Divide and Conquer: Efficient large-scale structure from motion using graph partitioning

Brojeshwar Bhowmick[1], Suvam Patra[1], Avishek Chatterjee[2], Venu Madhav Govindu[2], Subhashis Banerjee[1]

[1]Indian Institute of Technology Delhi, New Delhi, India

[2]Indian Institute of Science, Bengaluru, India

brojeshwar|suvam|suban@cse.iitd.ac.in, avishek|venu@ee.iisc.ernet.in

Divide and Conquer: Efficient large-scale structure from motion using graph partitioning

Brojeshwar Bhowmick[1], Suvam Patra[1], Avishek Chatterjee[2], Venu Madhav Govindu[2], Subhashis Banerjee[1]

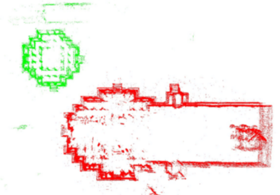
[1]Indian Institute of Technology Delhi, New Delhi, India

[2]Indian Institute of Science, Bengaluru, India

Introduction

- Contemporary large scale SfM methods use bundle adjustment.
- Reconstruction fails when:
  - Accumulated error in incremental reconstruction is large.
  - Number of 3D to 2D correspondences are insufficient.

(a) Reconstruction failure by VSFM.



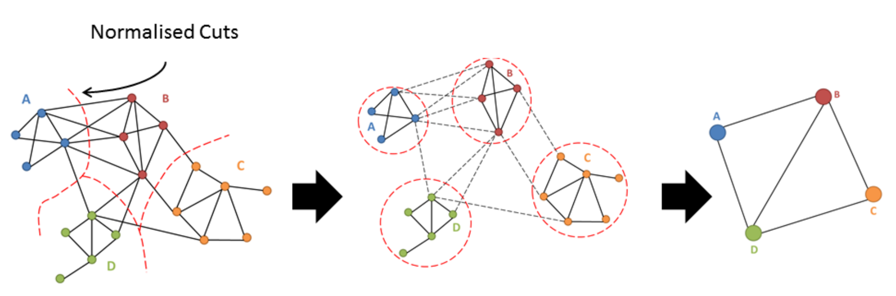
(b) Successful reconstruction by our method.

- Bundle adjustment is:
  - Computationally demanding.
  - Time consuming due to large number of images.

Our Contribution

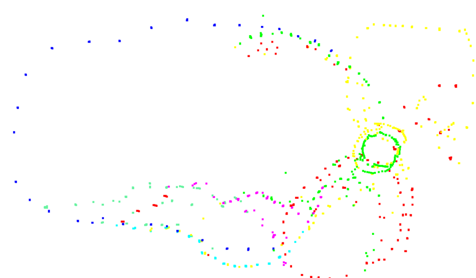
- Partition a large collection of images into disjoint connected components.
- Each component can be independently and reliably reconstructed.
- Identify connecting images between components to register the independent reconstructions.
- A method to register independent reconstructions using pairwise epipolar geometry.
- One order of magnitude speed improvement compared to state-of-the-art methods.

Dataset Decomposition



- Image acquisition from a site in an organised manner makes the problem of decomposition into smaller sets trivial.

- Images downloaded from the internet are referred to as unorganised images.
- Use multi-way normalised cut [3] to partition the match graph into individual components.



- The images that belong to a cut are used as connecting images.
- Each individual component is reconstructed in parallel using a sequential bundler [4].

Registration of Independent Component Reconstructions

- Let A and B be two independently reconstructed components, and $k \in \mathbb{C}_{AB}$ be a connecting camera between them.
- Let R_{Ak} and T_{Ak} denote the rotation and translation of camera k in the frame of reference of A
- For image $i \in A$, let R_{Ai} and T_{Ai} be the rotation of i in the frame of reference of A.

