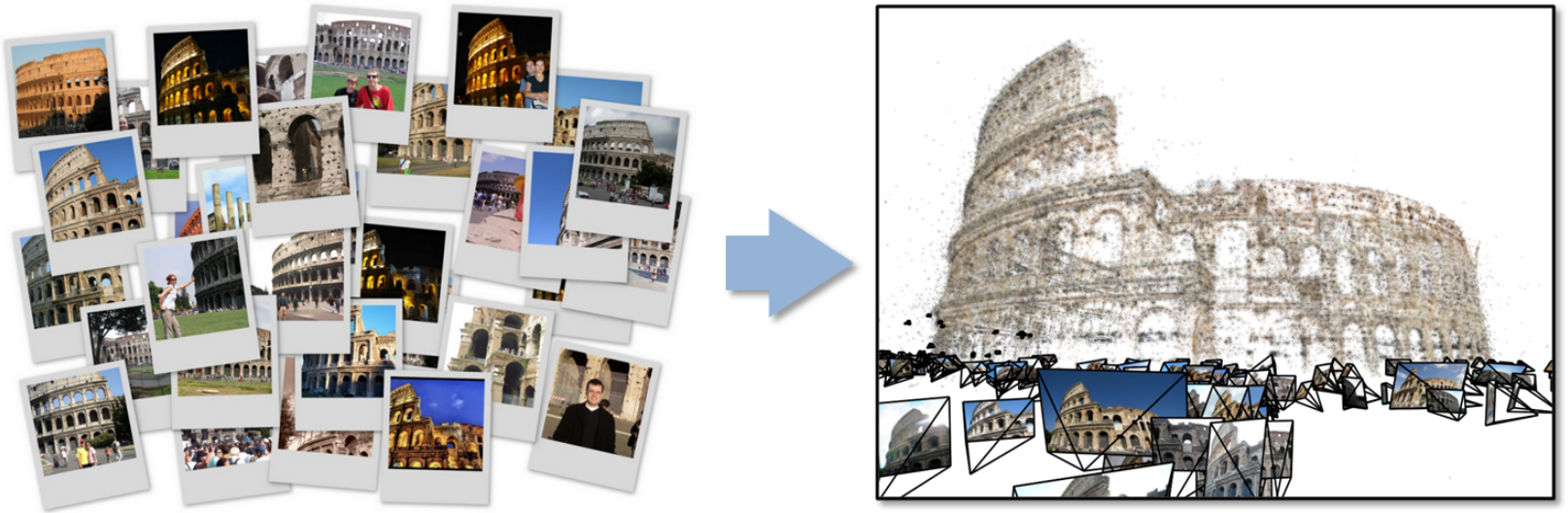
Discrete-Continuous Optimization for Large-scale Structure from Motion

David Crandall, Andrew Owens, Noah Snavely, Dan Huttenlocher

Presented by: Rahul Garg

SfM from Internet Images

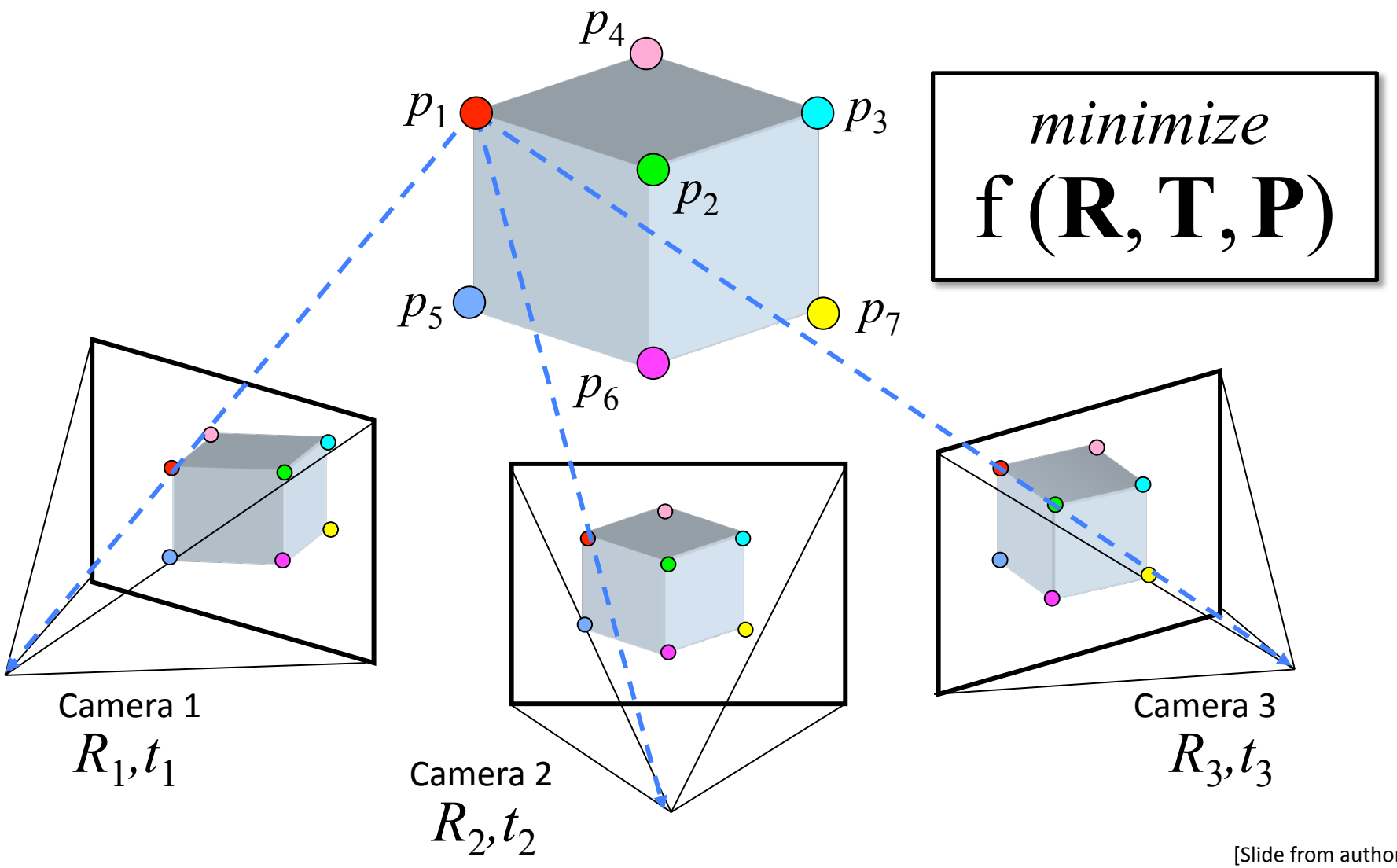
- Recent work has built 3D models from large, unstructured online image collections
  - [Snavely06], [Li08], [Agarwal09], [Frahm10], Microsoft's PhotoSynth, …



- SfM is a key part of these reconstruction pipelines

[Slide from authors]

Structure from Motion



[Slide from authors]

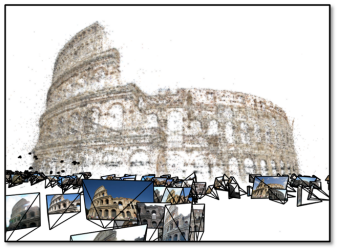
Reconstruction pipeline

Feature detection

Feature matching
Find scene points seen by multiple cameras

Initialization
Robustly estimate camera poses and/or scene points

Bundle adjustment
Refine camera poses **R, T** and scene structure **P**



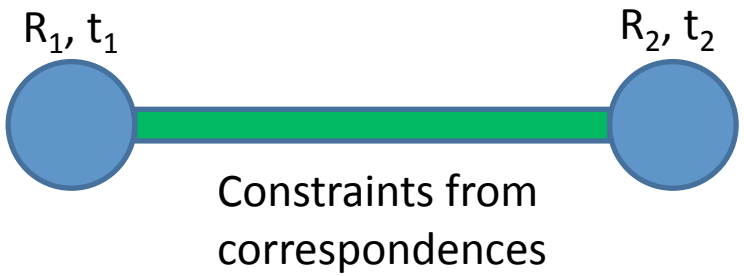
Incremental Bundle Adjustment (IBA)

[Slide from authors]

How IBA Works

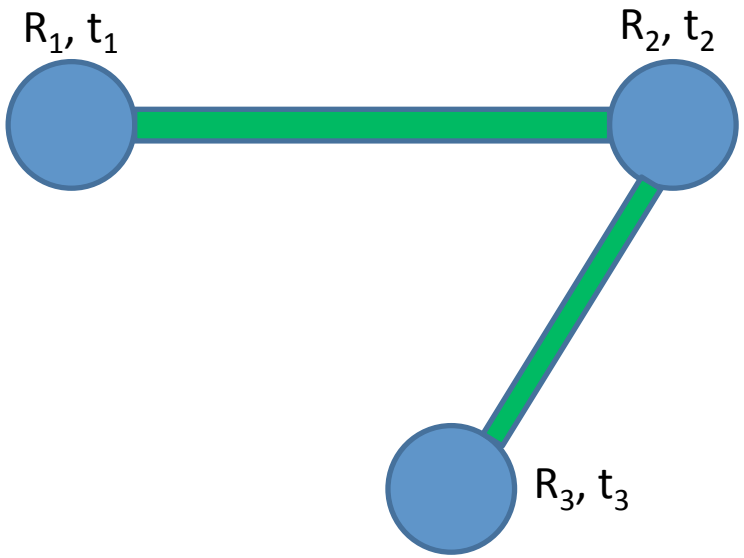
- Image graph with pairwise point correspondences

