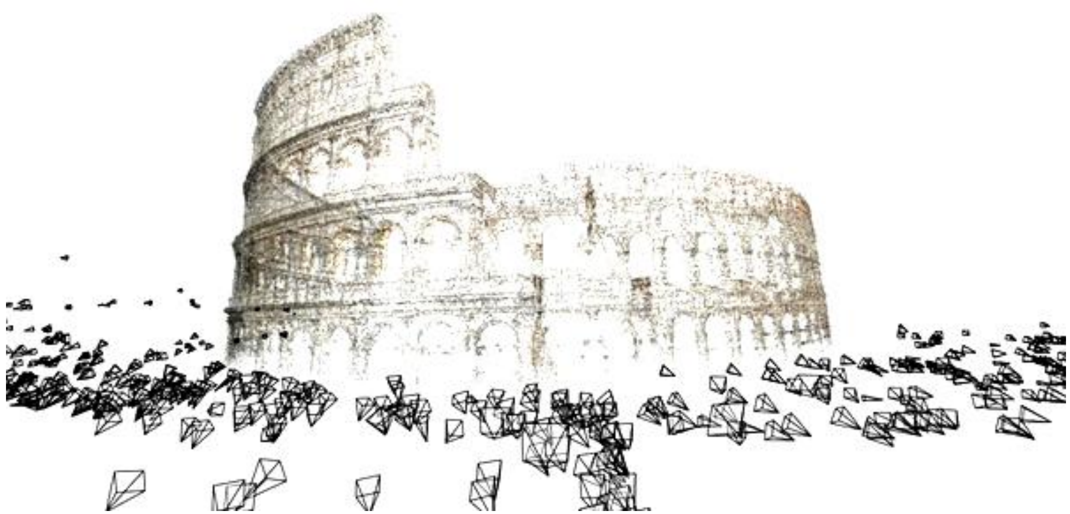
Building Rome in a Day

Agarwal, Sameer, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, and Richard Szeliski.

Presented by Ruohan Zhang



Source: Agarwal et al., Building Rome in a day.

Photo by e_vodkin

City of Dubrovnik, 4619 images, 3485717 points

Source: Agarwal et al., Building Rome in a day.

Outline

- A review of the method
- Reconstruction quality
  - How many images do we need?
  - How and why camera focal length help reconstruction
  - Number of keypoints
- Ambiguity: symmetry and repeated features
- More examples
- Computational cost breakdown

Method Overview

- The correspondence problem (distributed implementation)
  - SIFT + ANN (approximate nn) + ratio test + RANSAC (rigid scenes) to clean up matches
  - large scale matching: match graph
    - nodes are images, edges are matches
    - propose edges (matches) and then verify
    - proposal: whole image similarity (visual word) + query expansion
  - multiple images: feature track generation (connected component)

- The structure from motion (SFM) problem: given corresponding points, solve for 3D positions of the object interest points, camera orientations, positions, and focal lengths
  - practical purpose: skeletal set + incremental solution (bundle adjustment)
  - Multiview stereo to recover 3D geometries

Experiments

1. Datasets: objects with clean background, buildings, and street views
2. SIFT + ANN + ratio test + RANSAC
3. SFM Software : Bundler [7] sparse point clouds
4. Visualization: Meshlab [8]

Reconstruction quality: judge by eyes.

Outline

- A review of the method
- Reconstruction quality
  - How many images do we need?
  - How and why camera focal length help reconstruction
  - Number of keypoints
- Ambiguity: symmetry and repeated features
- More examples
- Computational cost breakdown

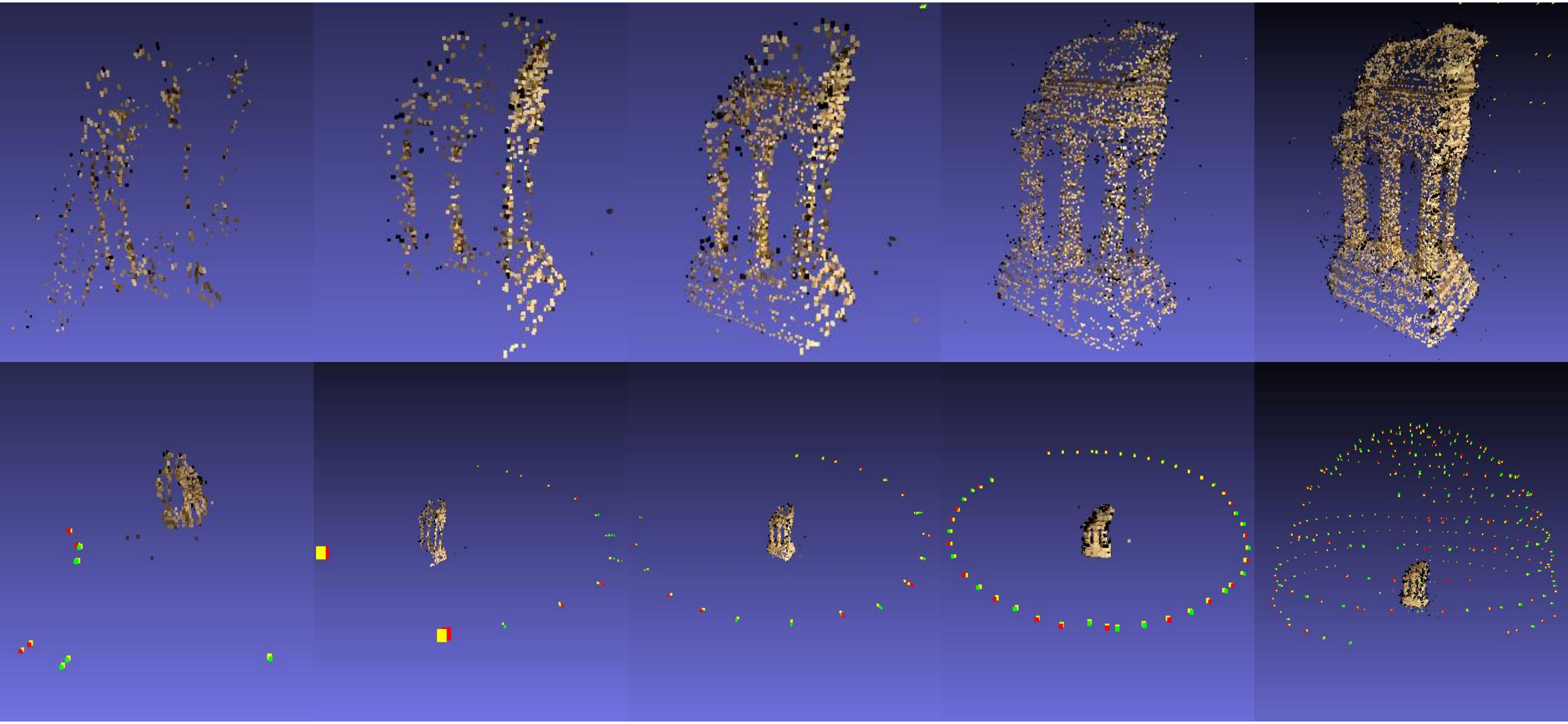
Reconstruction Quality: Image Overlaps

- How many images do we need to obtain a good reconstruction of an object?

Temple of the Dioskouroi, 317 images; Plaster stegosaurus, 363 images.

Source: Seitz et al., Multiview Stereo Evaluation Dataset.

Reconstruction Quality: Image Overlaps



Temple 8	Temple 16	Temple 24	Temple48	Temple Full
(45 degrees)	(22.5 degrees)	(15 degrees)	(7.5 degrees)	
10s	20s	34s	2m12s	40m46s

Reconstruction Quality: Image Overlaps

Dinosaur 16
(22.5 degrees)
13s

Dinosaur 24
(15 degrees)
19s

Dinosaur 48
(7.5 degrees)
45s

Dinosaur Full
15m52s

Reconstruction Quality: Image Overlaps

- General rule of thumb:
- Each point should be visible in 3+ images
- Every 15 degrees, 24 photos with a full 360 view

Source: Seitz et al., Multiview Stereo Evaluation Dataset.

