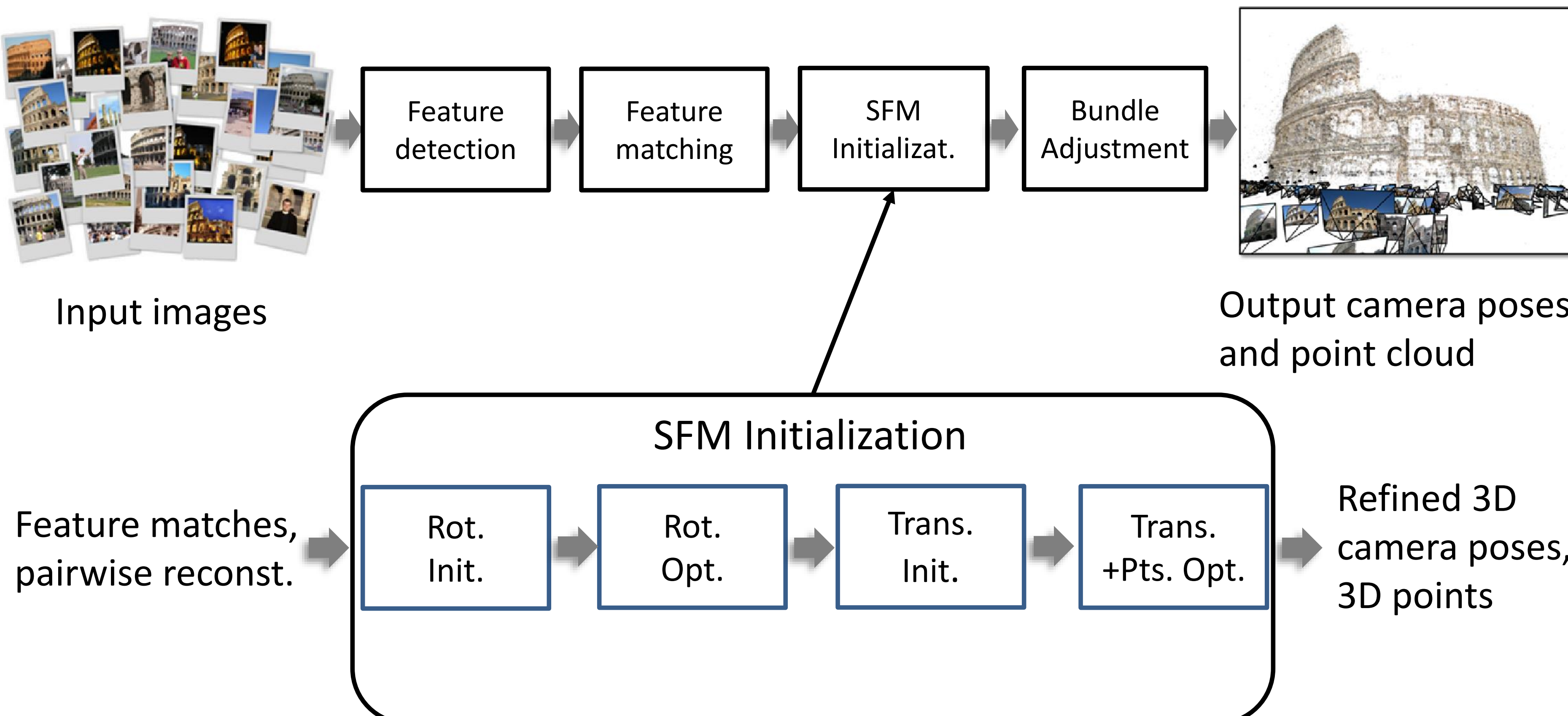
Quad
Reconstructed images: 5,233
Edges in MRF: 995,734

Next Lecture: Multi-View Stereo

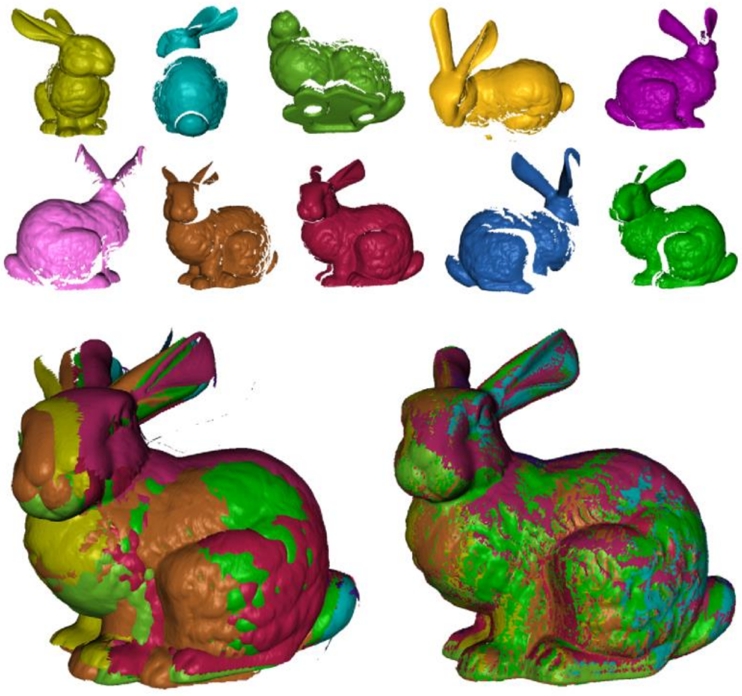


Multi-View SFM Pipeline

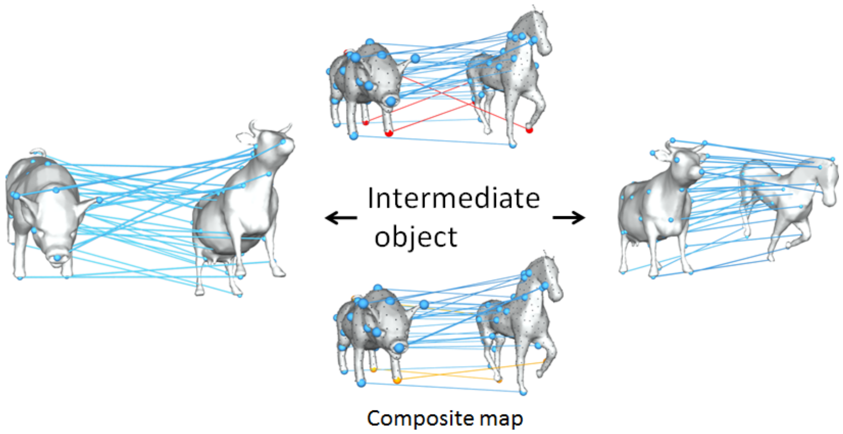
[Crandall et al. 13]



Similar Problems



Scan Alignment
[Gelfand et al. 05]



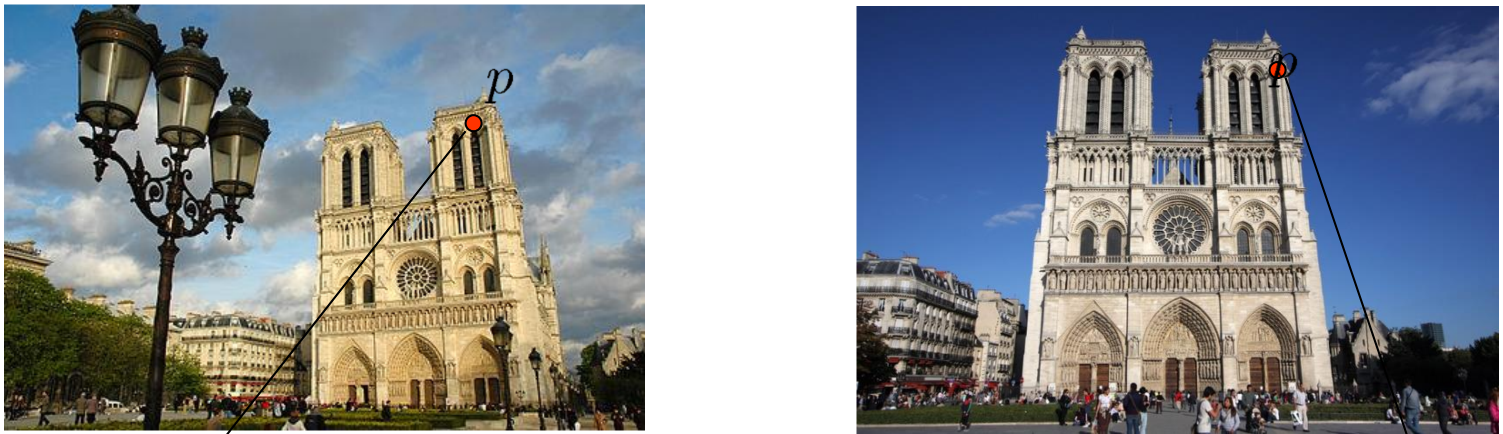
Data-Driven Map Computation
[Huang et al. 13]