How IBA Works



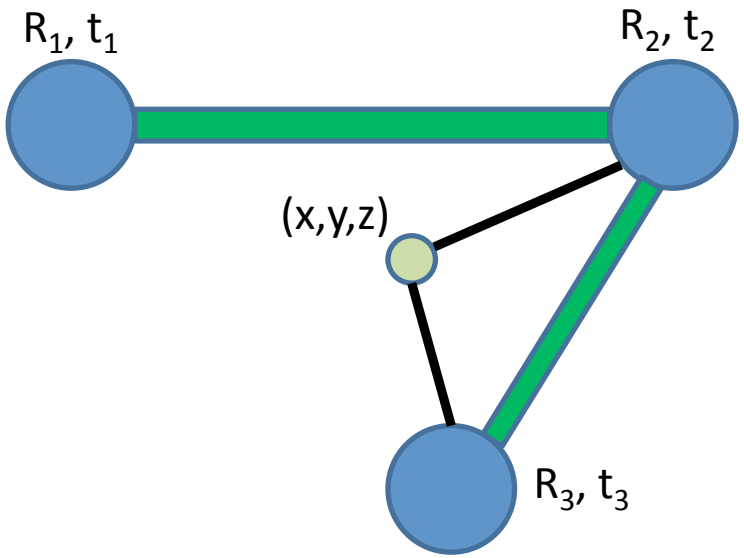
Start with a seed pair

How IBA works



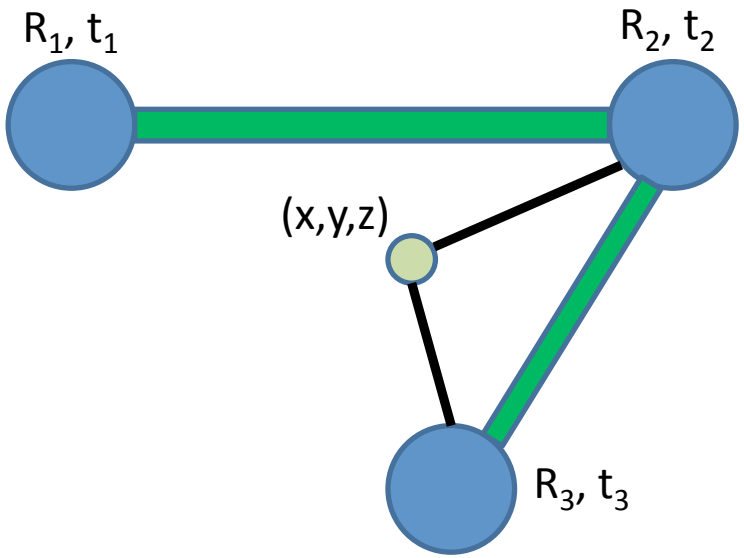
Add new cameras to the seed

How IBA works



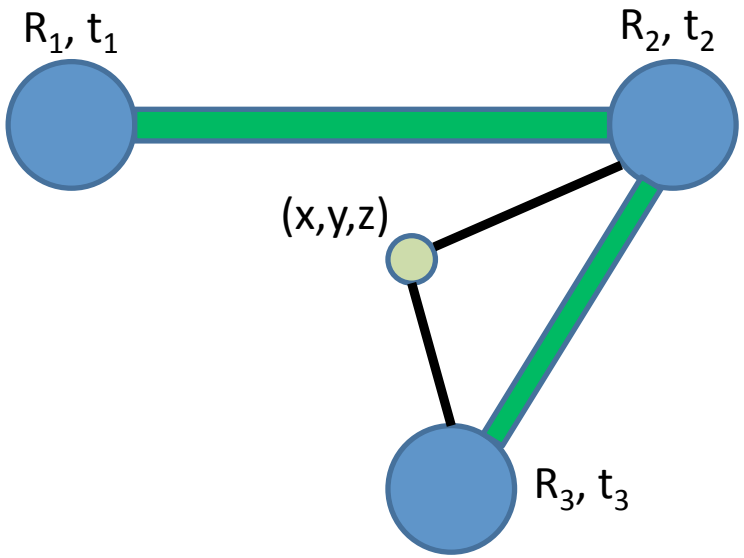
Add new cameras to the seed

How IBA works



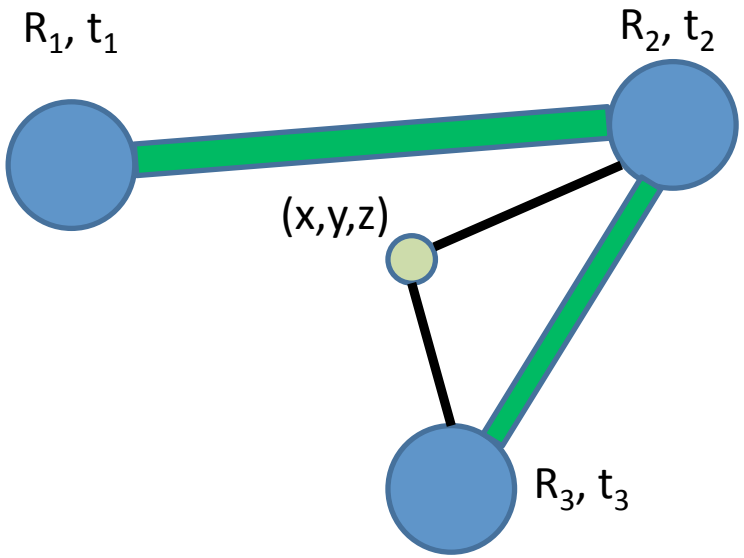
Also calculate 3D positions of points that cameras see

How IBA works



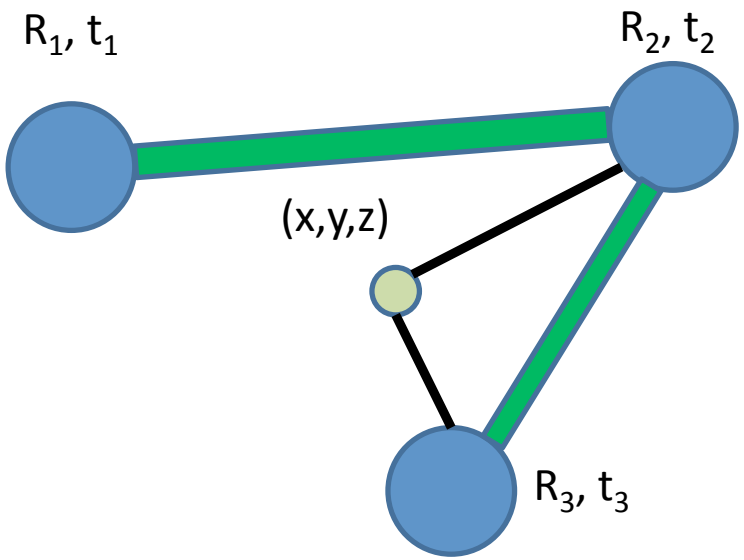
Wiggle solution periodically to get a better solution

How IBA works



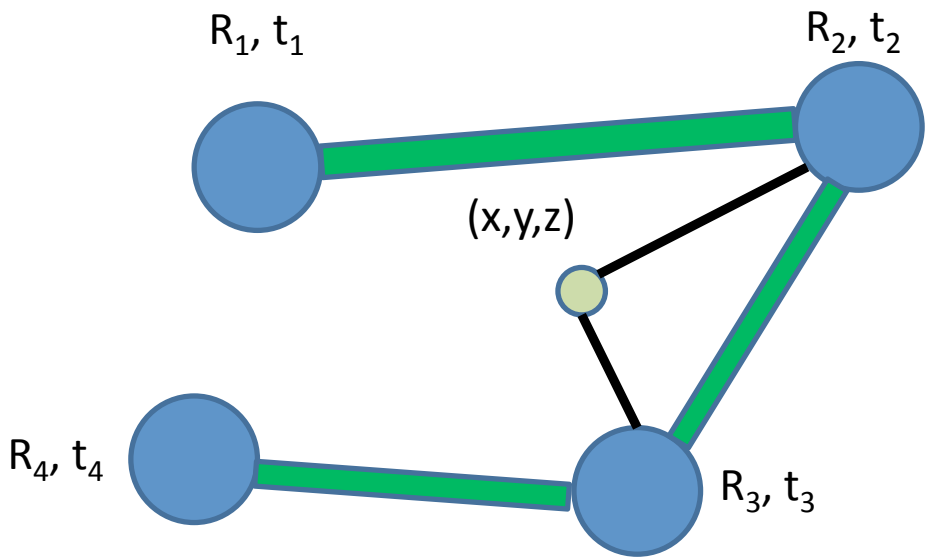
Wiggle solution periodically to get a better solution

How IBA works



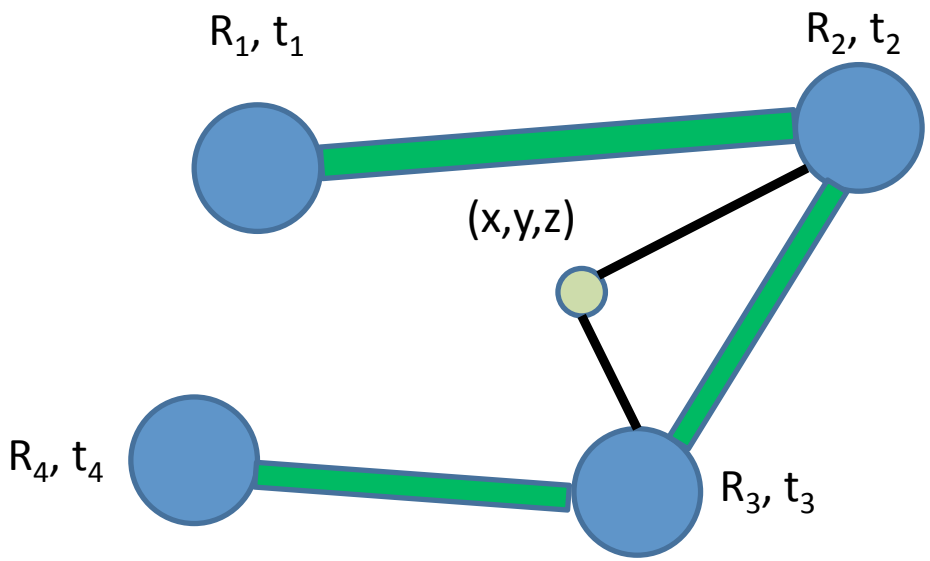
Wiggle (Bundle adjust) solution periodically to get a better solution