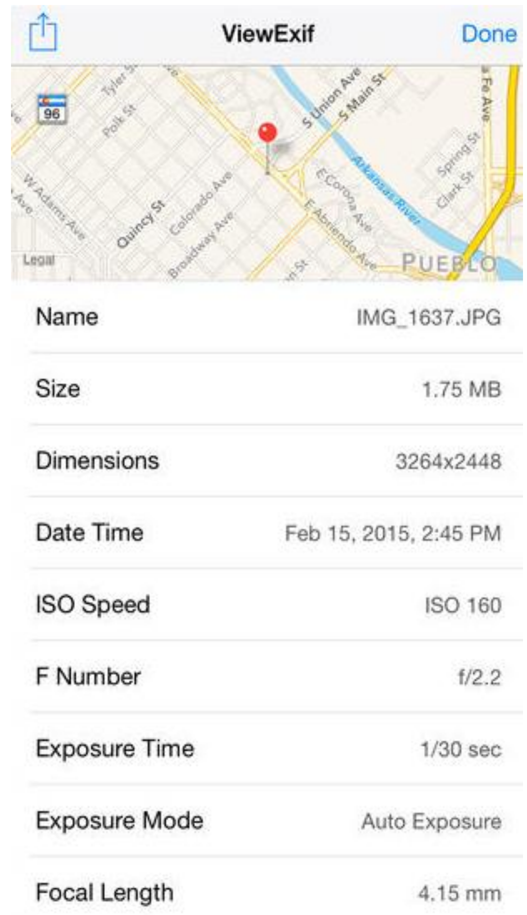
Outline

- A review of the method
- Reconstruction quality
  - How many images do we need?
  - How and why camera focal length help reconstruction
  - Number of keypoints
- Ambiguity: symmetry and repeated features
- More examples
- Computational cost breakdown

Reconstruction Quality: Camera Focal Length

- Usually obtained from the Exif tags in JPEG images.

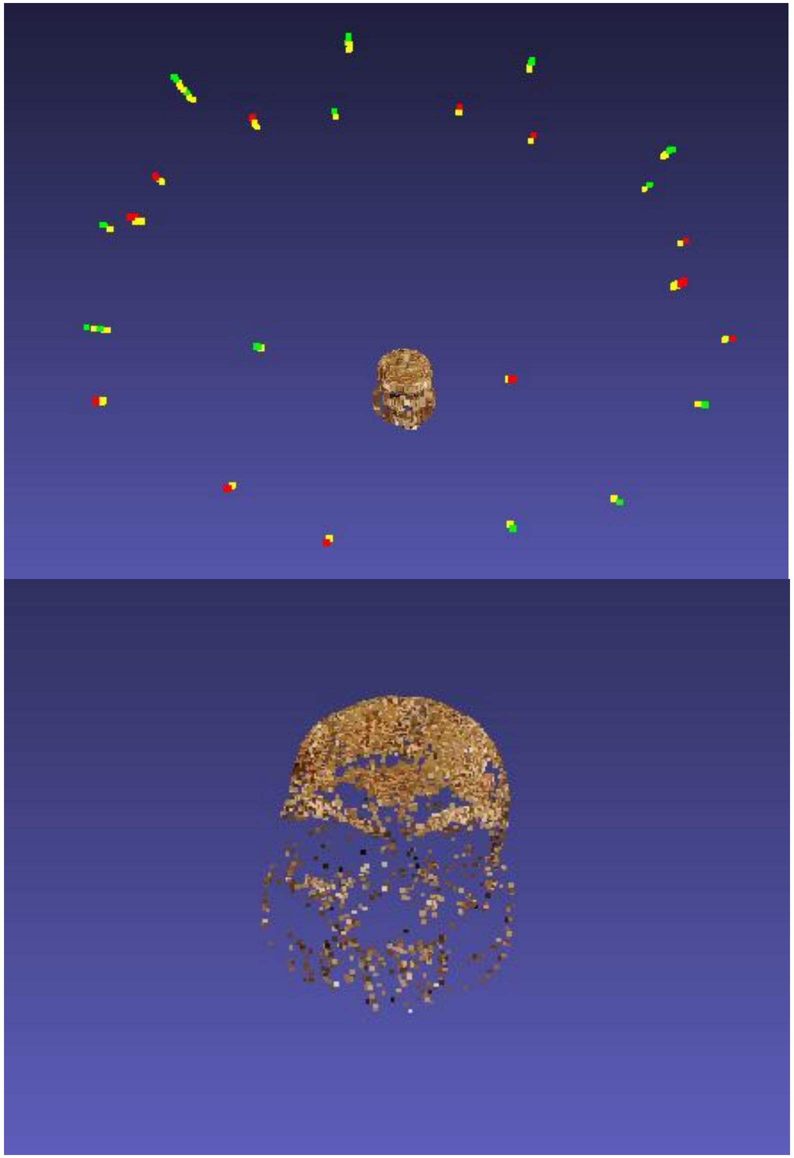
```
"Canon Canon DIGITAL IXUS 400"      => 7.176,
"Canon Canon DIGITAL IXUS 40"       => 5.76,
"Canon Canon DIGITAL IXUS 430"      => 7.176,
"Canon Canon DIGITAL IXUS 500"      => 7.176,
"Canon Canon DIGITAL IXUS 50"       => 5.76,
"Canon Canon DIGITAL IXUS 55"       => 5.76,
"Canon Canon DIGITAL IXUS 60"       => 5.76,
"Canon Canon DIGITAL IXUS 65"       => 5.76,
"Canon Canon DIGITAL IXUS 700"       => 7.176,
"Canon Canon DIGITAL IXUS 750"      => 7.176,
"Canon Canon DIGITAL IXUS 800 IS"   => 5.76,
"Canon Canon DIGITAL IXUS II"       => 5.27,
"Canon Canon EOS 10D"               => 22.7,
"Canon Canon EOS-1D Mark II"        => 28.7,
```



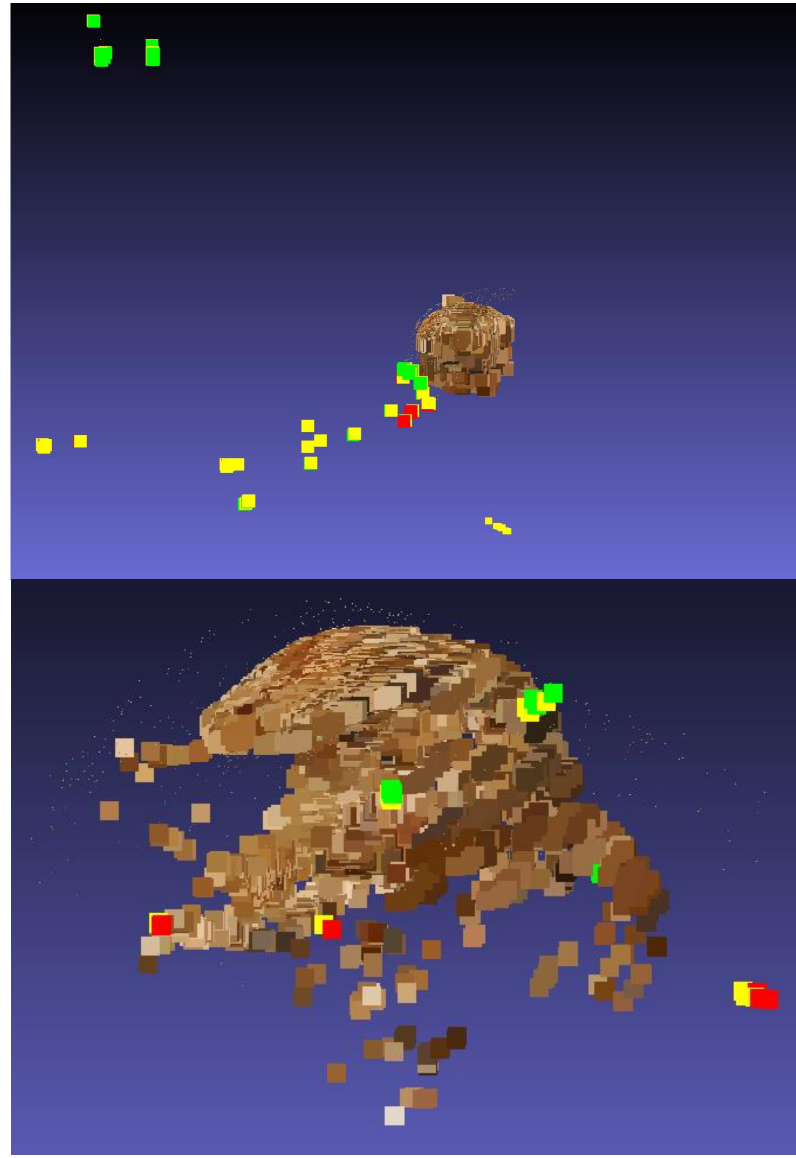
Focal Length Provided vs. Not

Skull, 24 images

Source: Furukawa & Ponce, 3D Photography Dataset.