Scale Estimation between a Pair of Reconstructions:

- R_{ik} and t_{ik} are estimated from the epipolar relationship between i and k.

$$R_{ik} \; = \; R_{Ak}R_{Ai}^T \Rightarrow R_{Ak} = R_{ik}R_{Ai}$$

Translation directions are related as described in [1]

$$t_{ik} \propto T_{Ak} - R_{ik}T_{Ai} \Rightarrow [t_{ik}]_\times (T_{Ak} - R_{ik}T_{Ai}) = 0$$

- Compute averaged rotation [2] and translation as:

$$\widehat{R}_{Ak} = \underset{i \in A}{\text{mean}} \left(R_{ik}R_{Ai}\right)$$

$$\widehat{T}_{Ak} = \underset{T_{Ak}}{\text{argmin}} \sum_{i \in A} \frac{\left\| [t_{ik}]_\times (T_{Ak} - R_{ik}T_{Ai}) \right\|^2}{\| T_{Ak} - R_{ik}T_{Ai} \|^2}$$

- Scale is calculated as:

$$\widehat{s}_{AB} = \underset{k_1, k_2 \in \mathbb{C}_{AB}}{\text{median}} \frac{\left\| -\widehat{R}_{Bk_1}\widehat{T}_{Bk_1} + \widehat{R}_{Bk_2}\widehat{T}_{Bk_2} \right\|}{\left\| -\widehat{R}_{Ak_1}\widehat{T}_{Ak_1} + \widehat{R}_{Ak_2}\widehat{T}_{Ak_2} \right\|}$$

Relative Rotation and Translation Estimation between Two Reconstructions:

- Using single epipolar relationship, rotation and translation between two reconstructions can be found as:

$$R_{AB} = R_B R_A^T = \widehat{R}_{Bk}^T \widehat{R}_{Ak}$$

$$T_{AB} = T_B - R_B R_A^T T_A = \widehat{s}_{AB} \widehat{R}_{Bk}^T \widehat{T}_{Ak} - \widehat{R}_{Bk}^T \widehat{T}_{Bk}$$

- As the above relations holds for all k,

$$\widehat{R}_{AB} = \underset{k \in \mathbb{C}_{AB}}{\text{mean}} \left(\widehat{R}_{Bk}^T \widehat{R}_{Ak} \right)$$

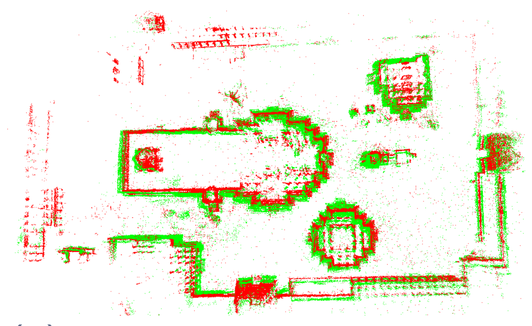
$$\widehat{T}_{AB} = \underset{T}{\text{argmin}} \sum_{k \in \mathbb{C}_{AB}} \left\| T - \left(\widehat{s}_{AB} \widehat{R}_{Bk}^T \widehat{T}_{Ak} - \widehat{R}_{Bk}^T \widehat{T}_{Bk} \right) \right\|_1$$

Experimental Results

- **Datasets:**

Dataset	No. of images	No. of components	No. of images reconstructed
Rome	13783	24	10534
Hampi	3017	7	2584
St Peter's Basilica	1275	5	1236
Colosseum	1164	3	1032