How IBA works



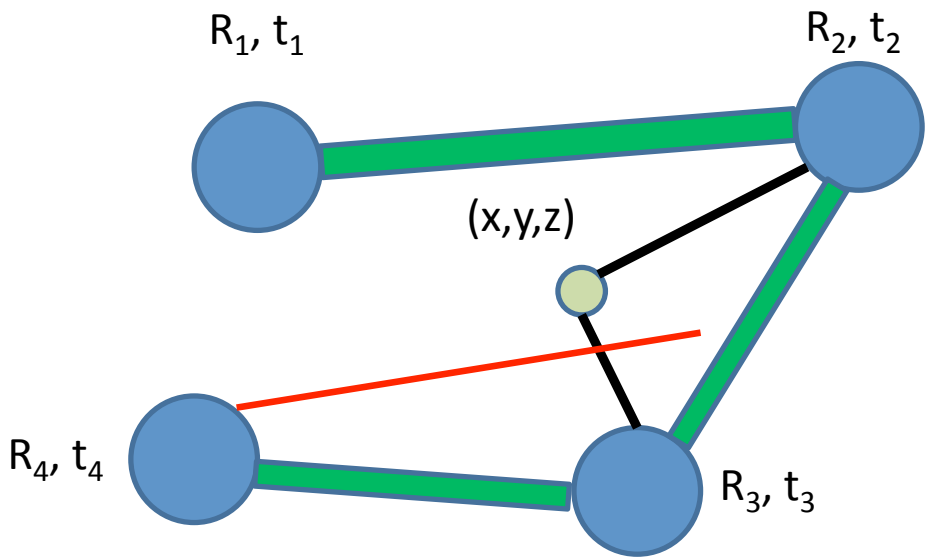
Keep adding more cameras

Problems with IBA



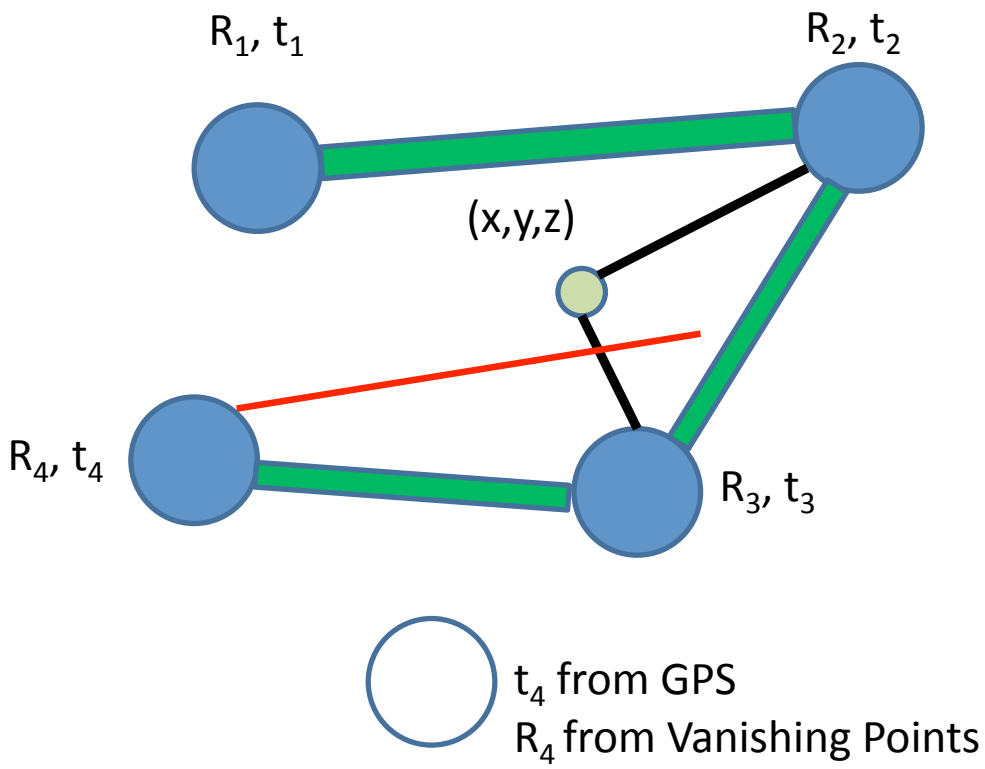
Incremental, prone to drifts and local minima

Problems with IBA



Incremental, prone to drifts and local minima

Problems with IBA



No easy way to incorporate priors (GPS info, etc.)

Initializing bundle adjustment

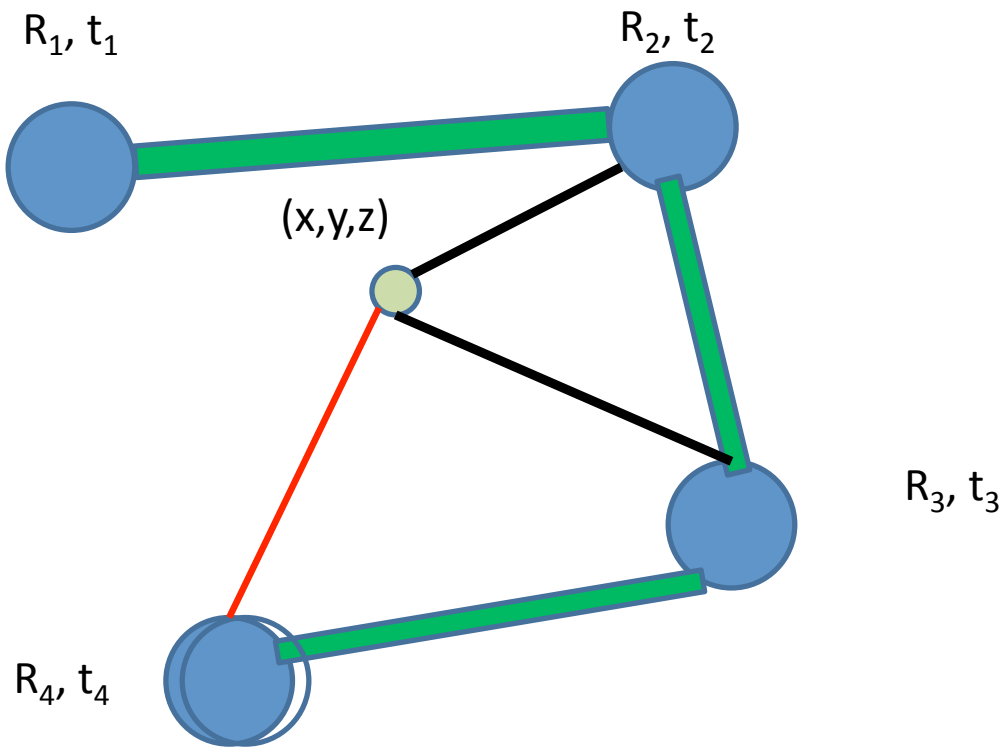
Incremental BA

- ✓ Works very well for many scenes

- ✗ Poor scalability, much use of bundle adjustment
- ✗ Poor results if a bad seed image set is chosen
- ✗ Drift and bad local minima for some scenes
- ✗ No way to add priors

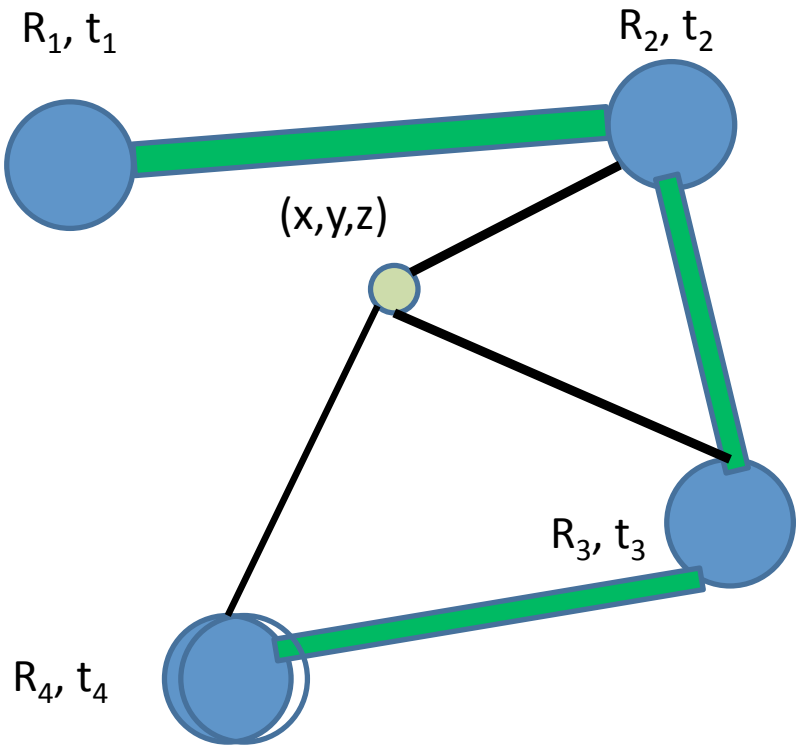
[Slide from authors]