Focal length provided.



Focal length not provided. Time: 5m7s

Reconstruction Quality: Camera Focal Length

- Why helpful? The optimization objective is a nonlinear least square:

$$\arg\min_{X_i, R_j, c_j, f_j} \sum_{i \sim j} \left\| x_{ij} - f_j \Pi\left(R_j (X_i - c_j)\right)\right\|^2.$$

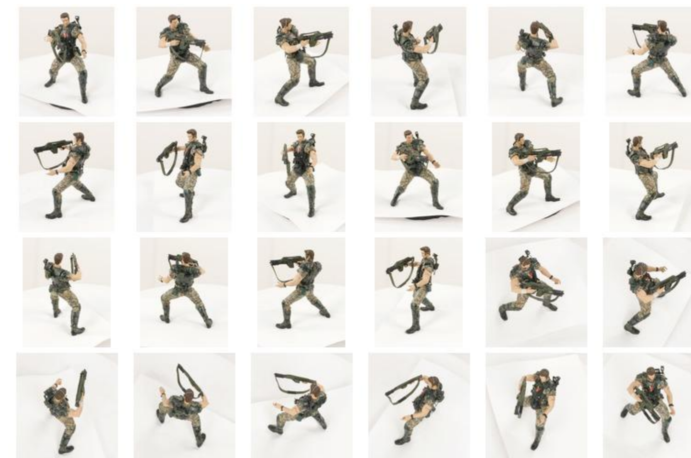
- For the original experiment, they use images both with or without this information, e.g., Notre Dame: 705 images (383 with focal length).

Outline

- A review of the method
- Reconstruction quality
  - How many images do we need?
  - How and why camera focal length help reconstruction
  - Number of keypoints
- Ambiguity: symmetry and repeated features
- More examples
- Computational cost breakdown

Reconstruction Quality: Keypoints

- Same number of images : 24 images
- Same camera angles
- Same background
- Different number of keypoints detected



Soldier: 1842±273 keypoints/image

Predator: 4663±1415 keypoints/image

Warrior: 2616±764 keypoints/image

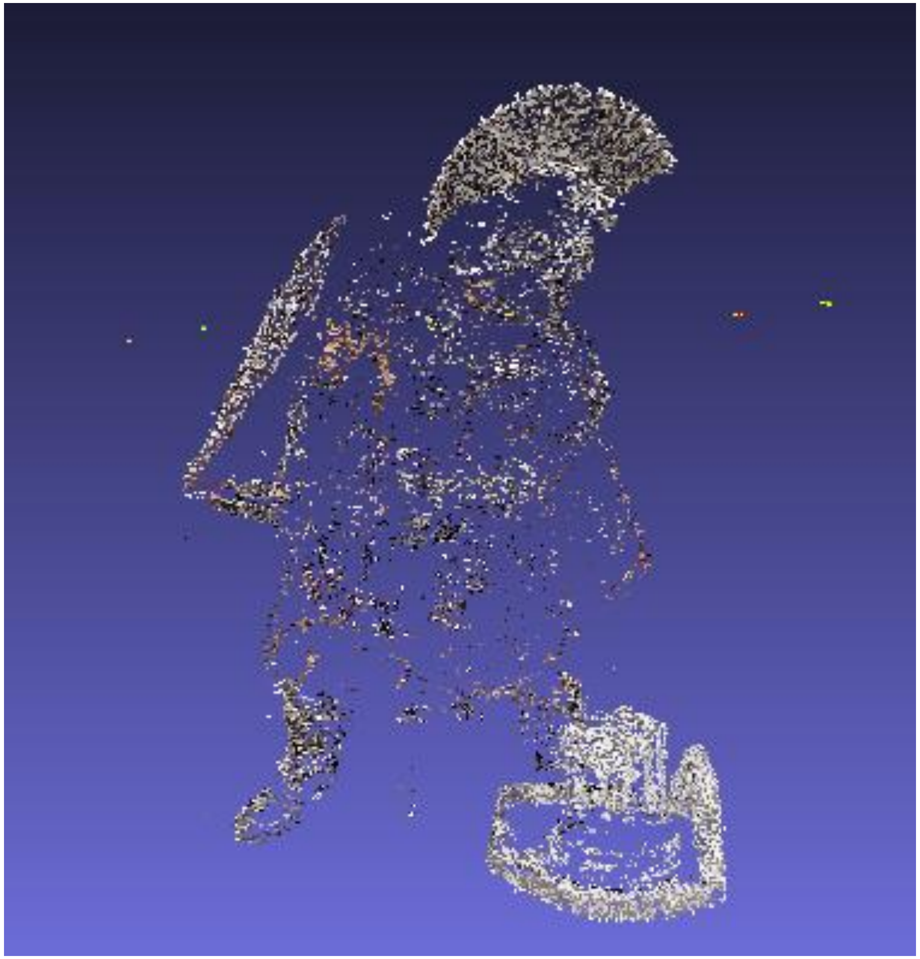
Source: Furukawa & Ponce, 3D Photography Dataset.

Solider: 1m56s
Warrior: 2m30s
Predator: 3m44s

Reconstruction Quality: Keypoints

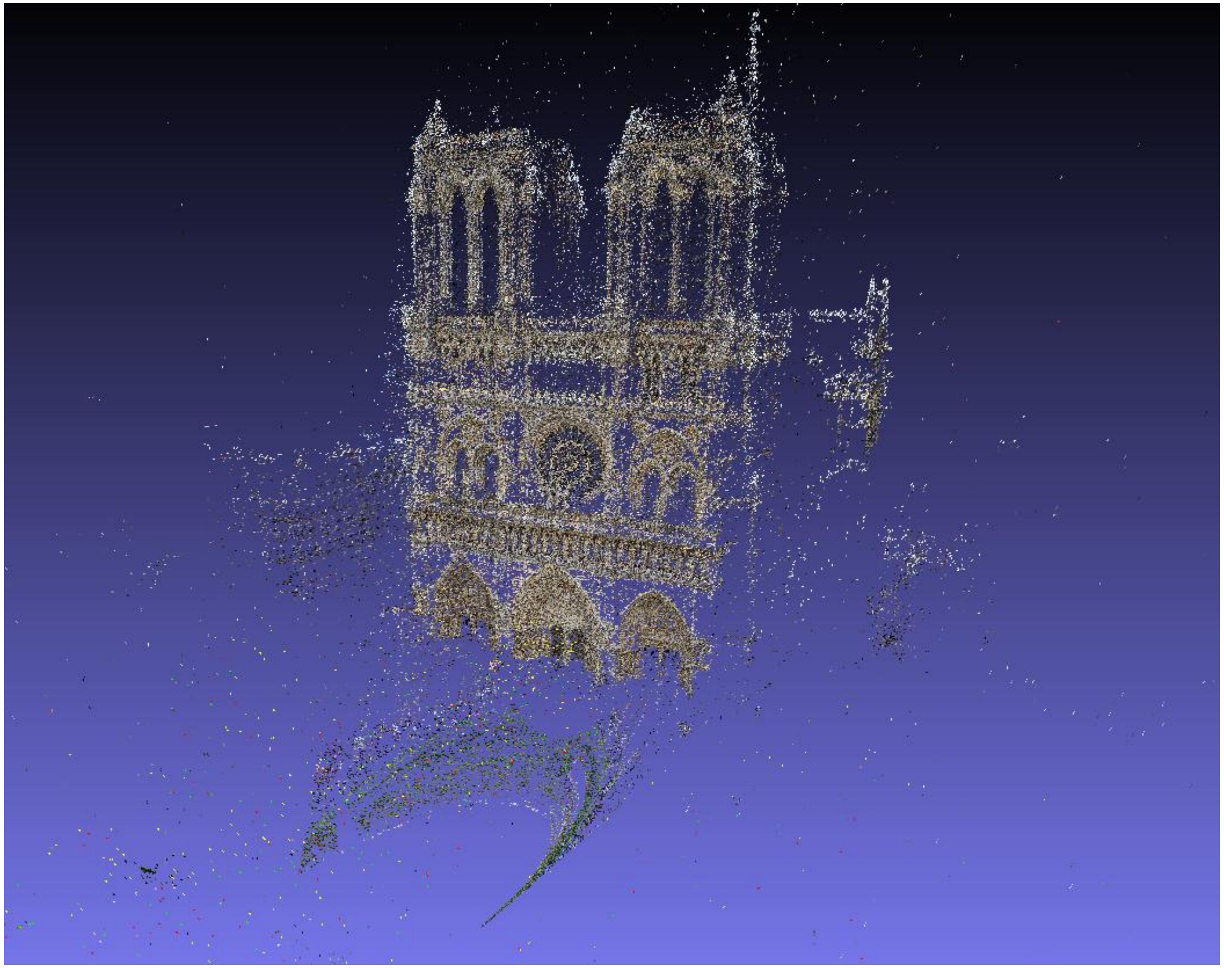
Armor: 48 images, 29407±12851 keypoints/image, 69min32s