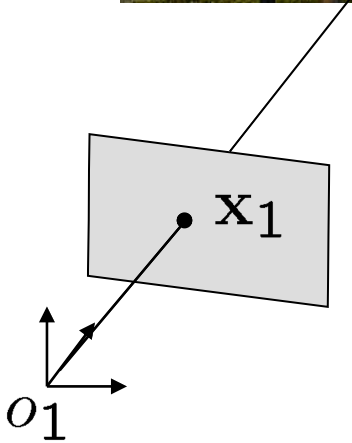
Image Features

SIFT Features [Lowe, IJCV 2004]

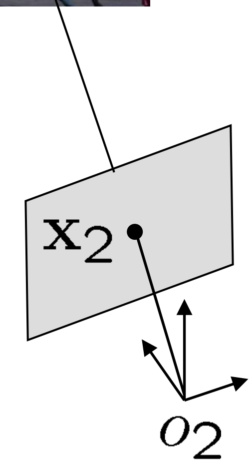
Pairwise Image Matching

Goal





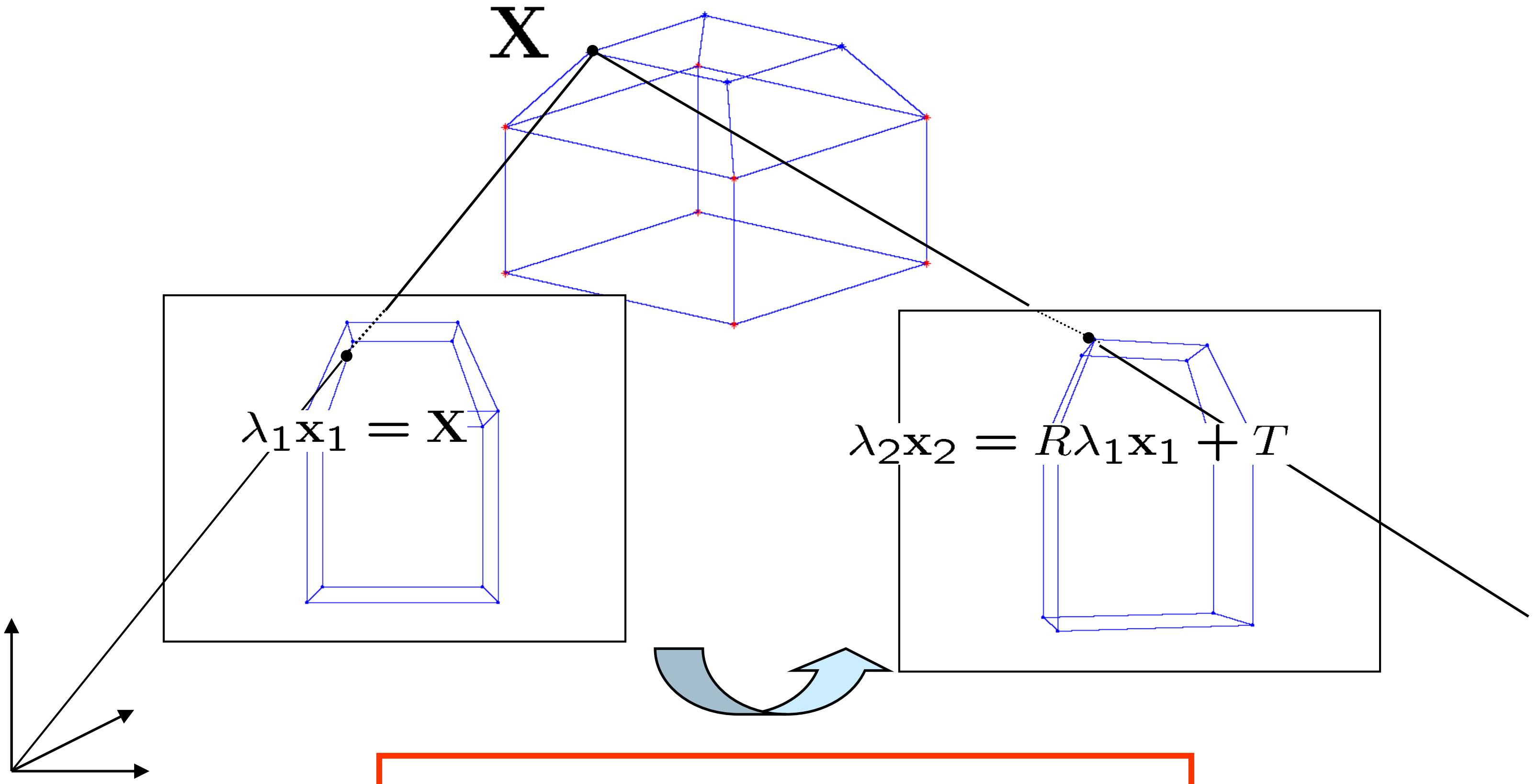
Given two views of the scene
recover the unknown
relative camera pose



Assume we known the intrinsic camera parameters

Five parameters to optimize

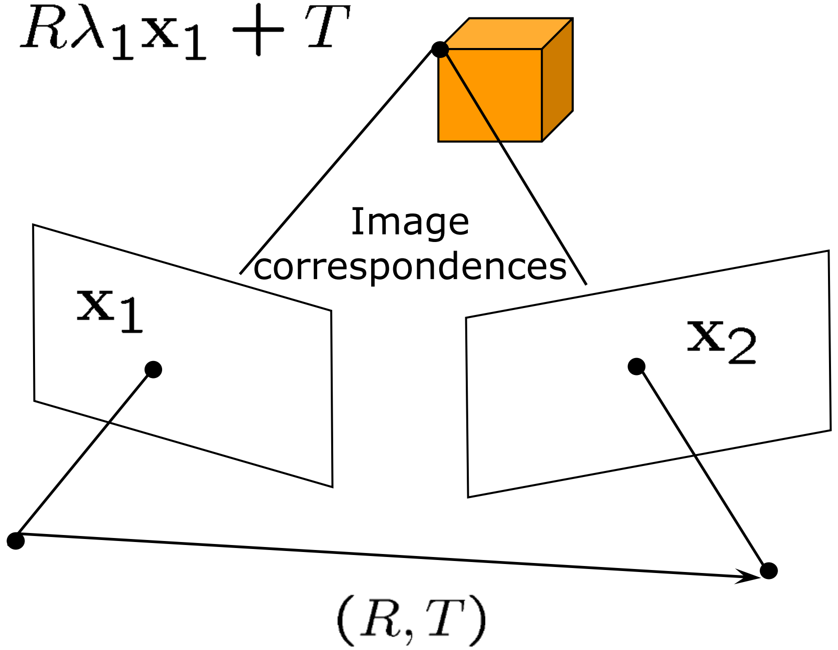
Rigid Body Motion --- Two Views



$$\boxed{\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T}$$

Epipolar Geometry

$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T$$



- Algebraic Elimination of Depth

[Longuet-Higgins '81]:

$$\mathbf{x}_2^T \underbrace{\hat{T}R}_{E} \mathbf{x}_1 = 0$$

- Essential matrix

$$E = \hat{T}R$$

Nister's Five-Point Method

$$\tilde{q}^\top \tilde{E} = 0$$

$$\tilde{q} \equiv \begin{bmatrix} q_1 q'_1 & q_2 q'_1 & q_3 q'_1 & q_1 q'_2 & q_2 q'_2 & q_3 q'_2 & q_1 q'_3 & q_2 q'_3 & q_3 q'_3 \end{bmatrix}^\top$$

$$\tilde{E} \equiv \begin{bmatrix} E_{11} & E_{12} & E_{13} & E_{21} & E_{22} & E_{23} & E_{31} & E_{32} & E_{33} \end{bmatrix}^\top$$