Our Approach



Solve for everything simultaneously

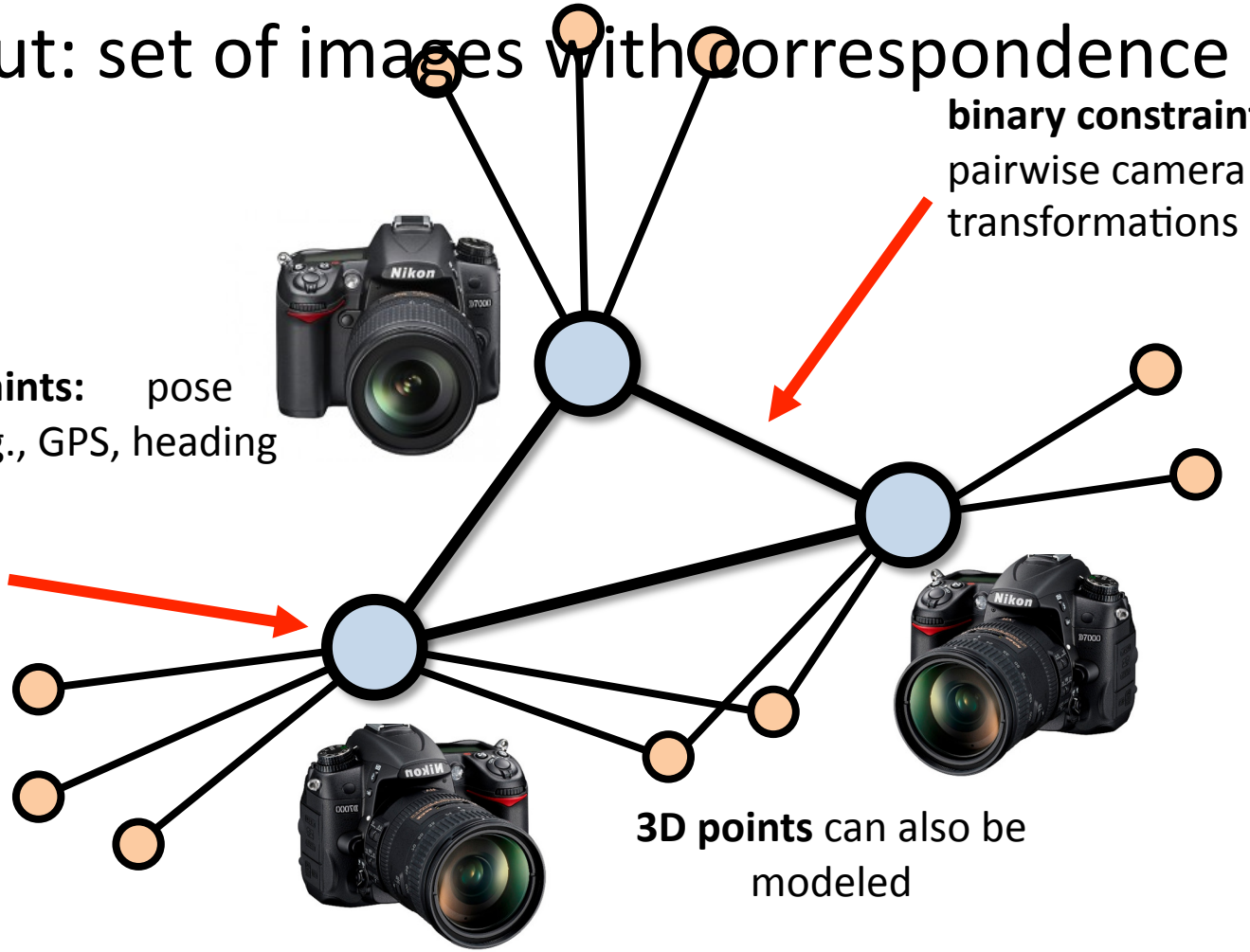
Our Approach



- Pose the problem as an MRF
- Nodes – Cameras and 3D Points
- Labels – Discretized Pose and 3D locations

The MRF model

- Input: set of images with correspondence



binary constraints: pairwise camera transformations

unary constraints: pose estimates (e.g., GPS, heading info)

3D points can also be modeled

[Slide from authors]

Making MRF Tractable

Label Space:

3 unknowns for rotation

3 unknowns for translation

Dimensionality of State space: 6

Making MRF Tractable

- First solve for rotation

Label Space:

3 unknowns for rotation

3 unknowns for translation

Dimensionality of State space: 3 + 3

Making MRF Tractable

- First solve for rotation
- Assume twist = 0

Label Space:

2 unknowns for rotation

3 unknowns for translation

Dimensionality of State space: 2 + 3

Making MRF Tractable

- First solve for rotation
- Assume twist = 0
- Assume cameras are on the ground

Label Space:

2 unknowns for rotation

2 unknowns for translation

Dimensionality of State space: 2 + 2

Making MRF Tractable

- First solve for rotation
- Assume twist = 0
- Assume cameras are near ground
- Solve using parallelizable BP on a cluster, use distance transforms to speed up message passing

Label Space:

2 unknowns for rotation

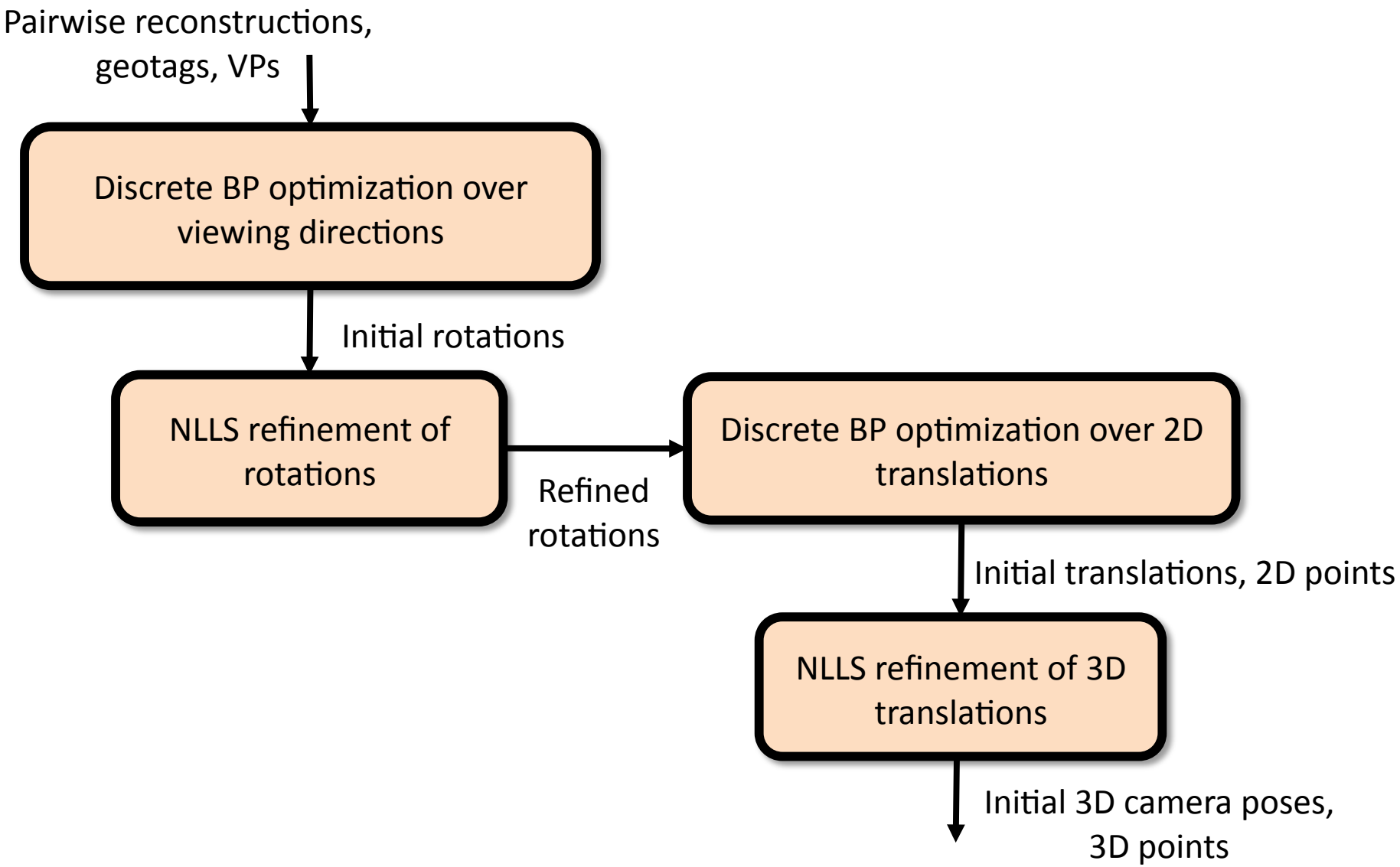
2 unknowns for translation

Dimensionality of State space: 2 + 2

Refining the MRF Solution

- Use Non Linear Least Squares to refine the MRF solution (Discrete => Continuous space)

Overall Approach



[Slide from authors]

SfM on unstructured photo collections

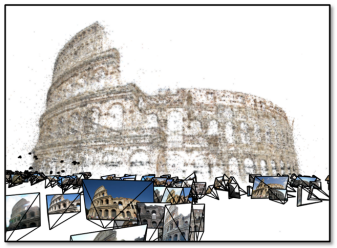
Feature point detection

Image matching
Find scene points seen by multiple cameras

Initialization
Robustly estimate camera poses and/or scene points

- Discrete optimization over viewing directions
- NLLS refinement of rotations
- Discrete optimization over 2D translations
- NLLS refinement of 3D translations