Source: Lazebnik, et al., Visual Hull Data Sets.



(Demo)

Reconstruction Quality: Notre Dame



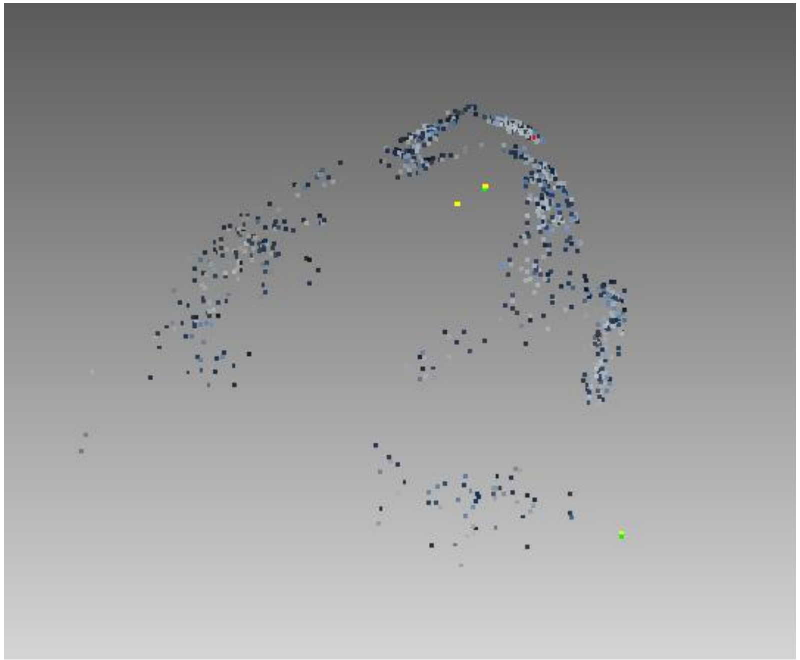
705 images (383 with focal length), 18760±16598 keypoints/frame, 5.625 days (Demo)

Source: Wilson & Snavely, Network principles for sfm: Disambiguating repeated structures with local context.

Outline

- A review of the method
- Reconstruction quality
  - How many images do we need?
  - How and why camera focal length help reconstruction
  - Number of keypoints
- Ambiguity: symmetry and repeated features
- More examples
- Computational cost breakdown

Ambiguity: Symmetry and Repeated Features

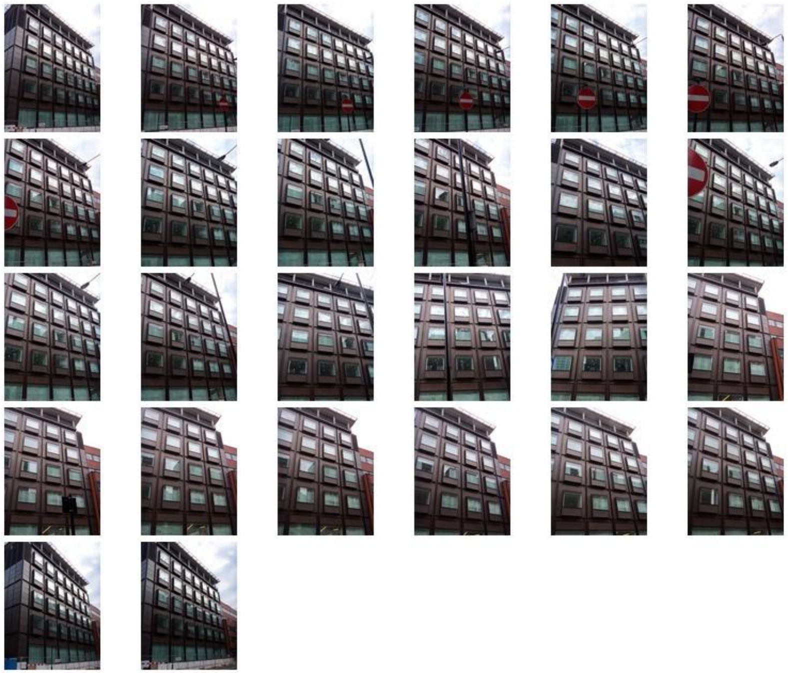


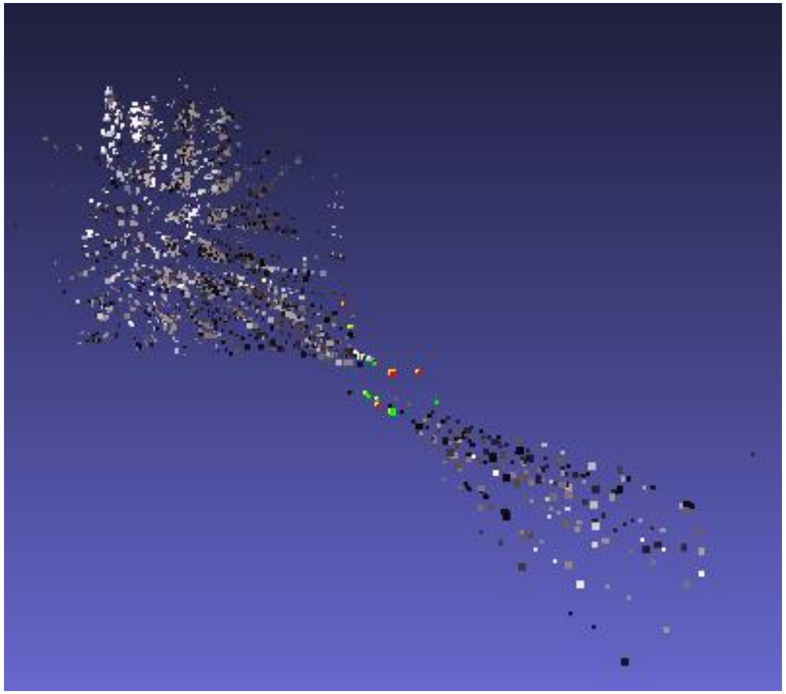
Bear: 20 images, 5773 ± 751 keypoints/image, 3m42s

Does ratio test help?

Source: Hao et al., Efficient 2D-to-3D Correspondence Filtering for Scalable 3D Object Recognition.