$$\downarrow$$

$$E = xX + yY + zZ + wW$$

$$\downarrow$$

$$EE^\top E - \frac{1}{2} trace(EE^\top)E = 0$$

$$\downarrow$$

$$E$$

$$\downarrow$$

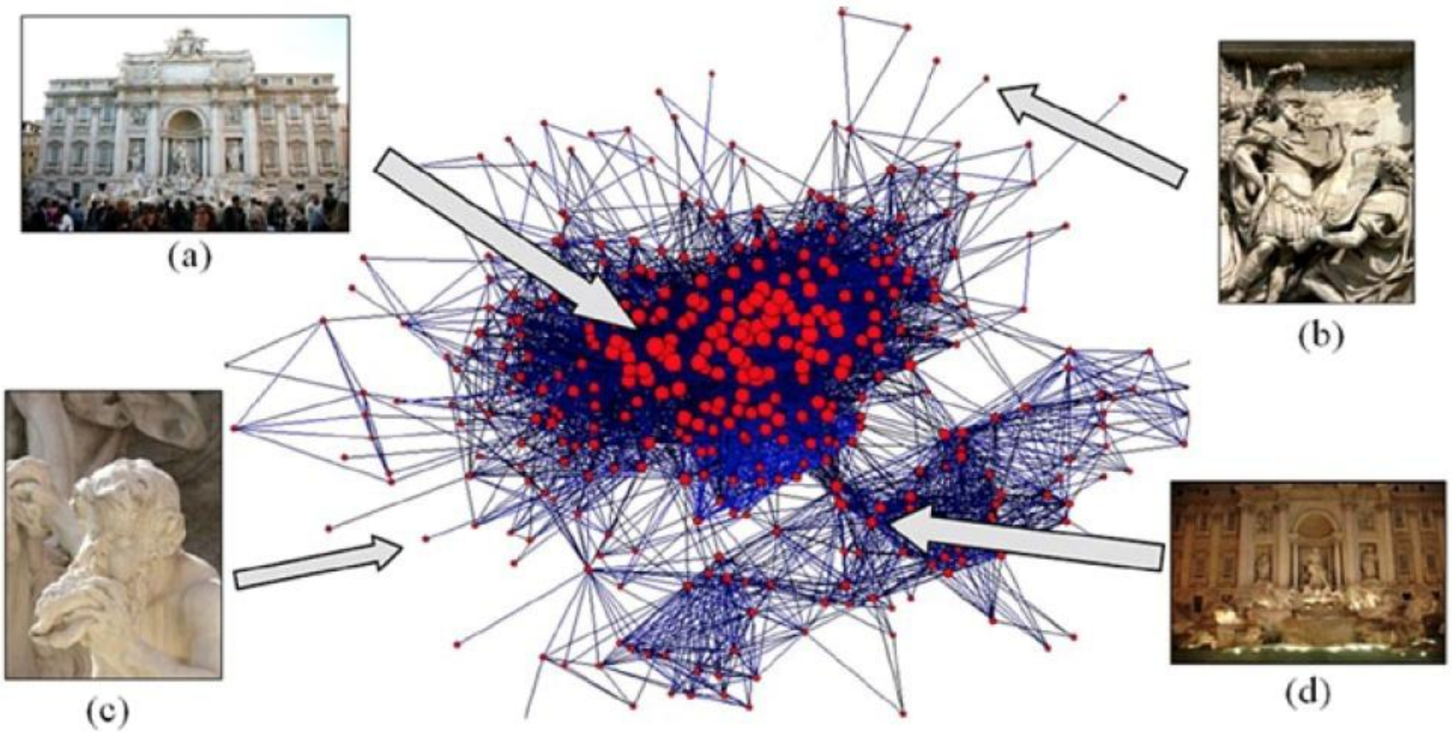
$$R \quad T$$

RANSAC [Fischler and Bolles' 81]

- Pick five feature matches
- Estimate the essential matrix and count the matched SIFT features
- Return the one with the most matched SIFT features

Which Pairs of Images to Match?

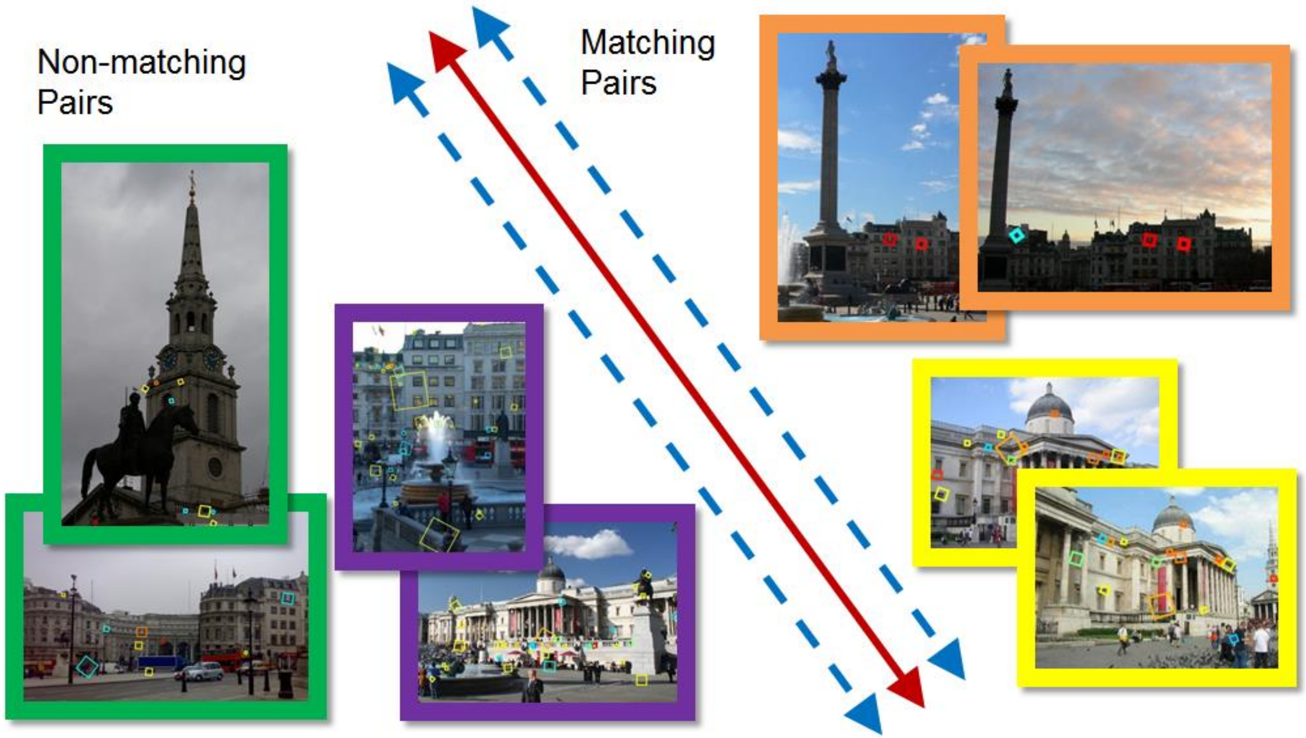
Nearest Neighbors in Image Descriptors (e.g., GIST and HOG)



Only works for images that significantly overlap

Train a Classifier to Differentiate Good/Bad Matches

[Cao and Snavely' 12]

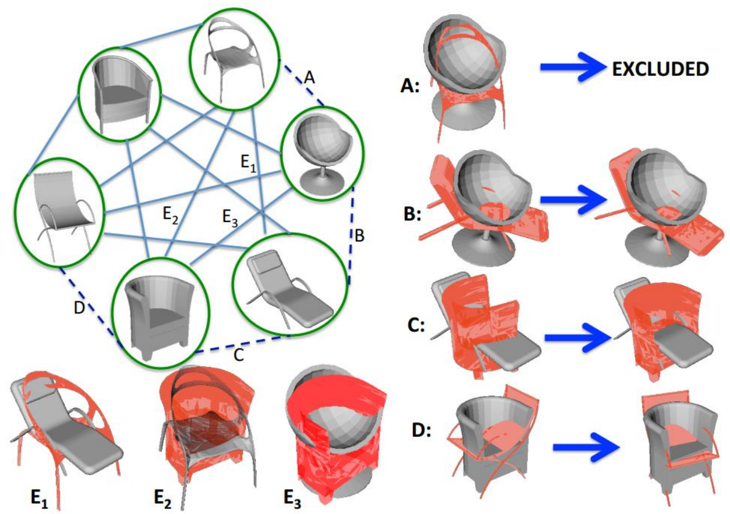


Iterative Matching/Learning

- Start from matching image descriptors
- Verification via image matching
- Train a classifier to find more potential image pairs and iterate

Graph Connectivity Optimization --- maximizing $\lambda_2(G)$

Imageweb [Heath et al 10]



Fuzzy correspondences on shapes [Kim et al 12]

Multi-View Pose Estimation

What We Have So Far

- A graph of images

- Along each edge
  - Noisy relative poses
  - Matched SIFT features

Three Approaches of Multi-View Pose Estimation (Goal: Remove Bad Matches)