- Hampi dataset

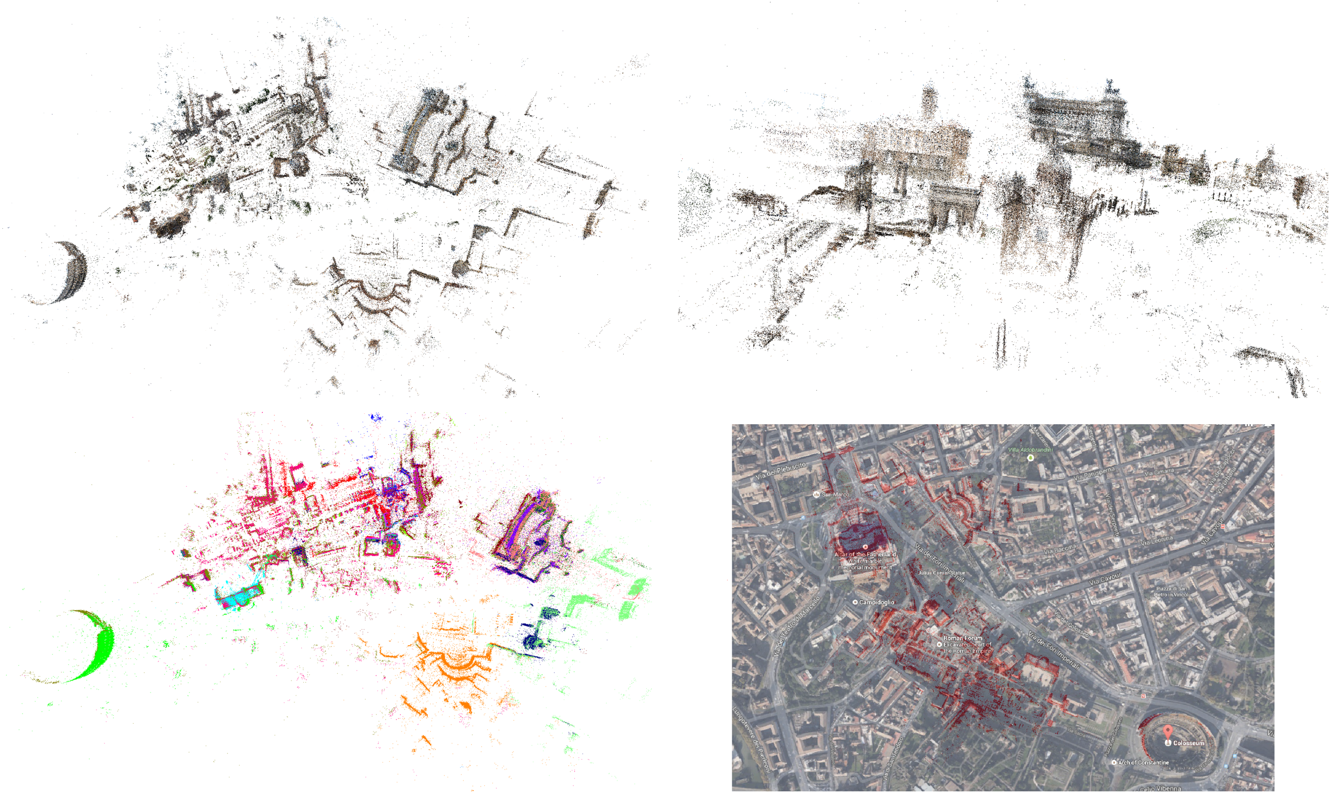


(a) Comparison with VSFM (red) and our method (green).

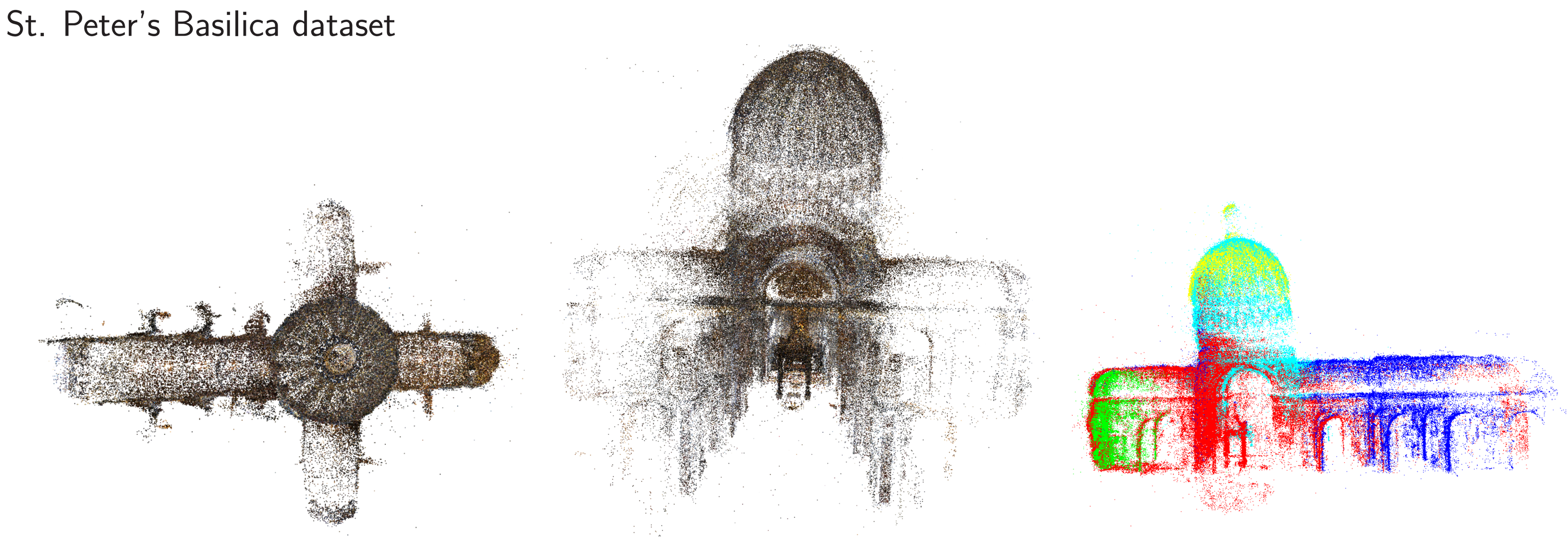
(b) Comparison with VSFM (red) and our method with epipolar robustness (green).