Ambiguity: Symmetry and Repeated Features

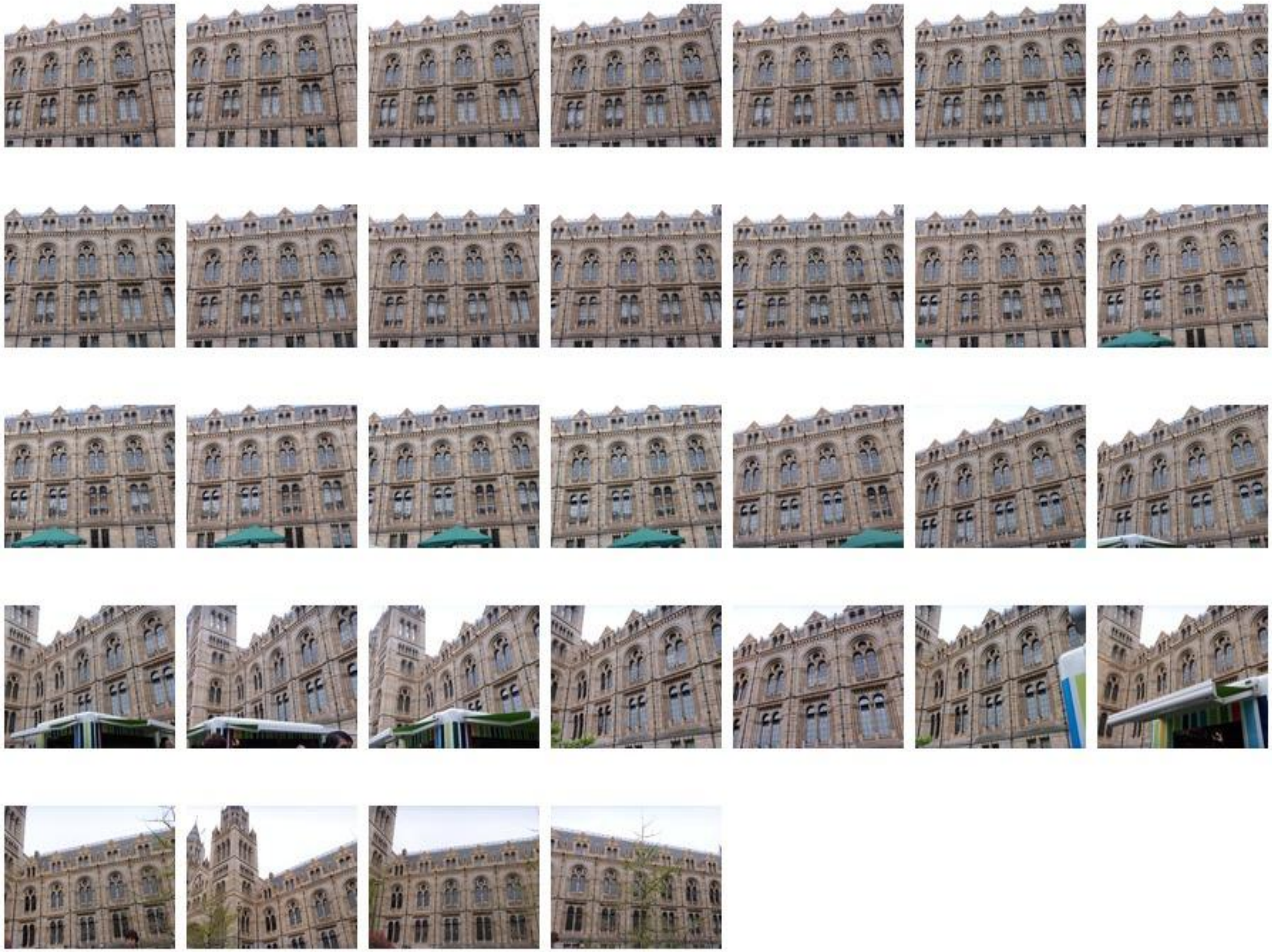




Building 1, 26 images, 18973±2513 keypoints/image, 12m29s

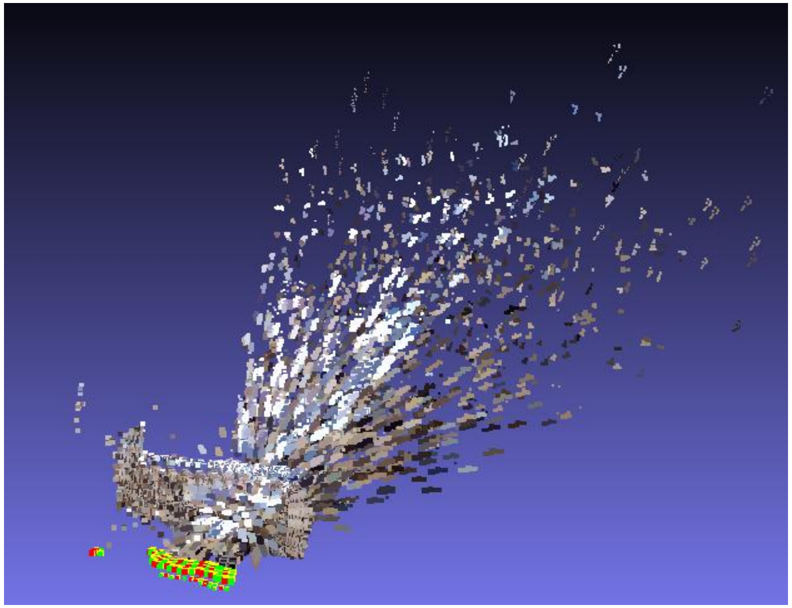
Source: Ceylan et al., Coupled structure-from-motion and 3D symmetry detection for urban facades.

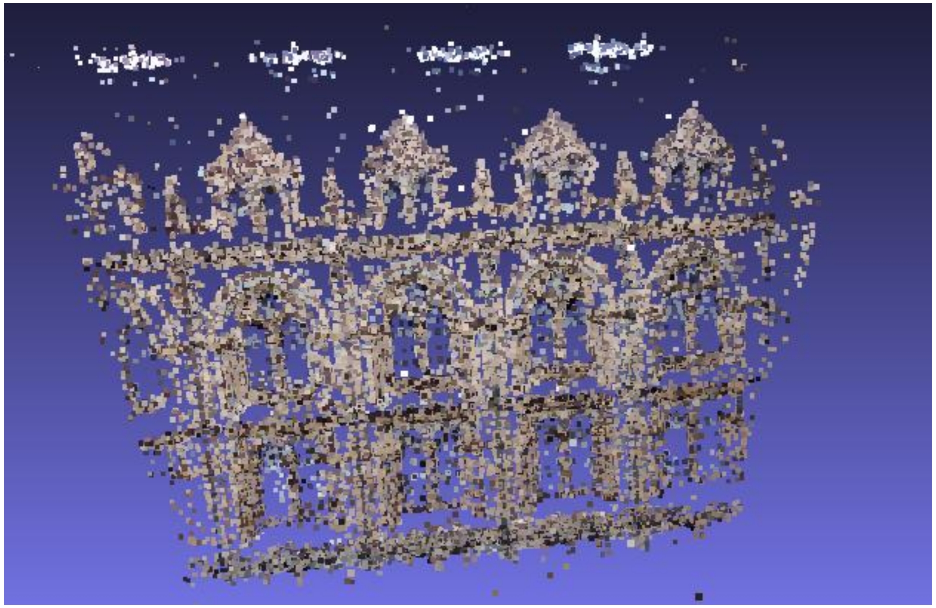
Ambiguity: Symmetry and Repeated Features



Source: Ceylan et al., Coupled structure-from-motion and 3D symmetry detection for urban facades.