Bundle adjustment
Refine camera poses R, T and scene structure P

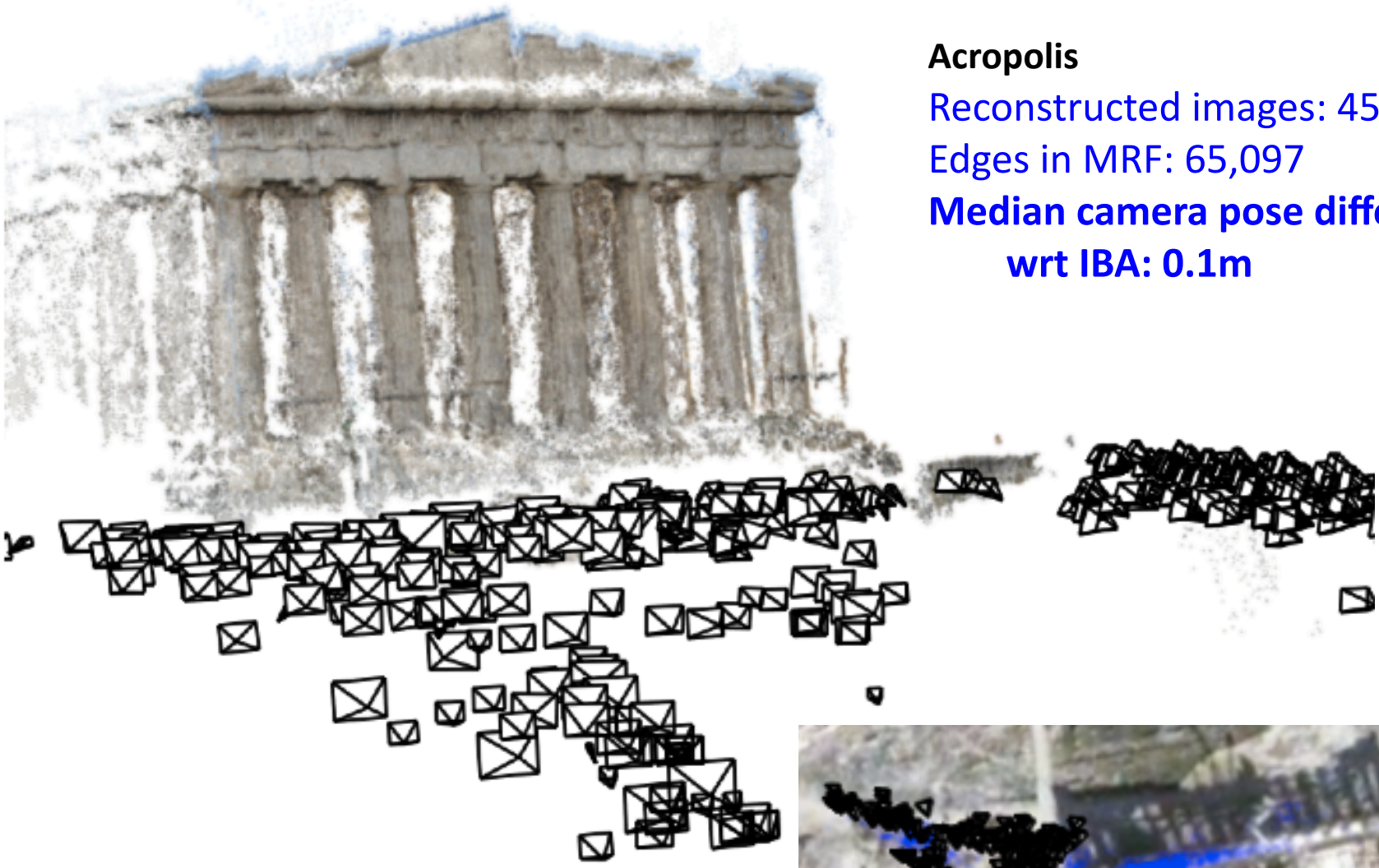


[Slide from authors]

Experimental results

- Evaluated on several Flickr datasets
  - Download photos (highest resolution available) & geotags (if available)
  - Removed images with missing EXIF focal lengths
  - Removed panoramic and high-twist images

[Slide from authors]

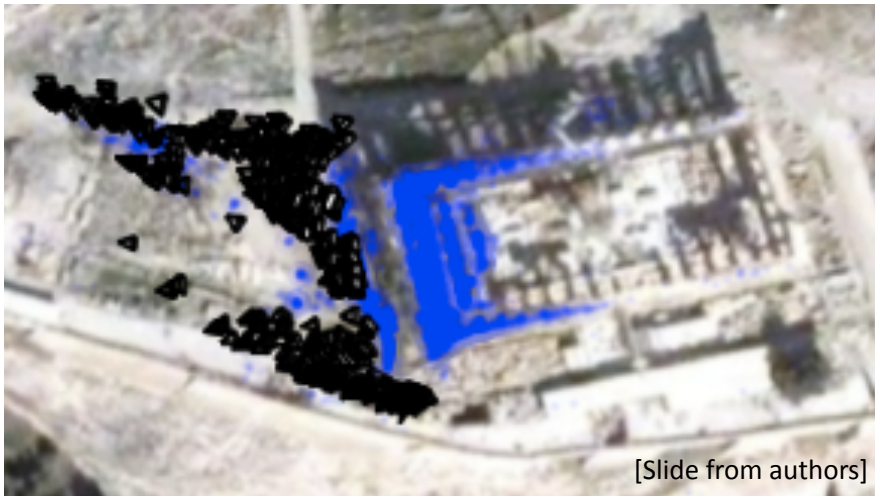


Acropolis

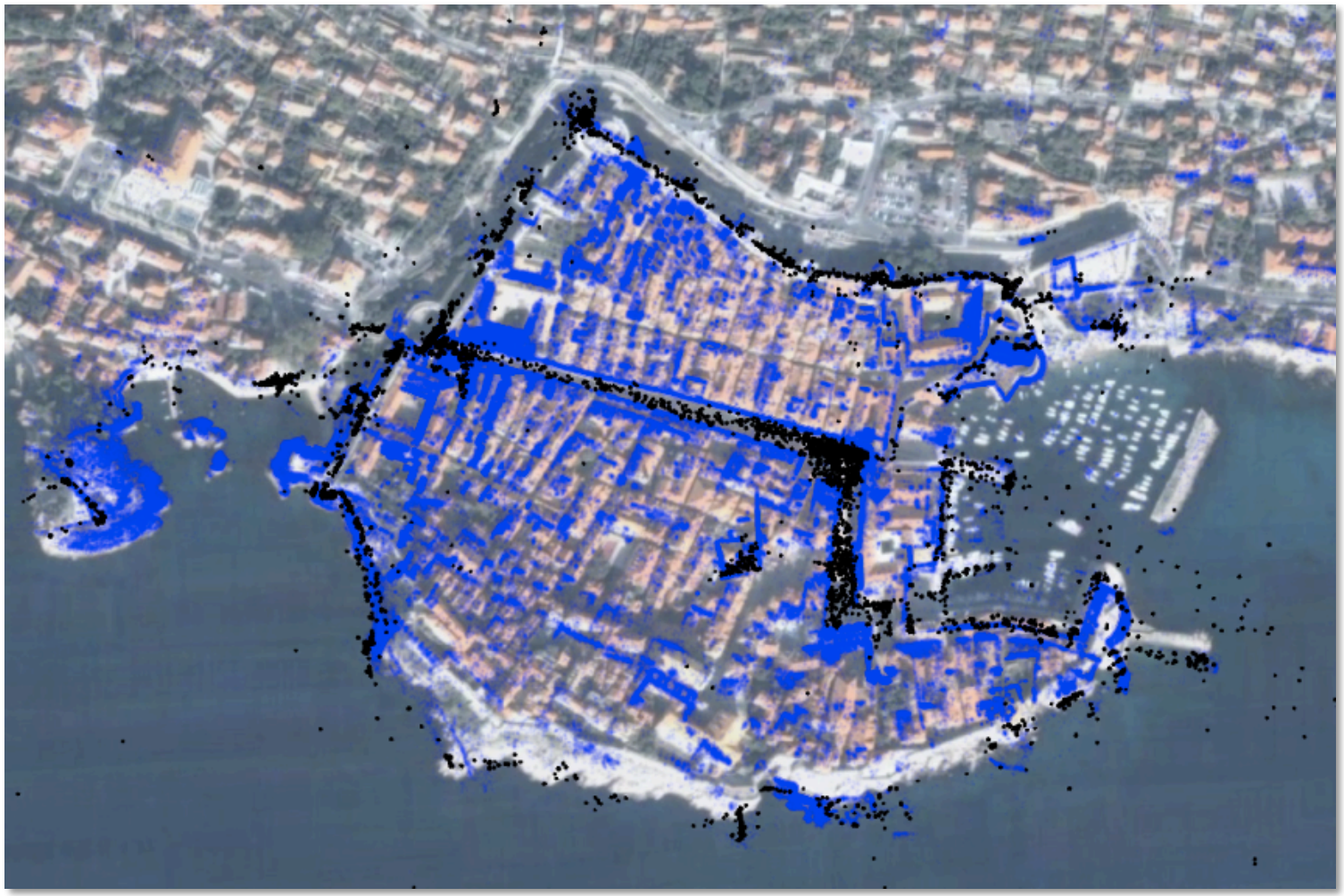
Reconstructed images: 454

Edges in MRF: 65,097

Median camera pose difference wrt IBA: 0.1m



[Slide from authors]



Dubrovnik (Croatia)