- Combinatorial Optimization

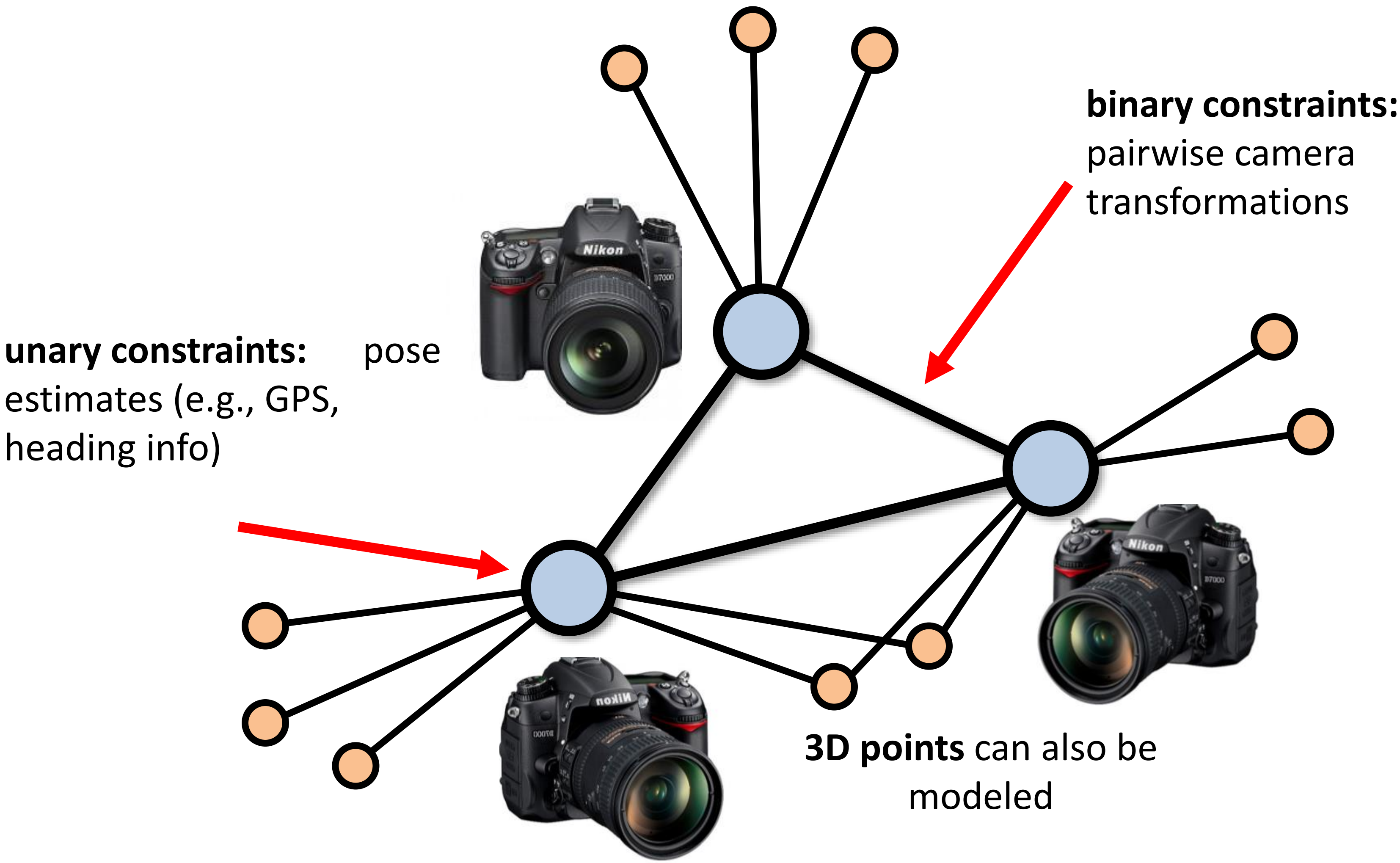
- Convex/Nonconvex Optimizations

- MRF-Based Formulation

[Crandall et al. 13]

The MRF model

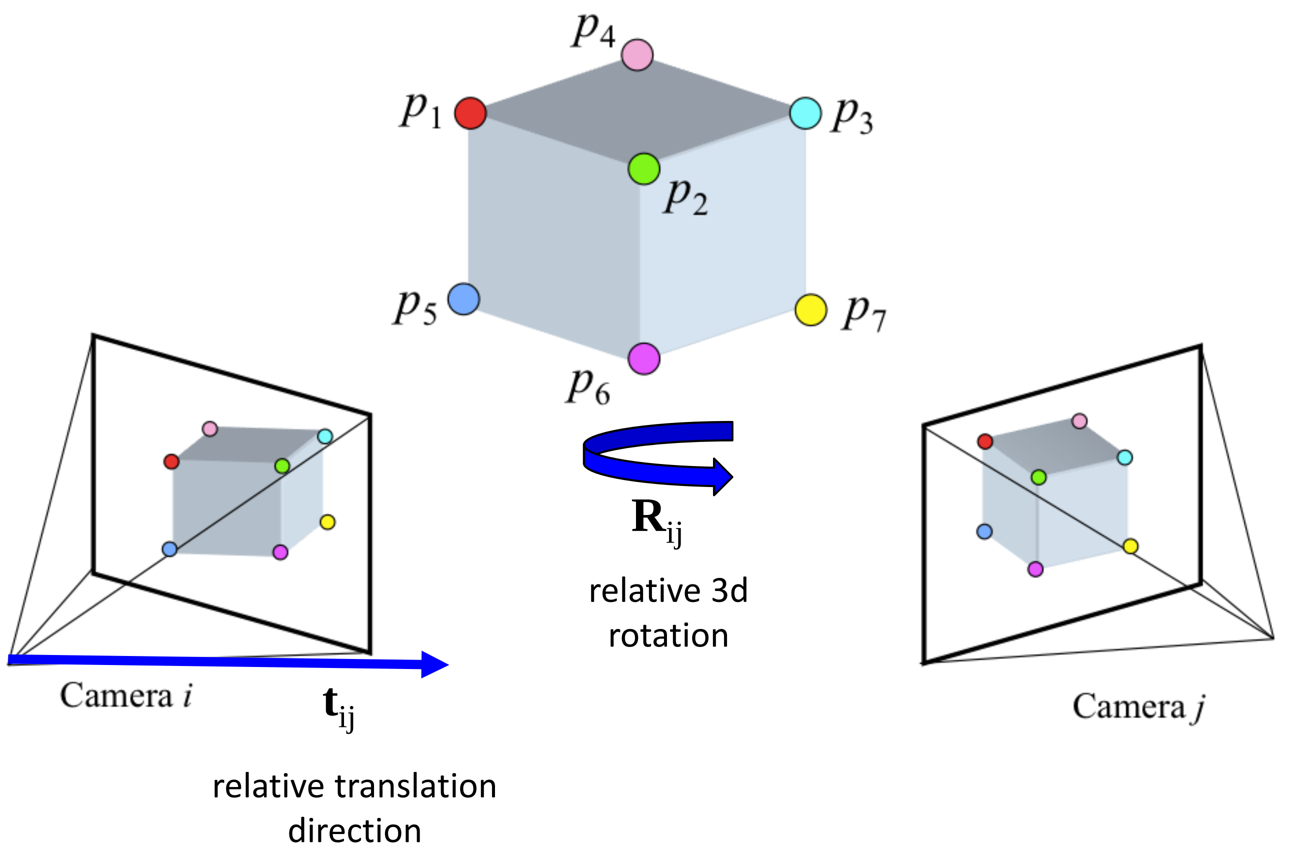
- Input: set of images with correspondence



[Crandall et al. 13]

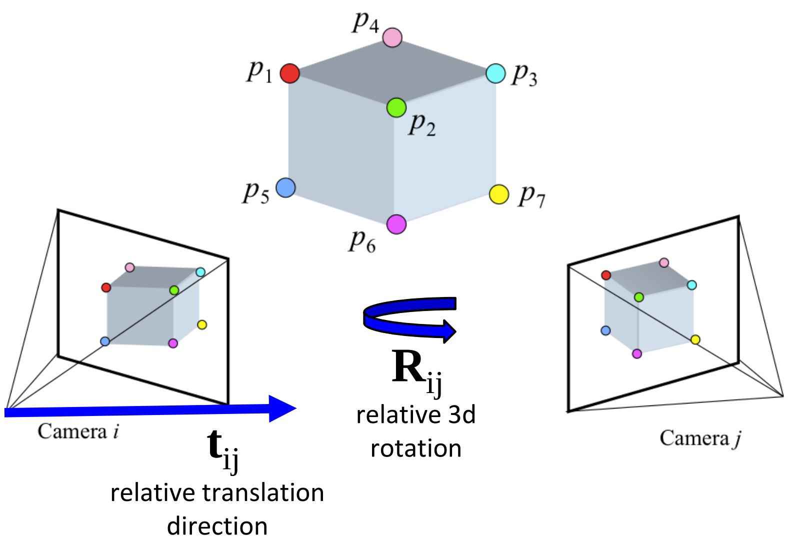
Constraints on camera pairs

- Compute relative pose between camera pairs using 2-frame SfM [Nister04]



[Crandall et al. 13]