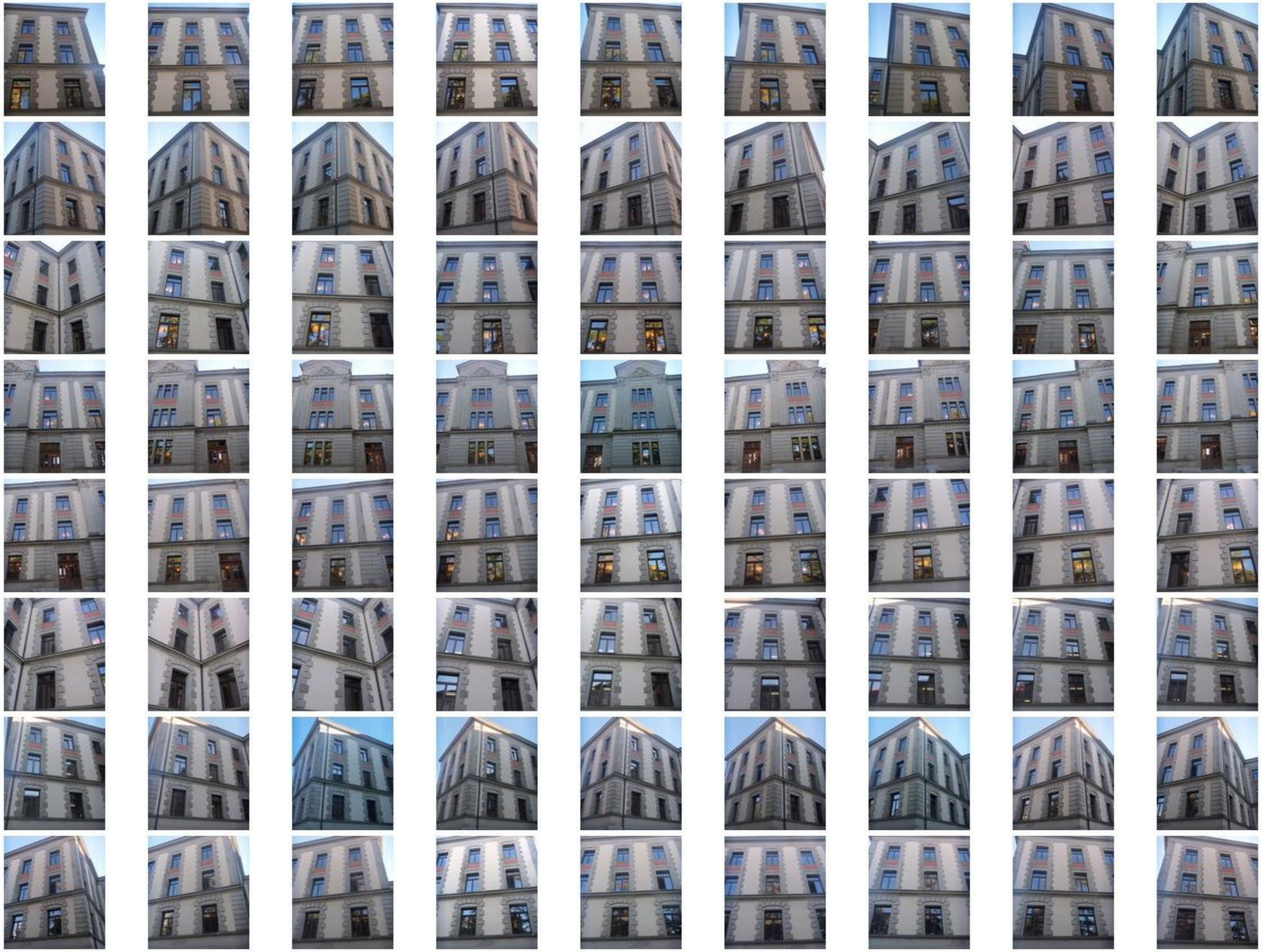
Ambiguity: Symmetry and Repeated Features





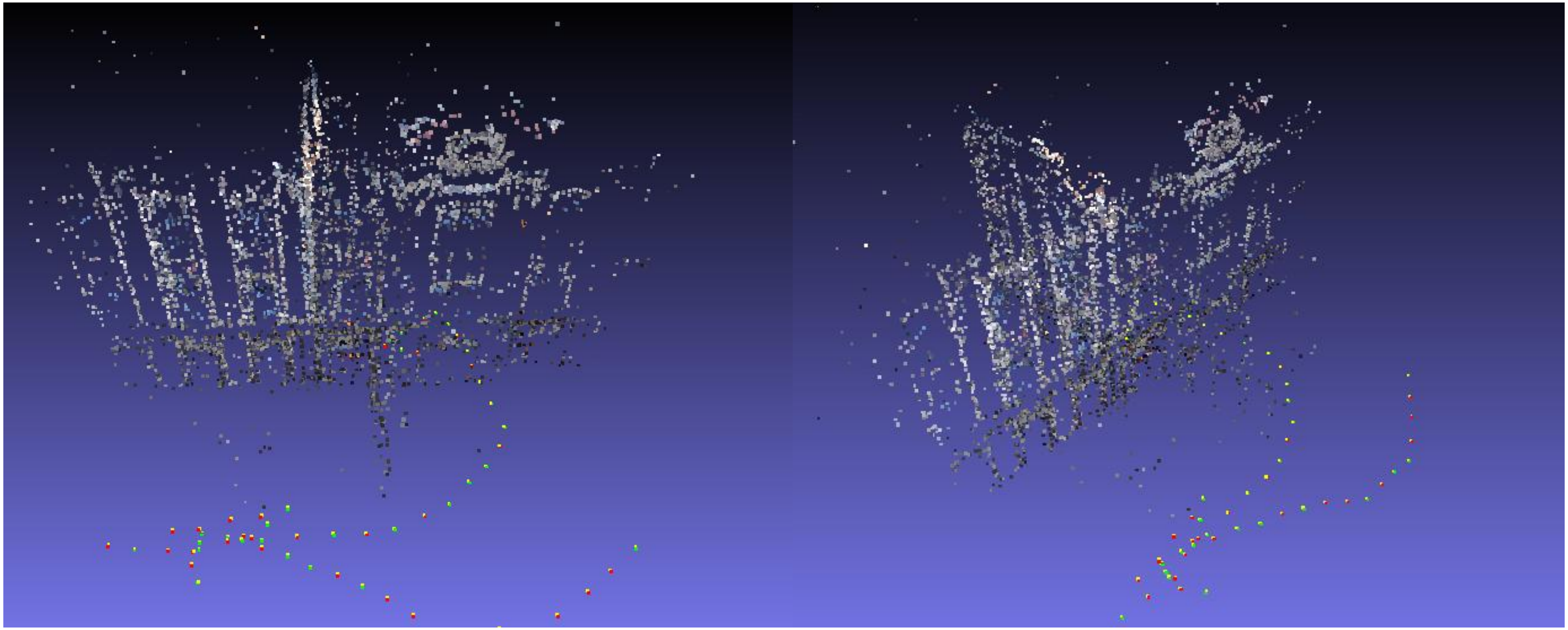
Building 6, 32 images, 56324±6941 keypoints/image, 67m54s

Ambiguity: Symmetry and Repeated Features



Source: Ceylan et al., Coupled structure-from-motion and 3D symmetry detection for urban facades.