Reconstructed images: 6,532

Edges in MRF: 1,835,488

Median camera pose difference wrt IBA: 1.0m

[Slide from authors]

Quad

Total images: 6,514

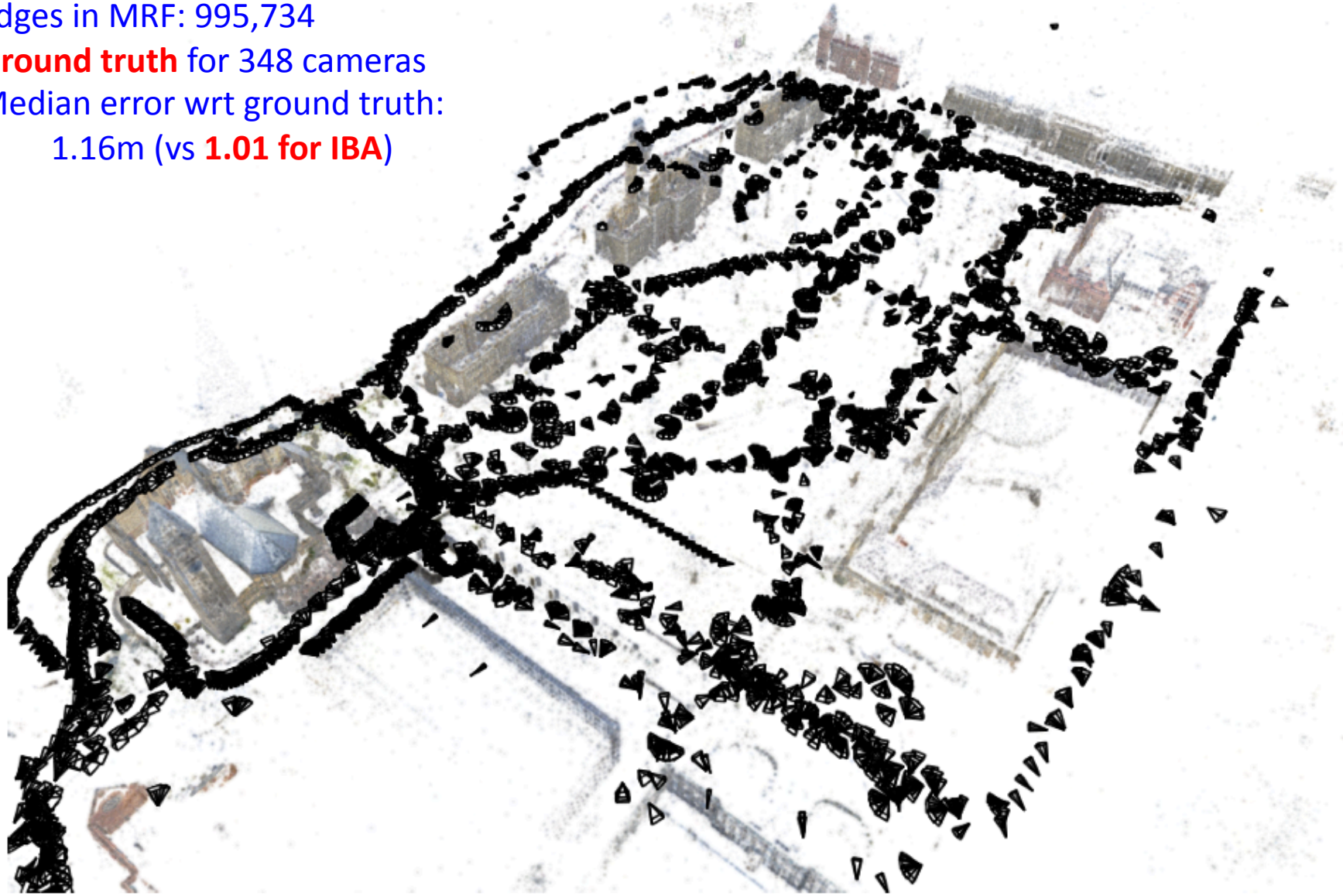
Reconstructed images: 5,233

Edges in MRF: 995,734

Ground truth for 348 cameras

Median error wrt ground truth:

1.16m (vs **1.01 for IBA**)



[Slide from authors]

Quad results

% geotags	BP	NLLS	Final BA
80%	7.50m	7.24m	**1.16m**
40%	7.67m	7.37m	**1.21m**
16%	7.66m	7.63m	**1.22m**
8%	8.27m	8.06m	**1.53m**
4%	18.25m	16.56m	**5.01m**

Table 4. Median error in camera position with respect to ground truth for the Quad dataset, with geotags for about 40% of images. The median error of IBA was 1.01m.

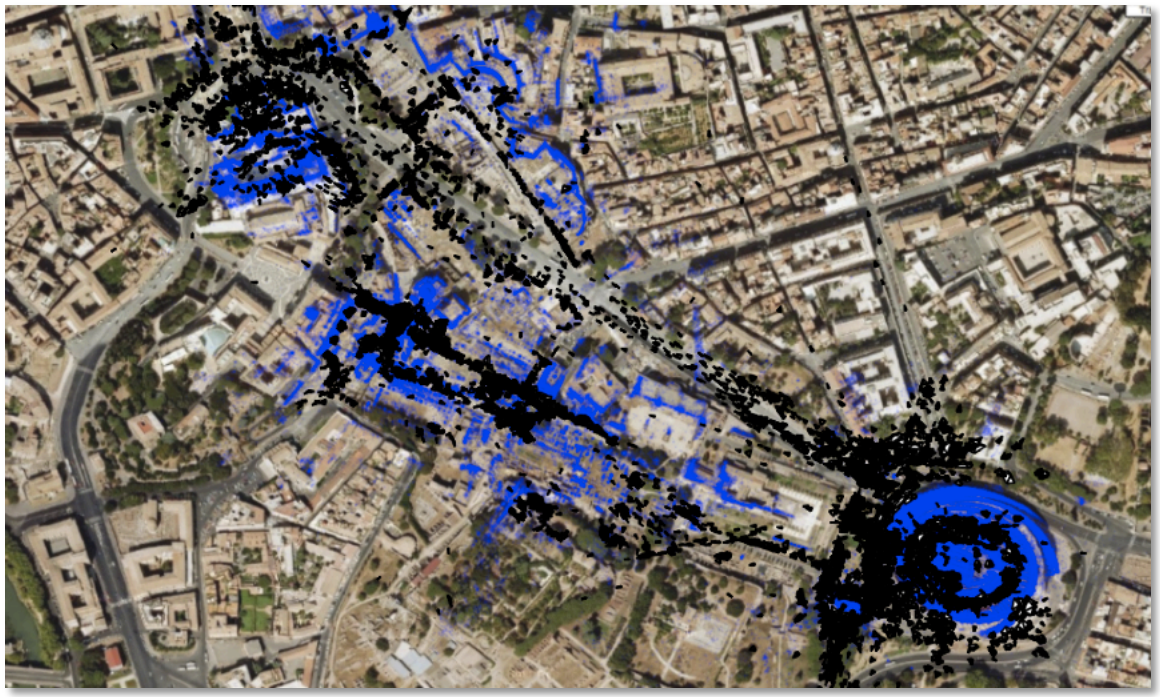
[Slide from authors]

Central Rome

Reconstructed images: 14,754

Edges in MRF: 2,258,416

[Slide from authors]