Constraints on camera pairs



- Find absolute camera poses ($\mathbf{R}_i$, $\mathbf{t}_i$) and ($\mathbf{R}_j$, $\mathbf{t}_j$) that agree with these pairwise estimates:

$$\boxed{\mathbf{R}_{ij} \quad = \quad \mathbf{R}_i^\top \mathbf{R}_j}$$

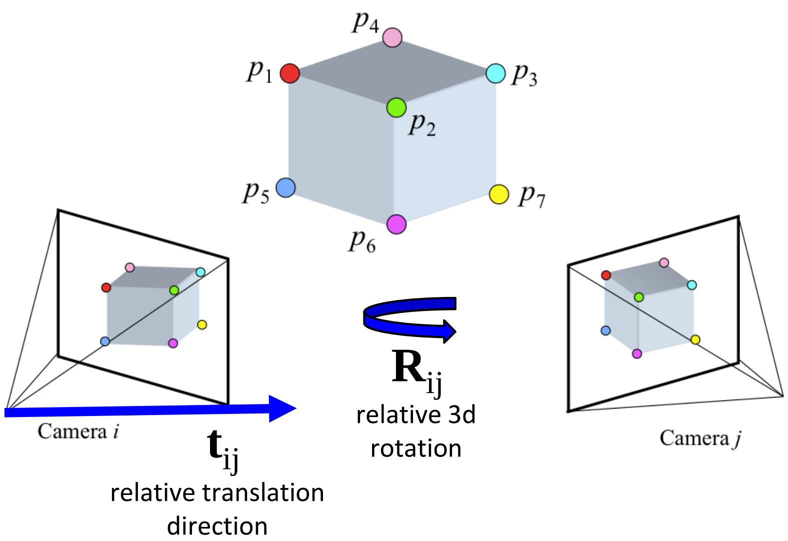
rotation consistency

$$\boxed{\lambda_{ij}\mathbf{t}_{ij} \quad = \quad \mathbf{R}_i^\top(\mathbf{t}_j - \mathbf{t}_i)}$$

translation direction consistency

[Crandall et al. 13]

Constraints on camera pairs



- Define robustified error functions to use as pairwise potentials:

$$d^{\mathbf{R}}(\mathbf{R}_{ij}, \mathbf{R}_i^\top \mathbf{R}_j)$$

$$d^{\mathbf{R}}(\mathbf{R}_a, \mathbf{R}_b) = \rho_R(||\mathbf{R}_a - \mathbf{R}_b||)$$

rotation consistency

$$d^{\mathbf{T}}(\mathbf{t}_j - \mathbf{t}_i, \mathbf{R}_i \mathbf{t}_{ij})$$

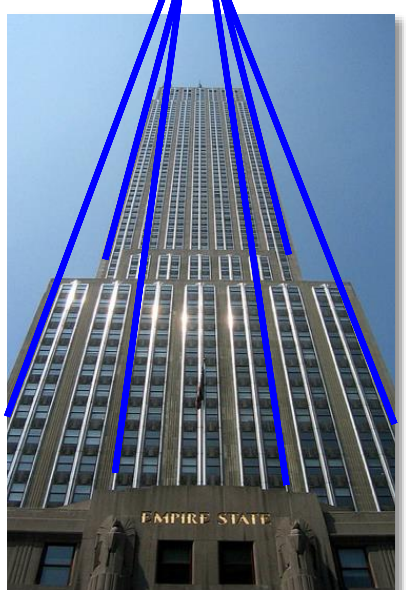
$$d^{\mathbf{T}}(\mathbf{t}_a, \mathbf{t}_b) = \rho(\mathrm{angleof}(\mathbf{t}_a, \mathbf{t}_b))$$

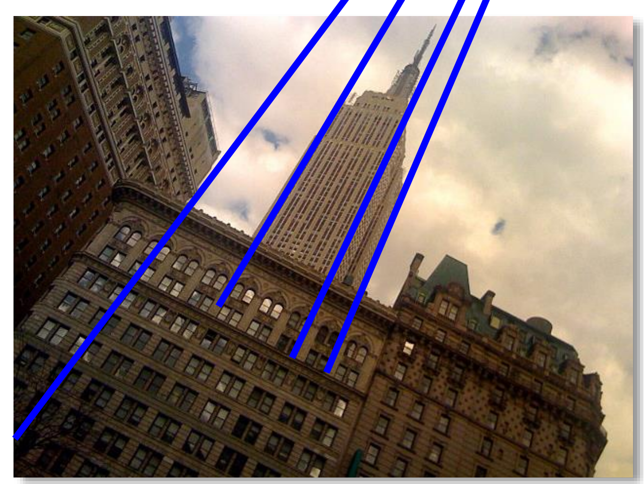
translation direction consistency

[Crandall et al. 13]

Prior pose information

- *Noisy* absolute pose info for some cameras
  - 2D positions from geotags (GPS coordinates)
  - Orientations (tilt & twist angles) from vanishing point detection [Sinha10]





Overall optimization problem

- Given pairwise and unary pose constraints, solve for absolute camera poses simultaneously
  - for *n* cameras, estimate

    $\mathcal{R} = (\mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_n})$ and $\mathcal{T} = (\mathbf{t_1}, \mathbf{t_2}, \ldots, \mathbf{t_n})$

    so as to minimize total error over the entire graph

$$D^{\mathbf{R}}(\mathcal{R}) = \sum_{e_{ij} \in E_C} \underbrace{\boxed{d^{\mathbf{R}}\left(\mathbf{R}_{ij}, \mathbf{R}_i^\top \mathbf{R}_j\right)}}_{\text{pairwise rotation consistency}} + \alpha_1 \sum_{I_i \in \mathcal{I}} \underbrace{\boxed{d_i^{\mathbf{O}}(\mathbf{R}_i)}}_{\text{unary rotation consistency}}$$

$$D^{\mathbf{T}}(\mathcal{T}, \mathcal{R}) = \sum_{e_{ij} \in E_C} \underbrace{\boxed{d^{\mathbf{T}}(\mathbf{t}_j - \mathbf{t}_i, \mathbf{R}_i \mathbf{t}_{ij})}}_{\text{pairwise translation consistency}} + \alpha_2 \sum_{I_i \in \mathcal{I}} \underbrace{\boxed{d_i^{\mathbf{G}}(\mathbf{t}_i)}}_{\text{unary translation consistency}}$$

MRF Inference

- Convert continuous optimization into a labeling problem:

$$E(\mathbf{x}) = \sum_i f_i(x_i) + \sum_{(i,j)\in\mathcal{E}} f_{ij}(x_i, x_j)$$

- A well studied problem with efficient solvers

Incorporate Points

- Point tracks --- interest points across multiple images that have similar SIFT descriptors
- Select point tracks that cover each camera-camera edge five times and each image ten times
- Relation between 3D location and image coordinates:

$$\mu_{ik}\mathbf{x}_{ik} = \mathbf{K}_i\mathbf{R}_i(\mathbf{X}_k - \mathbf{t}_i)$$

Intrinsic camera parameters

[Crandall et al. 13]

Solving the MRF

- Use discrete loopy belief propagation [Pearl88]
  - Up to 1,000,000 nodes (cameras and points)
  - Up to 5,000,000 edges (constraints between cameras and points)
  - 6-dimensional label space for cameras (3-dimensional for points)

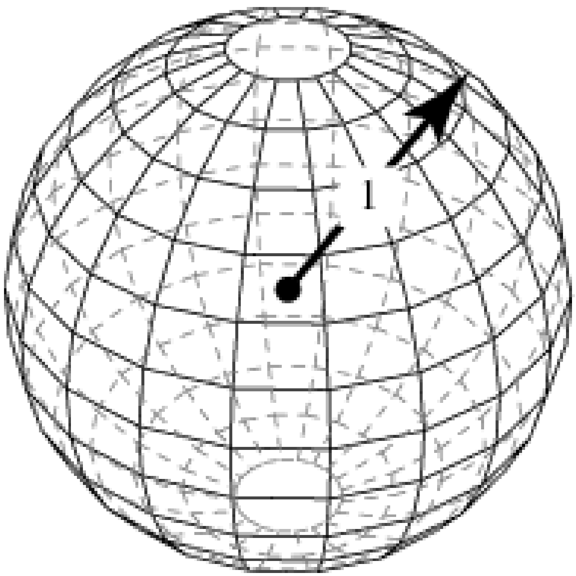
[Crandall et al. 13]

Solving the MRF

- Reduce 6-dimensional label space by…
  - Solving for rotations & translations independently [Martinec07], [Sim06], [Sinha08]
  - Assuming camera twist angles are near 0
  - Initially solving for 2D camera positions
- Speed up BP by…
  - Using a parallel implementation on a cluster
  - Using distance transforms (aka min convolutions) to compute BP messages in O(L) time in # of labels (instead of $O(L^2)$) [Felzenszwalb04]

[Crandall et al. 13]

Discrete BP: Rotations

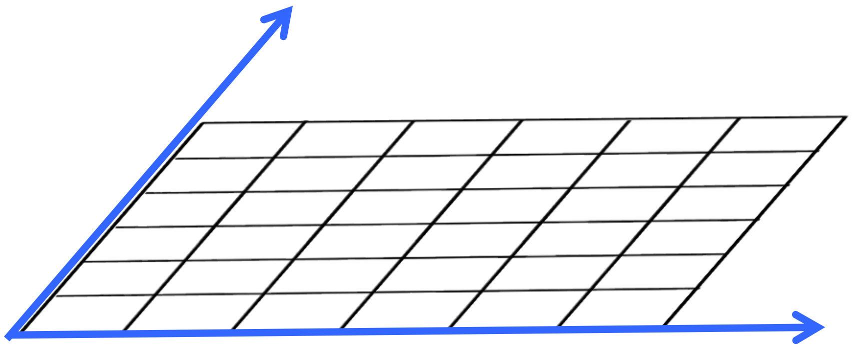


- Parameterize viewing directions as points on unit sphere
  - Discretize into 10x10x10 = 1,000 possible labels
  - Measure rotational errors as robust Euclidean distances on sphere (to allow use of distance transform)