Ambiguity: Symmetry and Repeated Features

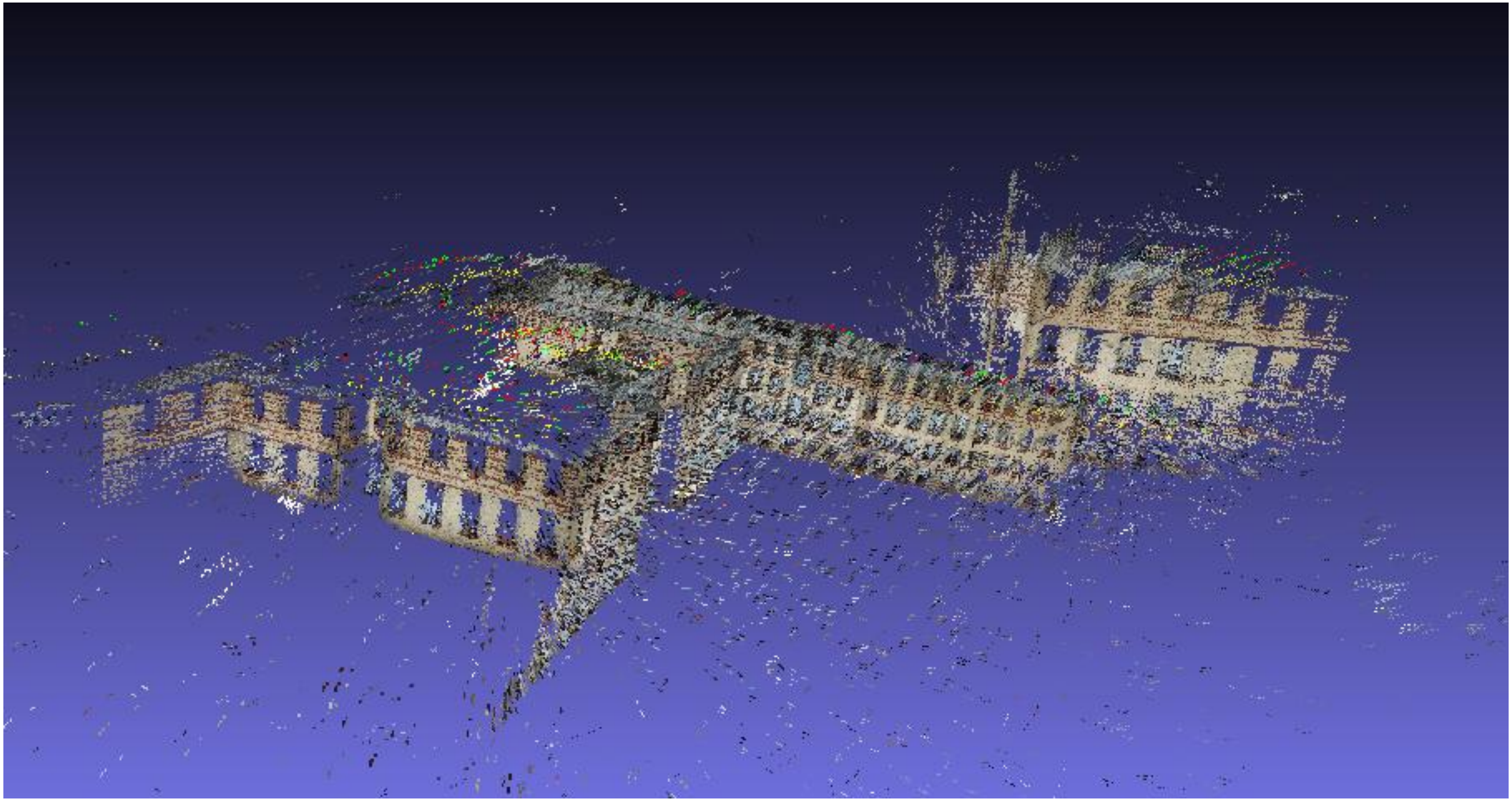


Buildings 8, 72 images, 9283±2977 keypoints/image, 39m30s.
Note the two walls that are misplaced.

Ambiguity: Symmetry and Repeated Features

Source: Cohen et al., Discovering and exploiting 3d symmetries in structure from motion.

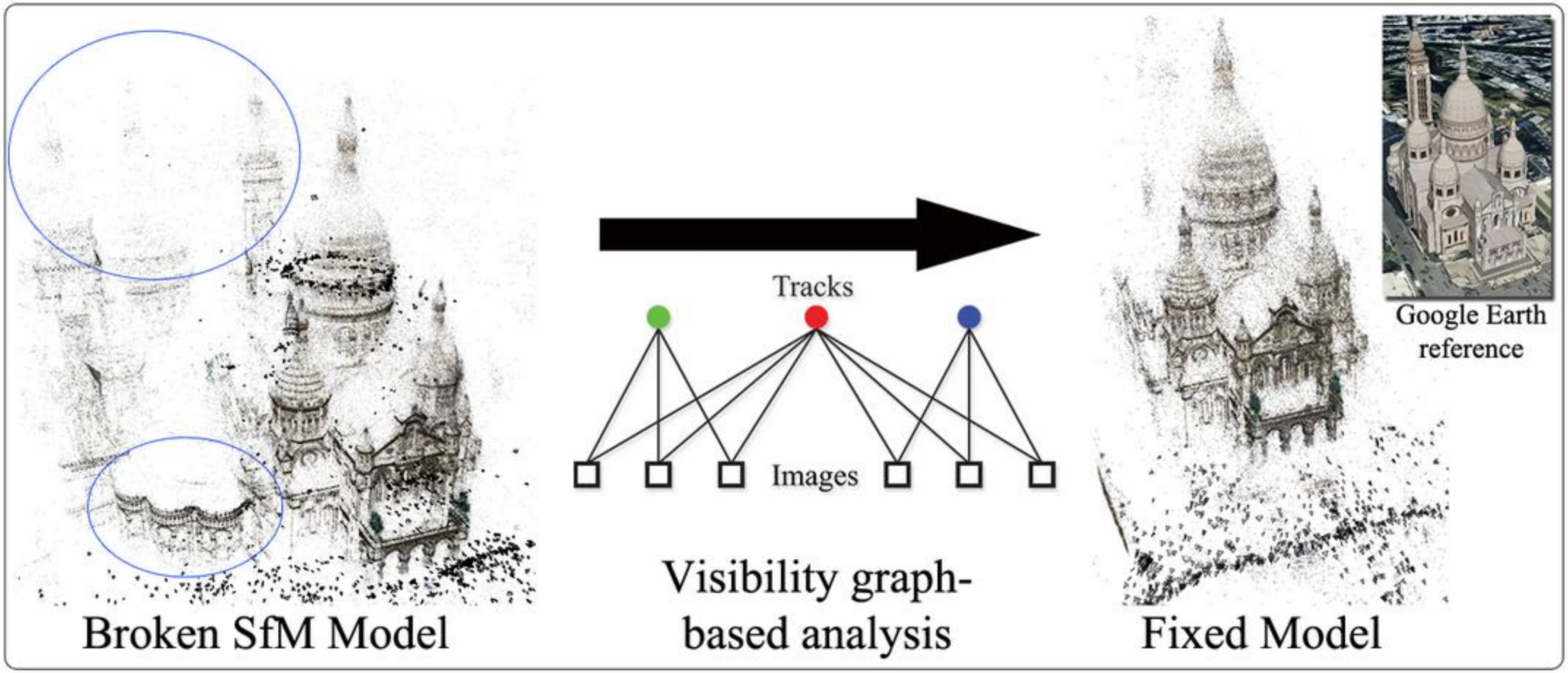
Ambiguity: Symmetry and Repeated Features



Street, 312 images, 14144 ± 5145 keypoints/image, 997m31s

Disambiguation

Network Principles for SfM: Disambiguating Repeated Structures with Local Context