(c) Overlaid on Google map.

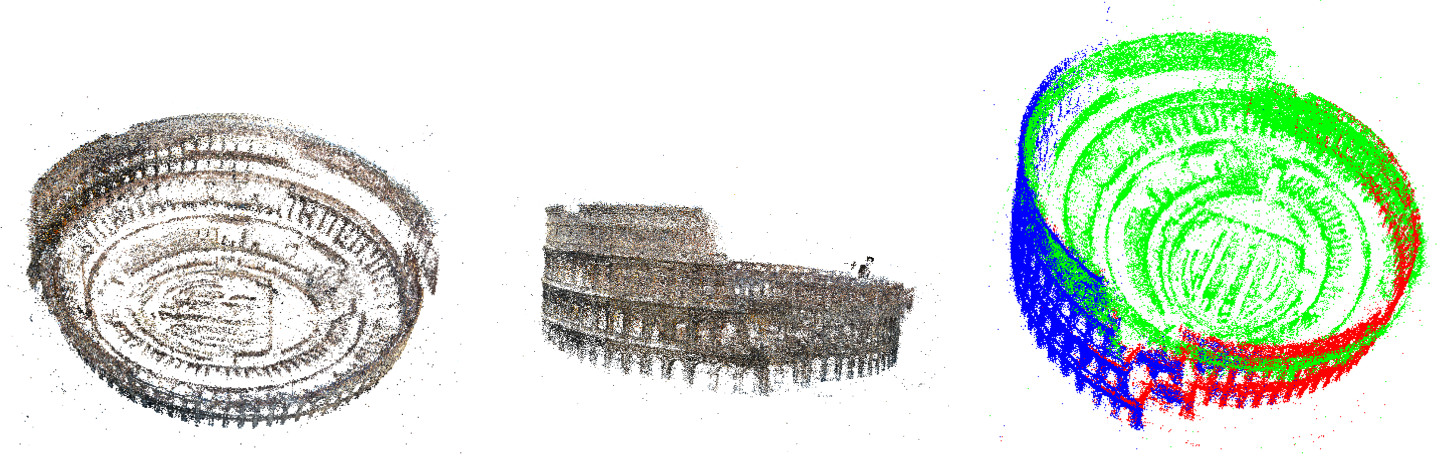
- Central Rome dataset



- St. Peter's Basilica dataset



- Colosseum dataset



Time Comparison

Dataset	Match graph creation using vocabulary tree (mins)	Pairwise matching (mins)	Reconstruction and registration (mins)	Total time by us (mins)	Pairwise matching by VSFM (mins)	Reconstruction by VSFM (mins)	Total time by VSFM (mins)
Rome	768	502	**27**	**1297**	N/A	N/A	N/A
Hampi	481	424	**8**	**913**	9522	59	9581
St Peter's Basilica	98	22	**4**	**124**	1385	10	1395
Colosseum	83	24	**3**	**110**	1394	9	1403

Comparison of our Method against VisualSFM for Hampi Dataset

Error entity	Error unit	Mean error	Median error	RMS error
Camera rotation	Degrees	1.93	1.57	2.66
Camera translation	Ratio of graph diameter	0.012	0.0091	0.041

References

[1] V. M. Govindu.
Combining two-view constraints for motion estimation.
In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 218–225, 2001.

[2] V. M. Govindu.
Lie-algebraic averaging for globally consistent motion estimation.
In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[3] J. Shi and J. Malik.
Normalized cuts and image segmentation.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8):888–905, 2000.

[4] C. Wu.
Towards linear-time incremental structure from motion.
In *Proceedings of the International Conference on 3D Vision*, 3DV '13, pages 127–134, 2013.

brojeshwar|suvam|suban@cse.iitd.ac.in, avishek|venu@ee.iisc.ernet.in