[Crandall et al. 13]

Discrete BP: Translations

- Parameterize positions as 2D points in plane
  - Use approximation to error function
    (to allow use of distance transforms)
  - Discretize into up to 300 x 300 = 90,000 labels



Bundle Adjustment

Rotation Optimization

$$D^{\mathbf{R}}(\mathcal{R}) = \sum_{e_{ij} \in E_C} d^{\mathbf{R}}\left(\mathbf{R}_{ij}, \mathbf{R}_i^\top \mathbf{R}_j\right)$$

x²

$$d^{\mathbf{R}}(\mathbf{R}_a, \mathbf{R}_b) = \rho_R(||\mathbf{R}_a - \mathbf{R}_b||)$$

Using quadratic loss after removing outlier rotations

Fix one image (or use geotags which provide pose priors)

Nonlinear Least Squares

↑

Gauss-Newton Method

Gauss-Newton Method

- The Gauss–Newton algorithm is a method used to solve non-linear least squares problems

$$f(x) \equiv \frac{1}{2} \triangle z(x)^\top W \triangle z(x)$$

$$g \equiv \frac{df}{dx} = \triangle z^\top W J \qquad H \equiv \frac{d^2 f}{dx^2} = J^\top W J + \sum_i (\triangle z^\top W)_i \frac{d^2 z_i}{dx^2}$$

$$H \approx J^\top W J. \qquad \frac{d^2 z_i}{dx^2} \approx 0$$

$$(J^\top W J)\, \delta x = -J^\top W \triangle z$$

Linear Convergence for Practical Problems

Levenberg – Marquardt Heuristic

- The LMA interpolates between the Gauss–Newton algorithm (GNA) and the method of gradient descent

- When far from the minimum it acts as a steepest descent and it performs gauss newton iteration when near to the solution

$$(\mathsf{H} + \lambda \mathsf{W})\, \delta \mathsf{x} \; = \; -\mathsf{g}$$

Translation and Point Optimization

Camera-Camera Relation: $\lambda_{ij}\mathbf{t}_{ij} = \mathbf{R}_i^\top(\mathbf{t}_j - \mathbf{t}_i)$ $\qquad \hat{\mathbf{t}}_{ij} = \mathbf{R}_i\mathbf{t}_{ij}$

Camera-Point Relation: $\mu_{ik}\mathbf{x}_{ik} = \mathbf{K}_i\mathbf{R}_i(\mathbf{X}_k - \mathbf{t}_i)$ $\qquad \hat{\mathbf{x}}_{ik} = \mathbf{R}_i^\top\mathbf{K}_i^{-1}\mathbf{x}_{ik}$

$$D^{\mathbf{T}}(\mathcal{T}) = \alpha_2 \sum_{e_{ij}\in E_C} d^{\mathbf{T}}(\mathbf{t}_j - \mathbf{t}_i, \hat{\mathbf{t}}_{ij}) + d^{\mathbf{T}}(\mathbf{t}_i - \mathbf{t}_j, \hat{\mathbf{t}}_{ji}) \; +$$

$$\alpha_3 \sum_{e_{ik}\in E_F} d^{\mathbf{T}}(\mathbf{X}_k - \mathbf{t}_i, \hat{\mathbf{x}}_{ik})$$

$$d^{\mathbf{T}}(\mathbf{v}_a, \mathbf{v}_b) = \rho(\mathrm{angleof}(\mathbf{v}_a, \mathbf{v}_b))$$

x^2

Positional variables are decoupled!

Schur Trick

$$\begin{array}{cc} & \begin{array}{cc} \text{Trans.} & \text{Poss.} \end{array} \\ \begin{array}{c} \text{Trans.} \\ \text{Poss.} \end{array} & \begin{bmatrix} B & E \\ E^\top & C \end{bmatrix} \end{array} \begin{bmatrix} \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} v \\ w \end{bmatrix}$$