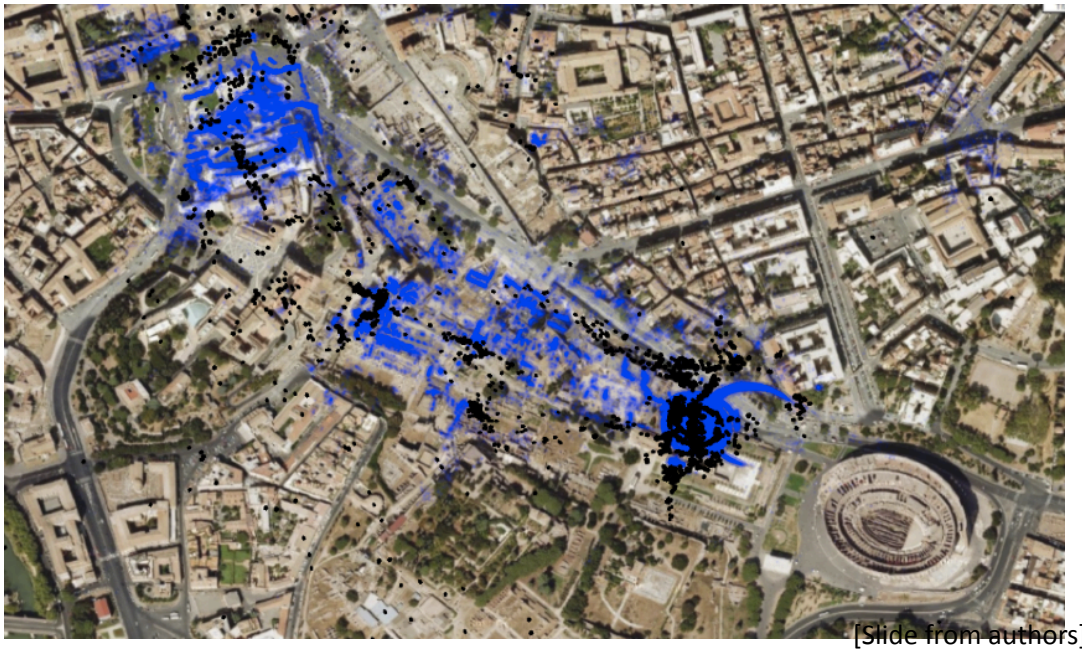
Central Rome

Reconstructed images: 14,754

Edges in MRF: 2,258,416

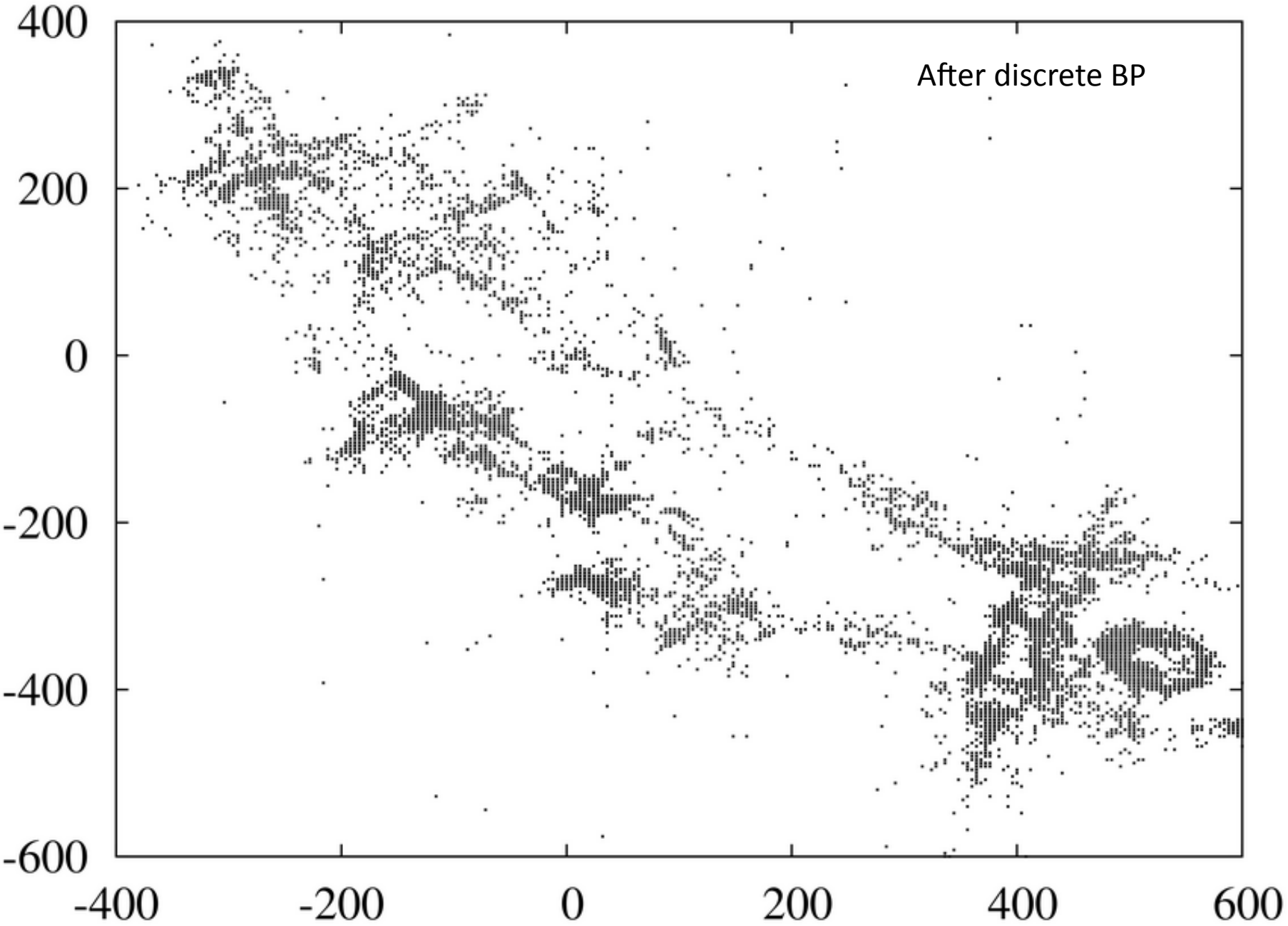
Median camera pose difference wrt IBA: 25.0m

Our result

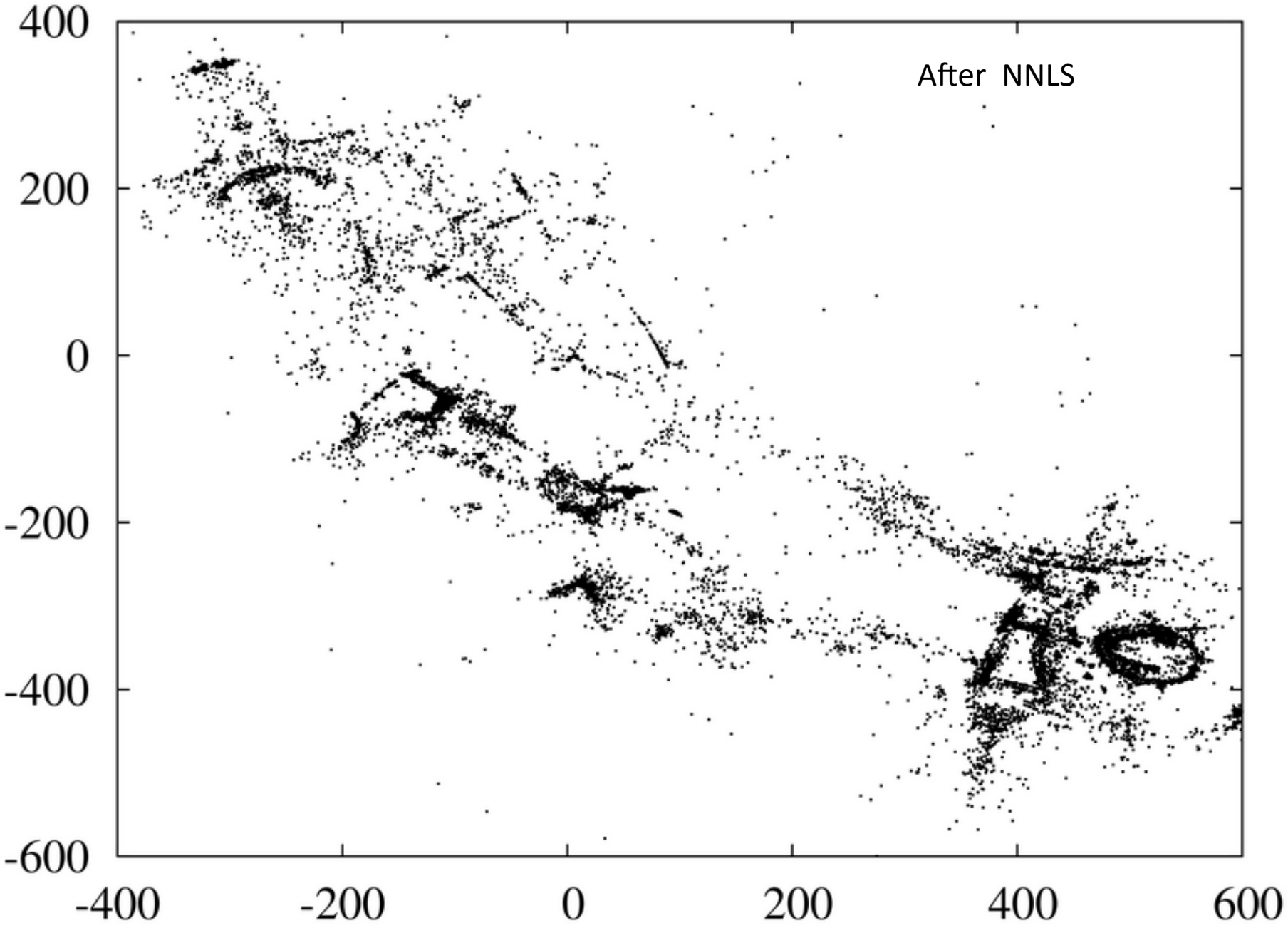


Incremental Bundle Adjustment [Agarwal09]