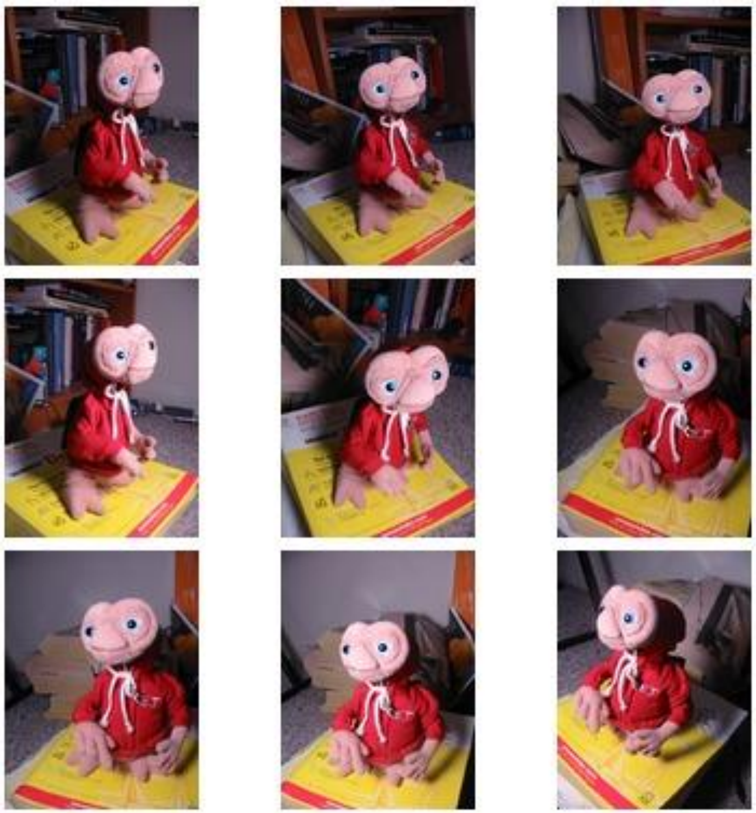


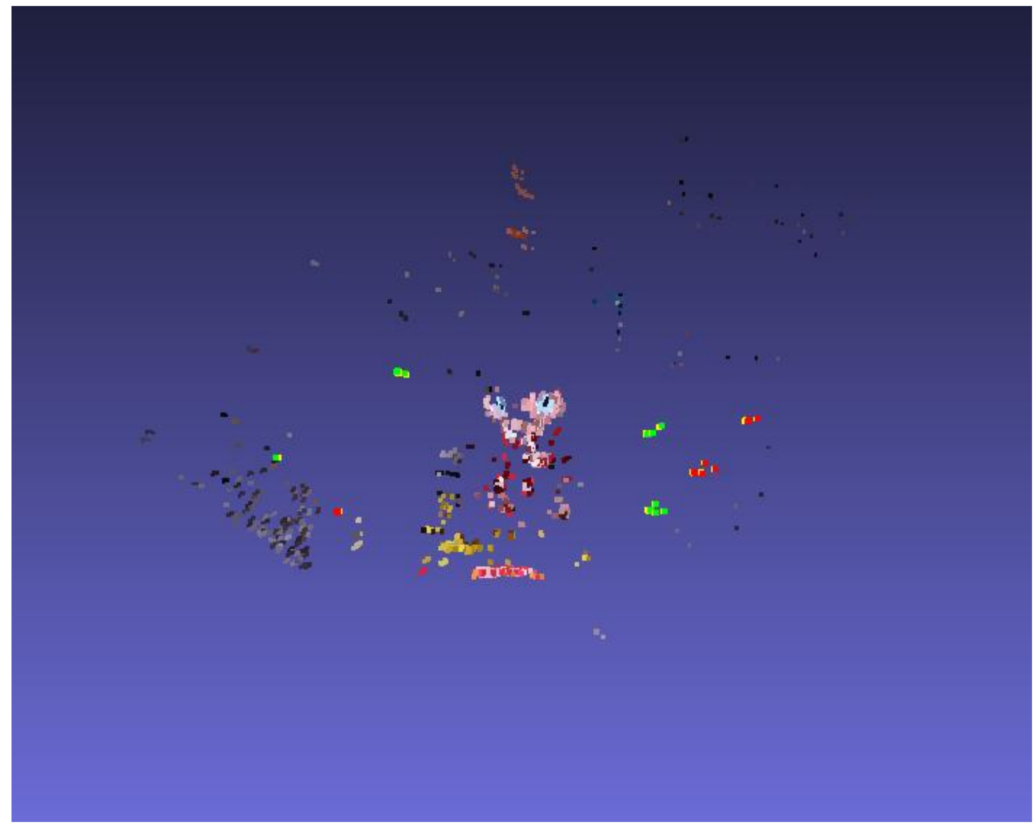
Source: Wilson & Snavely, Network principles for sfm: Disambiguating repeated structures with local context.

Outline

- A review of the method
- Reconstruction quality
  - How many images do we need?
  - How and why camera focal length help reconstruction
  - Number of keypoints
- Ambiguity: symmetry and repeated features
- More examples
- Computational cost breakdown

More Examples: ET

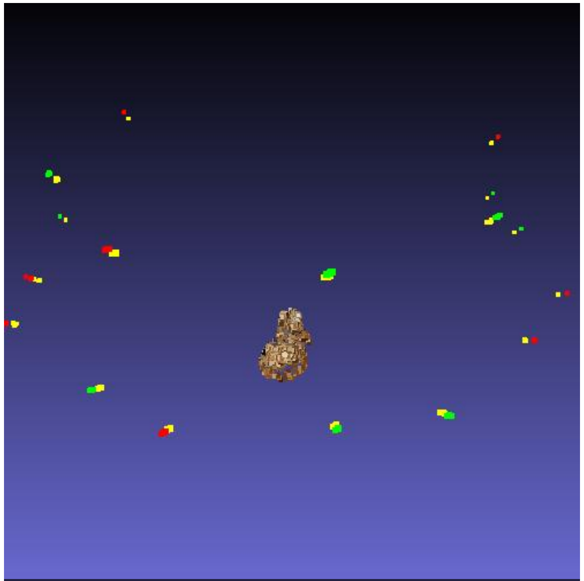


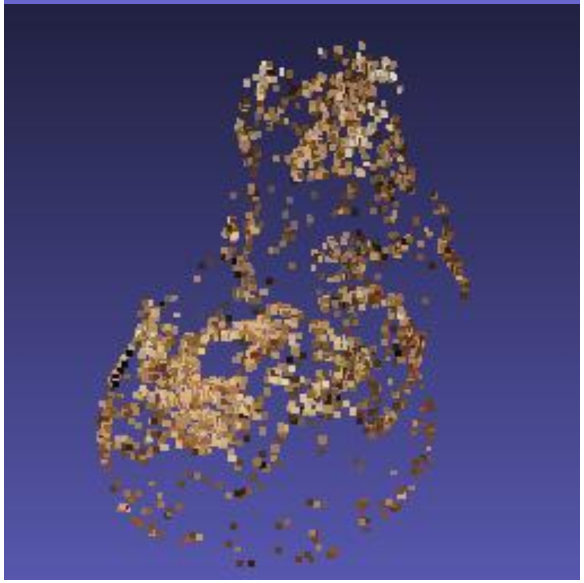


ET: 9 images, 1178±243 keypoints/image, 13s

Source: Snavely, Bundler: Structure from Motion (SfM) for Unordered Image Collections.

More Examples: Skull2





Skulls2, 24 images, 6324 ± 1778 keypoints/image, 5m24s

Source: Furukawa and Ponce, 3D Photography Dataset.

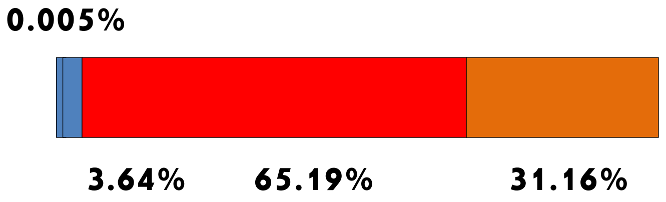
Outline

- A review of the method
- Reconstruction quality
  - How many images do we need?
  - How and why camera focal length help reconstruction
  - Number of keypoints
- Ambiguity: symmetry and repeated features
- More examples
- Computational cost breakdown

Computational Cost

- Number of keypoints
- Number of images

- Breakdown
  - Extract camera info from images
  - Keypoints detection
  - Pairwise keypoints matching (match graph, a key contribution)
  - SFM



- Hardware
  - Intel Core i7-5820K CPU 3.30GHZ x 12
  - 32 GB Memory
  - Geforce GTX 960

References and Resources

[1] Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., & Szeliski, R. (2011). Building rome in a day. *Communications of the ACM*, *54*(10), 105-112.

[2] 3D Photography Dataset. Yasutaka Furukawa and Jean Ponce. Beckman Institute and Department of Computer Science, University of Illinois at Urbana-Champaign. http://www-cvr.ai.uiuc.edu/ponce_grp/data/mview/

[3] Visual Hull Data Sets. Svetlana Lazebnik, Yasutaka Furukawa and Jean Ponce. Beckman Institute and Department of Computer Science, University of Illinois at Urbana-Champaign. http://www-cvr.ai.uiuc.edu/ponce_grp/data/visual_hull/index.html

[4] Ceylan, D., Mitra, N. J., Zheng, Y., & Pauly, M. (2014). Coupled structure-from-motion and 3D symmetry detection for urban facades. *ACM Transactions on Graphics (TOG)*, *33*(1), 2. Dataset: http://www.duygu-ceylan.com/duygu-ceylan/symmCalib.html

[5] Multiview Stereo Evaluation Dataset. Steve Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Rick Szeliski. http://grail.cs.washington.edu/projects/mview/

[6] MSR-Object3D-300 Dataset. http://research.microsoft.com/en-us/projects/3d_reconstruction_recognition/3d_obj_recognition.aspx. Qiang Hao, Rui Cai, Zhiwei Li, Lei Zhang, Yanwei Pang, Feng Wu, and Yong Rui. "Efficient 2D-to-3D Correspondence Filtering for Scalable 3D Object Recognition". in Proc. of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2013), pp.899-906, Portland, Oregon, USA. June 23-28, 2013.

[7] Bundler: Structure from Motion (SfM) for Unordered Image Collections. Noah Snavely. http://www.cs.cornell.edu/~snavely/bundler/

[8] MeshLab. http://meshlab.sourceforge.net/

[9] Wilson, K., & Snavely, N. (2013). Network principles for sfm: Disambiguating repeated structures with local context. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 513-520).

[10] Cohen, A., Zach, C., Sinha, S. N., & Pollefeys, M. (2012, June). Discovering and exploiting 3d symmetries in structure from motion. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 1514-1521). IEEE. Dataset: https://www.inf.ethz.ch/personal/acohen/papers/symmetryBA.php

More SFM datasets at http://riemenschneider.hayko.at/vision/dataset/index.php?filter=+sfm