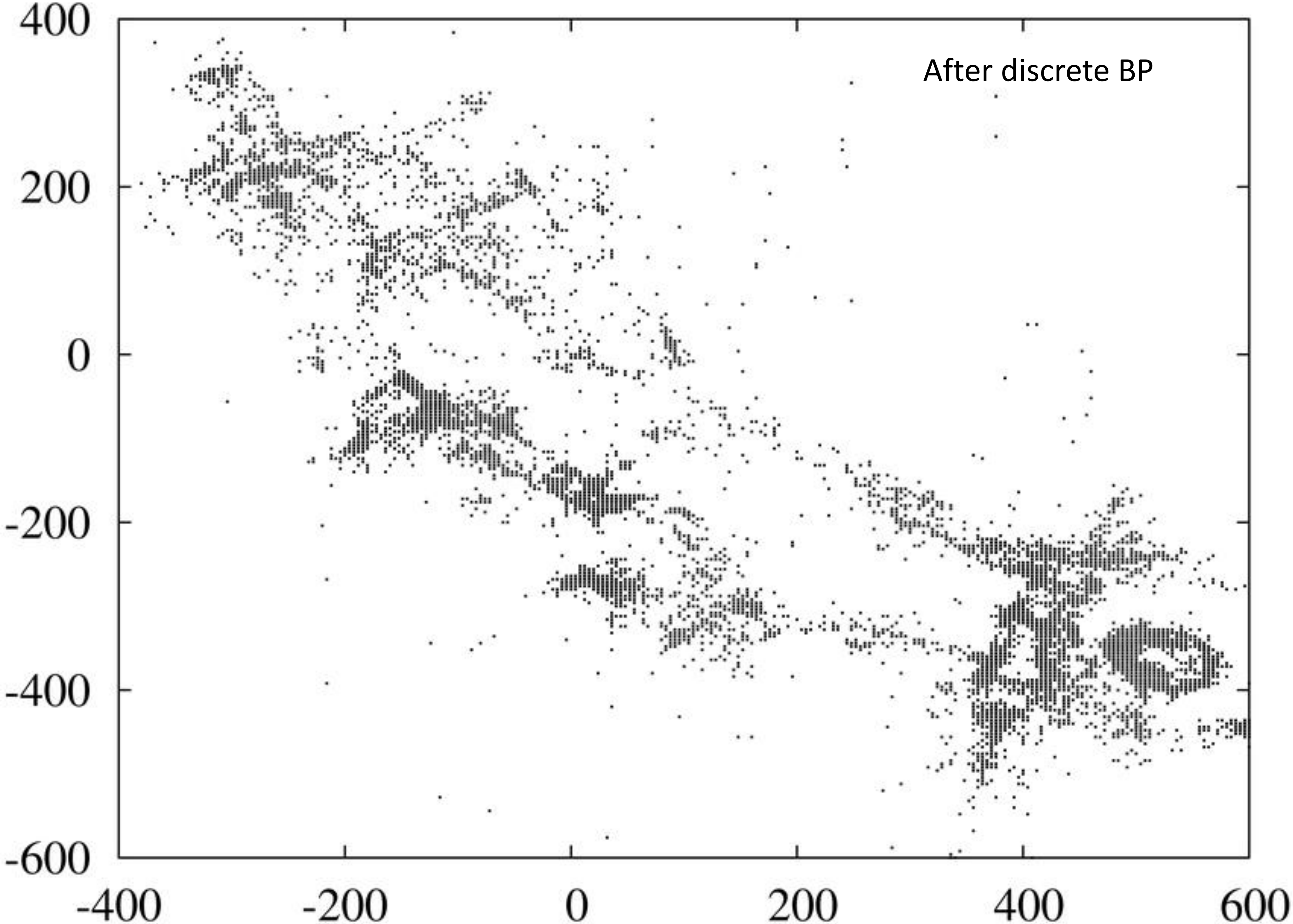
Block Diagonal Matrix

$$\Delta z = C^{-1}(w - E^\top \Delta y)$$

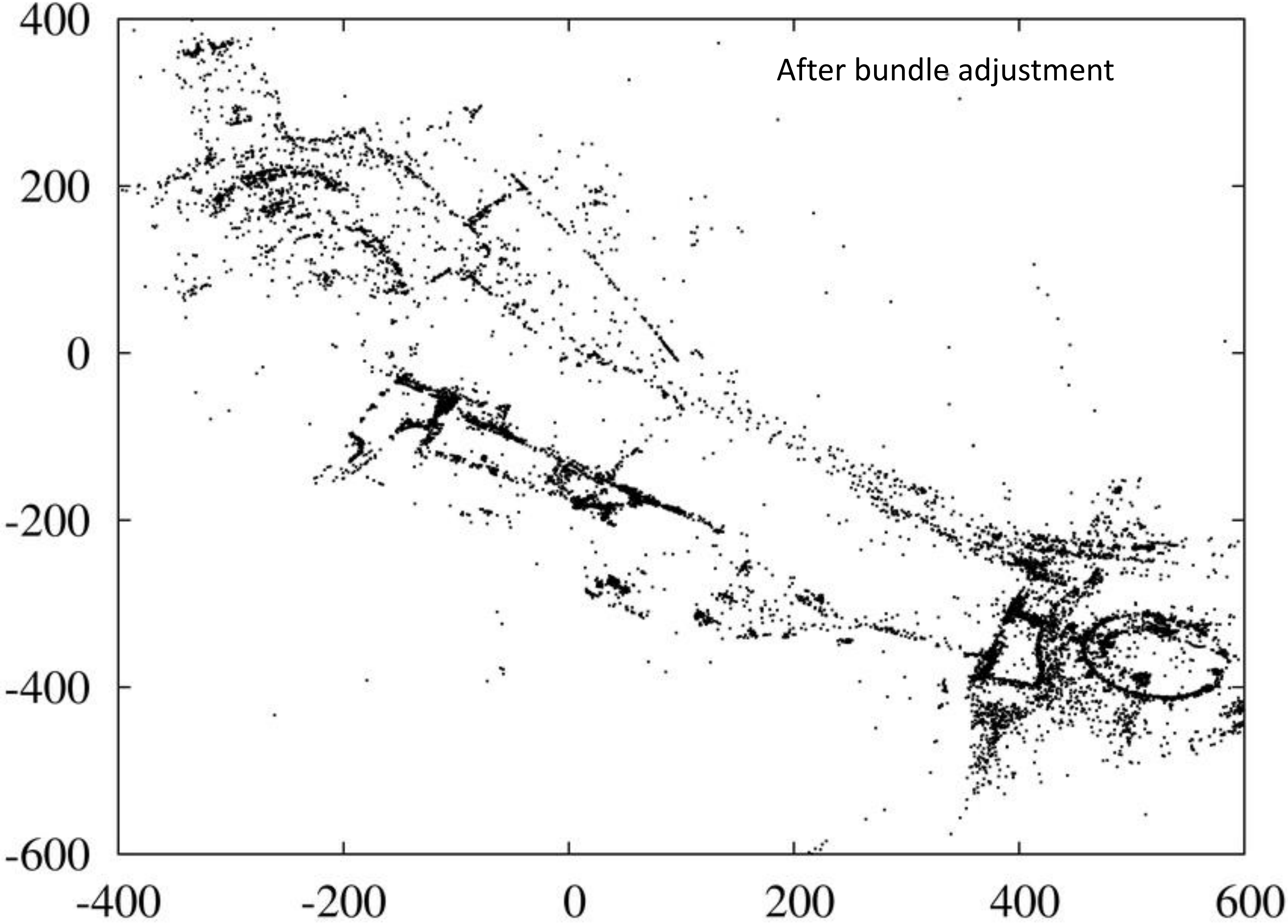
$$\left[B - EC^{-1}E^\top\right] \Delta y = v - EC^{-1}w$$

Small-scale linear system

[Crandall et al. 13]



[Crandall et al. 13]



Experimental Results

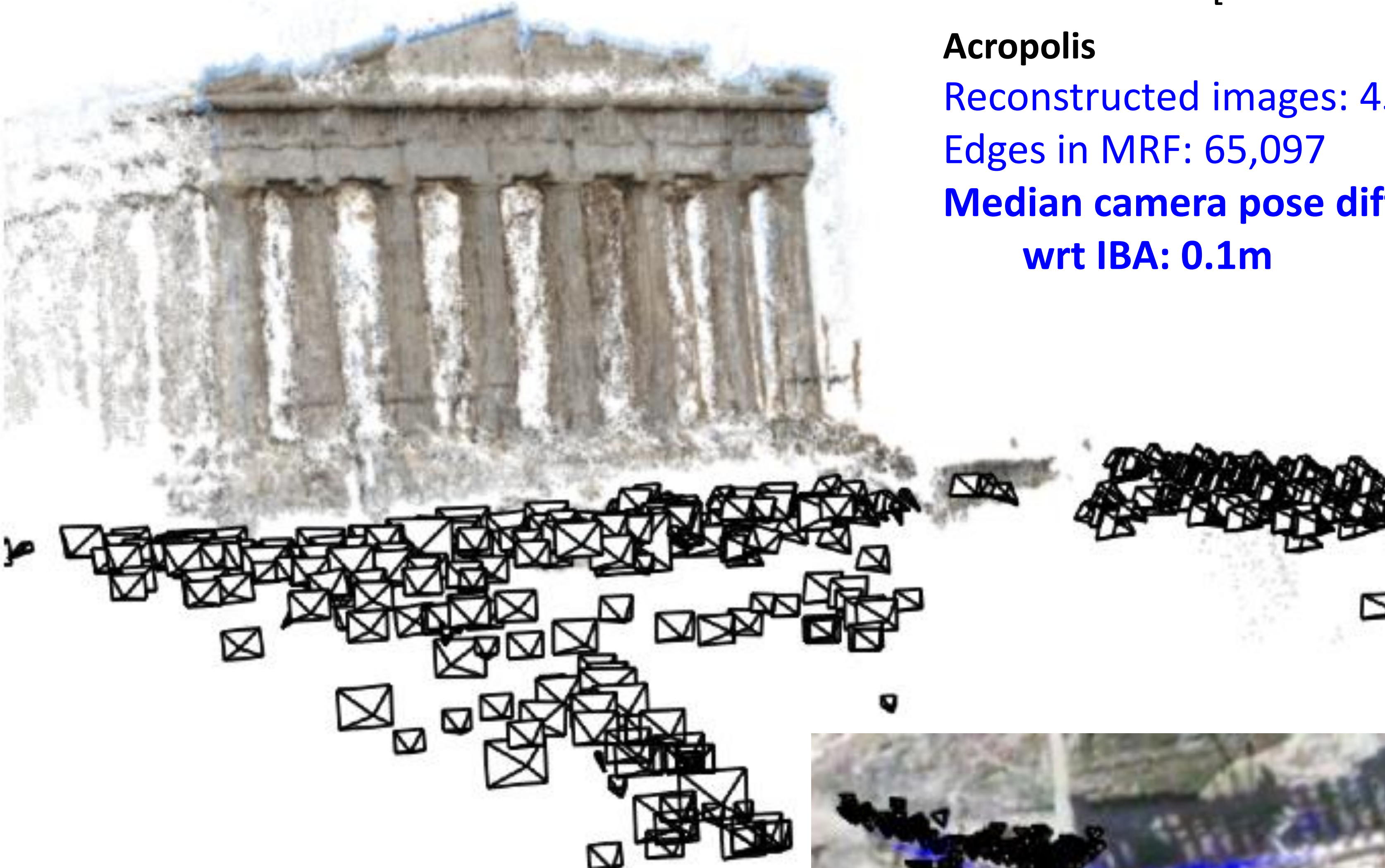
[Crandall et al. 13]