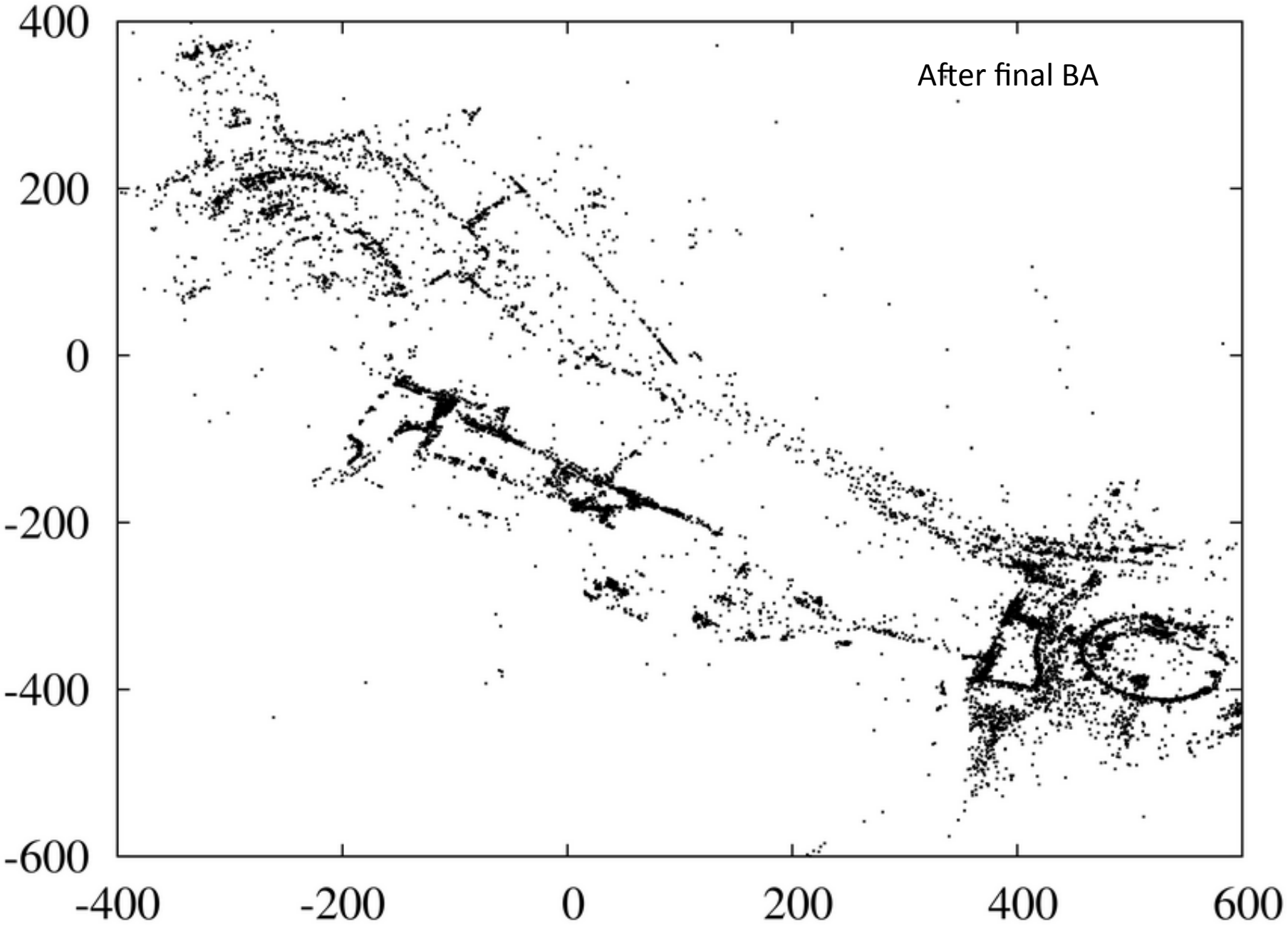
[Slide from authors]



[Slide from authors]



[Slide from authors]



[Slide from authors]

Running times

- Our results are about as good as or better than IBA, but at a fraction of the computational cost
  - Favorable asymptotic complexity
  - BP is easy to parallelize on a cluster, unlike BA

	Our approach						Incremental
Dataset	**Rot BP**	**Rot NLLS**	**Trans BP**	**Trans NLLS**	**Bund Adj**	**Total**	**BA**
Acropolis	50s	16s	7m 24s	49s	5m 36s	**0.2 hours**	**0.5 hours**
Quad	40m 57s	8m 46s	53m 51s	40m 22s	5h 18m 00s	**7.7 hours**	**62 hours**
Dubrovnik	28m 19s	8m 28s	29m 27s	7m 22s	4h 15m 57s	**5.5 hours**	**28 hours**
CentralRome	1h 8m 24s	40m 0s	2h 56m 36s	1h 7m 51s	7h 20m 00s	**13.2 hours**	**82 hours**

IBA vs Our Approach

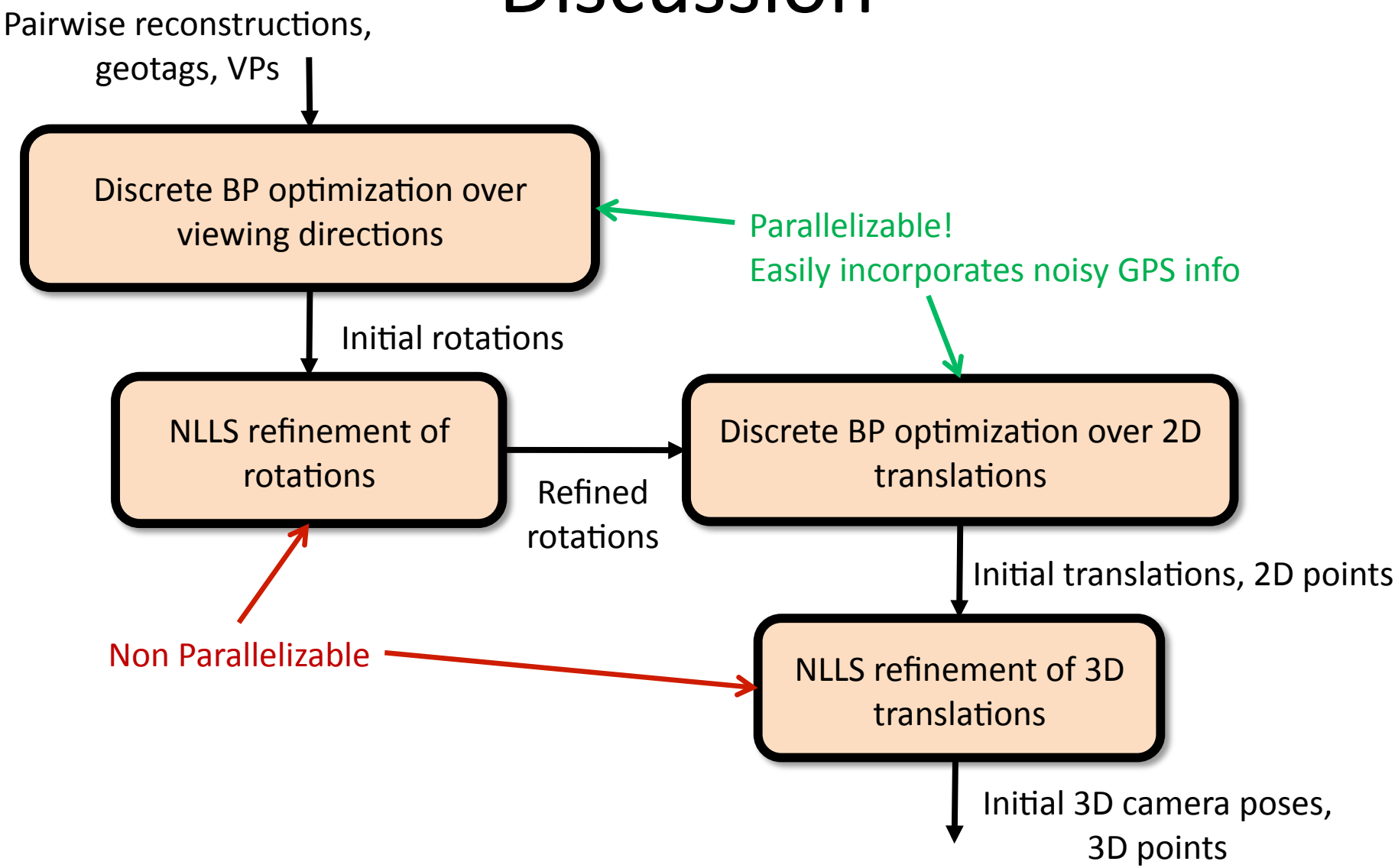
IBA

- Prone to drift, local minima
- Dependent upon seed set
- Not robust to outliers
- No easy way to add priors (GPS tags, etc.)
- Not Parallelizable

Our Approach

- Simultaneously solve for global optima
- Objective function robust to outliers
- Easy to add priors
- Parallelizable

Discussion


Pairwise reconstructions, geotags, VPs
Discrete BP optimization over viewing directions
Parallelizable!
Easily incorporates noisy GPS info
Initial rotations
NLLS refinement of rotations
Refined rotations
Discrete BP optimization over 2D translations
Initial translations, 2D points
Non Parallelizable
NLLS refinement of 3D translations
Initial 3D camera poses, 3D points

Discussion

Dataset	Our approach						Incremental
	Rot BP	Rot NLLS	Trans BP	Trans NLLS	Bund Adj	Total	BA
Acropolis	50s	16s	7m 24s	49s	5m 36s	**0.2 hours**	**0.5 hours**
Quad	40m 57s	8m 46s	53m 51s	40m 22s	5h 18m 00s	**7.7 hours**	**62 hours**
Dubrovnik	28m 19s	8m 28s	29m 27s	7m 22s	4h 15m 57s	**5.5 hours**	**28 hours**
CentralRome	1h 8m 24s	40m 0s	2h 56m 36s	1h 7m 51s	7h 20m 00s	**13.2 hours**	**82 hours**

	Rotational difference					Translational difference				Point difference
	Our approach			Linear approach [8]		Our approach				Our approach
Dataset	BP	NLLS	Final BA	Linear	NLLS	Geotags	BP	NLLS	Final BA	Final BA
Acropolis	14.1°	1.5°	**0.2°**	1.6°	1.6°	12.9m	8.1m	2.4m	**0.1m**	**0.2m**
Quad	4.7°	4.6°	**0.2°**	41°	41°	15.5m	16.6m	14.2m	**0.6m**	**0.5m**
Dubrovnik	9.1°	4.9°	**0.1°**	11°	6°	127.6m	25.7m	15.1m	**1.0m**	**0.9m**
CentralRome	6.2°	3.3°	**1.3°**	27°	25°	413.0m	27.3m	27.7m	**25.0m**	**24.5m**

Table 2. Median differences between our camera pose estimates and those produced by incremental bundle adjustment.

- Two step process for finding R and T. Is that why final BA is required?
- Final BA is still the bottleneck – not parallelizable

Thank You

References

- **Photo Tourism: Exploring Image Collections in 3D.**
  Noah Snavely, Steven M. Seitz, Richard Szeliski. SIGGRAPH 2006
- **Building Rome in a Day**
  Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski, ICCV 2009
- **Building Rome on a Cloudless Day**
  Jan-Michael Frahm, Pierre Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi-Hung Jen, Enrique Dunn, Brian Clipp, Svetlana Lazebnik, Marc Pollefeys, ECCV 2010