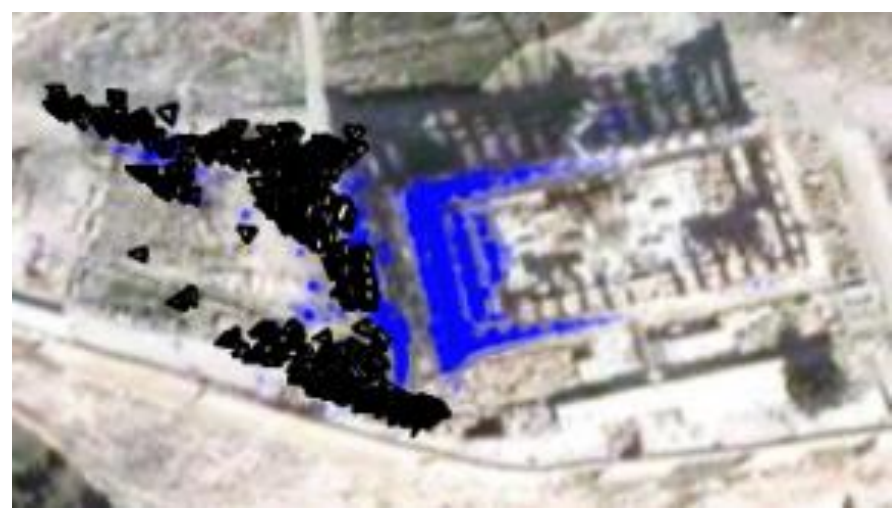
Acropolis

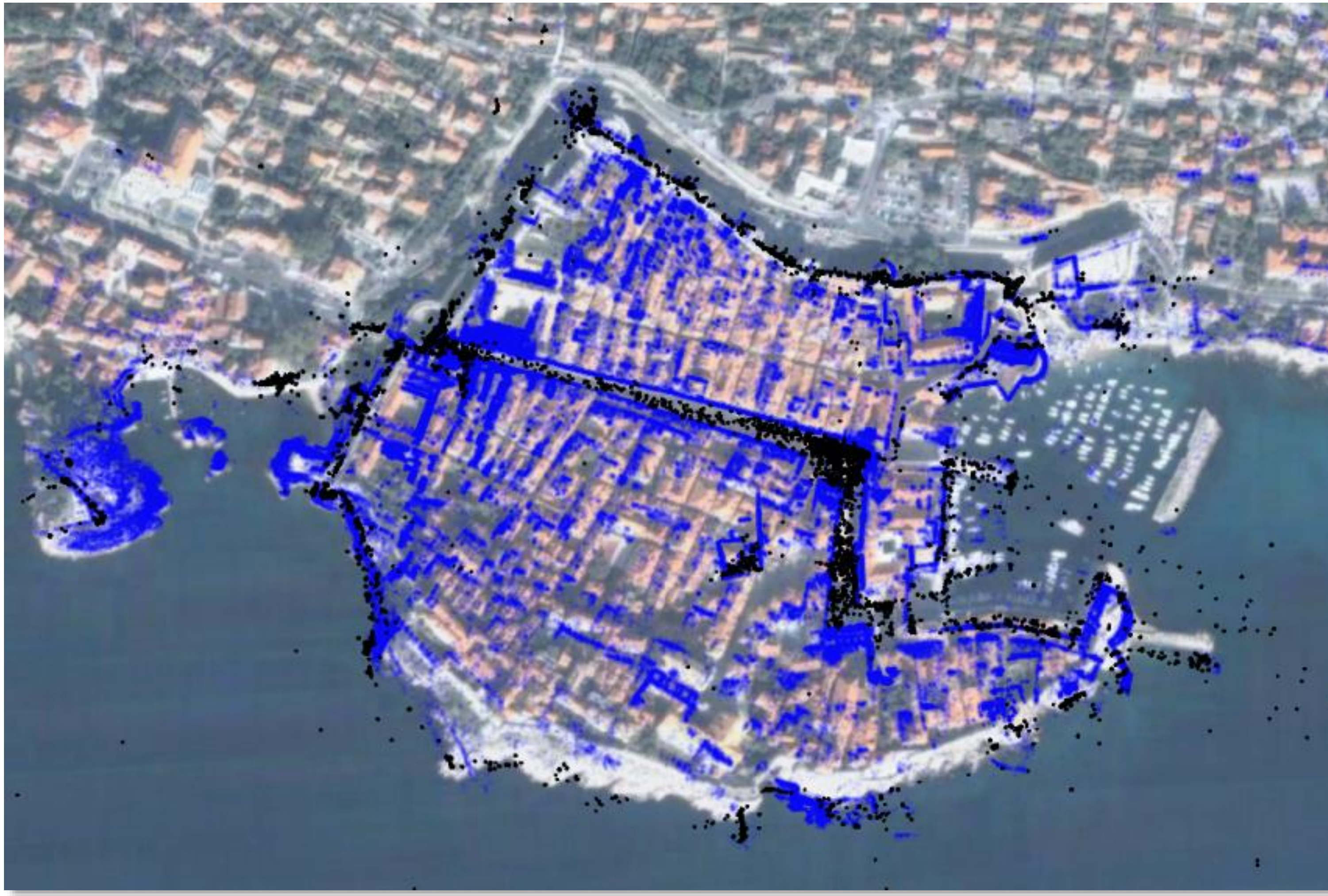
Reconstructed images: 454

Edges in MRF: 65,097

Median camera pose difference wrt IBA: 0.1m







Dubrovnik (Croatia)
Reconstructed images: 6,532
Edges in MRF: 1,835,488
Median camera pose difference wrt IBA: 1.0m

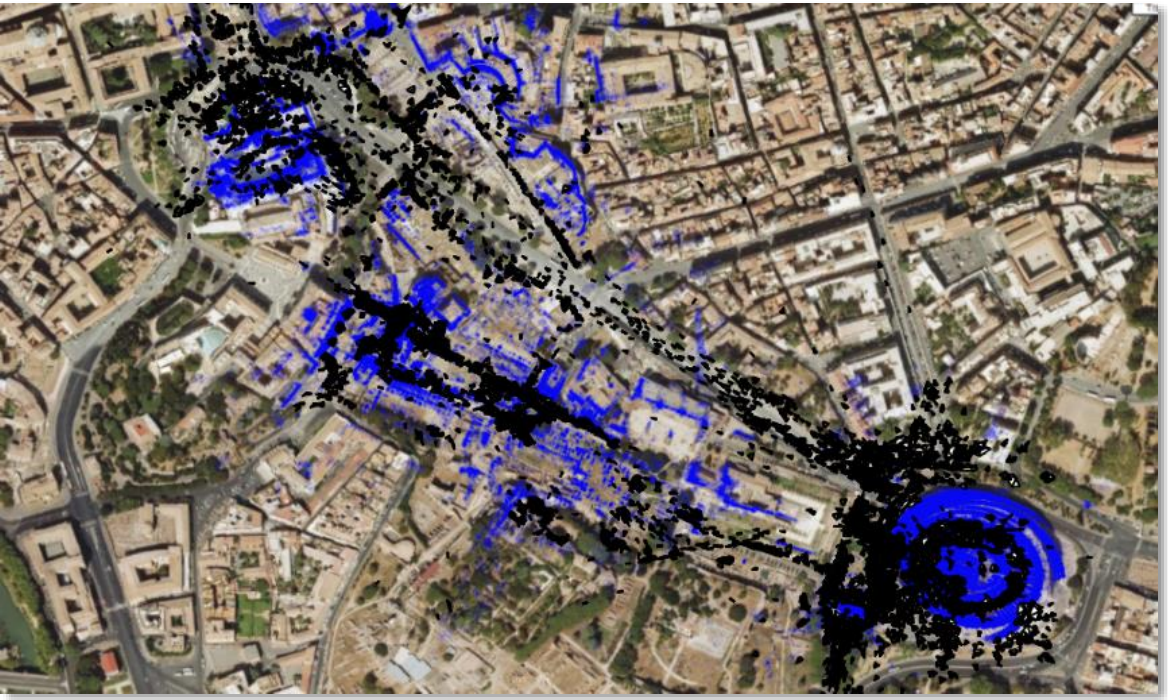
[Crandall et al. 13]

Central Rome

[Crandall et al. 13]

Reconstructed images: 14,754

Edges in MRF: 2,258,416



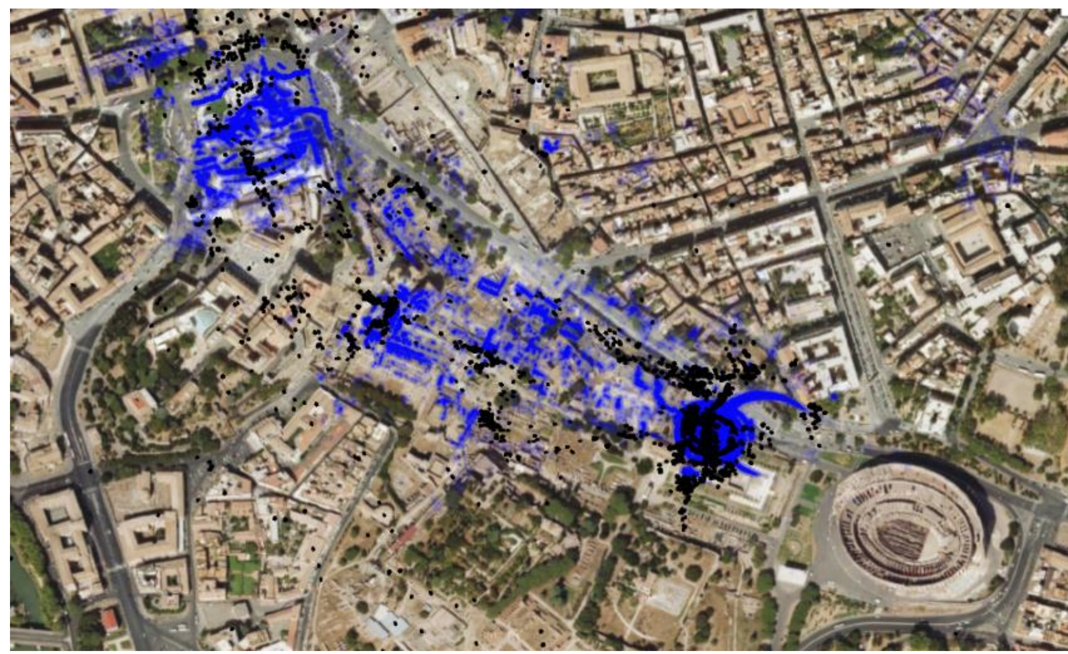
Central Rome

Reconstructed images: 14,754

Edges in MRF: 2,258,416

Median camera pose difference wrt IBA: 25.0m

Our result



Incremental Bundle Adjustment [Agarwal09]

[Crandall et al. 13]

How can Deep Learning Help?

Pipeline Steps

- Pairwise matching?

- Graph reconstruction?

- Multi-image matching?