Towards Linear-time Incremental Structure from Motion - Supplemental Material

The supplemental material includes

- https://www.youtube.com/watch?v=CsLhDaNS_xE demonstrates the sparse reconstructions for $r = 5\%, r' = 25\%$.
- **Thresholds** for **squared** reprojection errors are
  - 320 for creating new points and projections during normal triangulation;
  - 3200 for creating new points and projections during RT;
  - 8 for point filtering.
- **The remainder** of this document presents additional details of our experiments.
  1. The normalized reconstruction time when considering different density q/n.
  2. How the reprojection errors change during the reconstruction;
  3. How p and q increases with respect to n during the reconstruction;
  4. Qualitative comparison of the reconstructions under various settings.

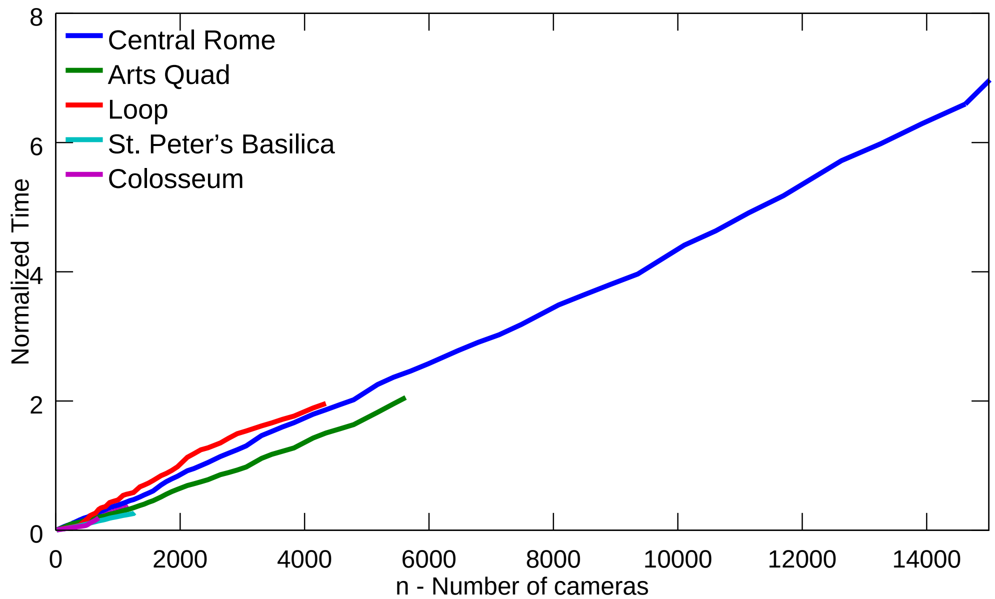


Figure 1: Normalized reconstruction time by multiplying the reconstruction time with the final n/q. Since the observations of length q is the longest data vector, the time complexity is determined by how the observation vector is accessed. It can be seen that the three large datasets have similar slopes for the normalized reconstruction time.

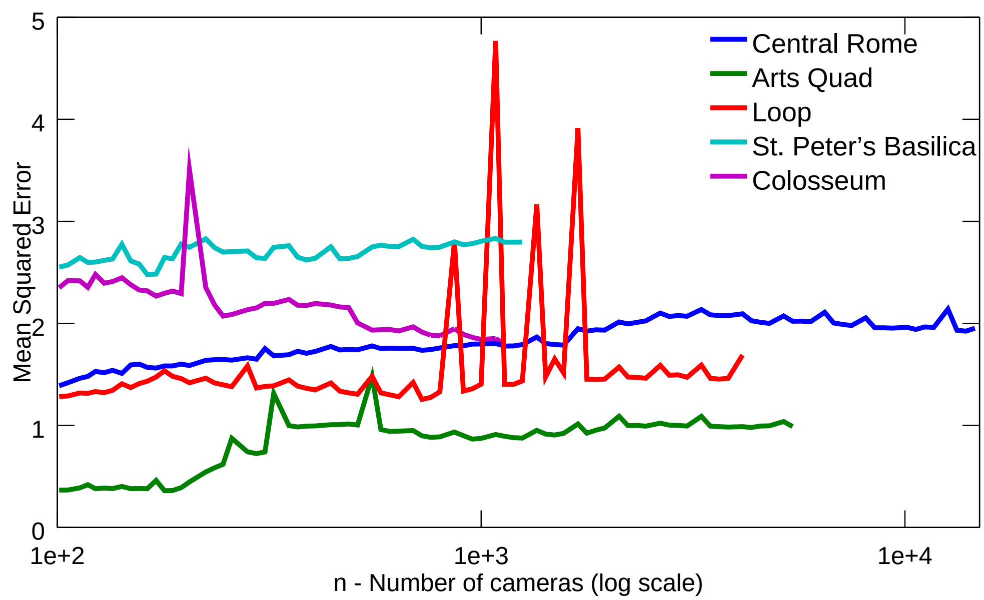


Figure 2: Mean squared error (MSE) after each full BA. The reprojection errors stay in a reasonable small range when the 3D models grow larger. Note that the spikes of MSE are caused by the re-triangulated matches in the RT step, where a larger threshold is used.

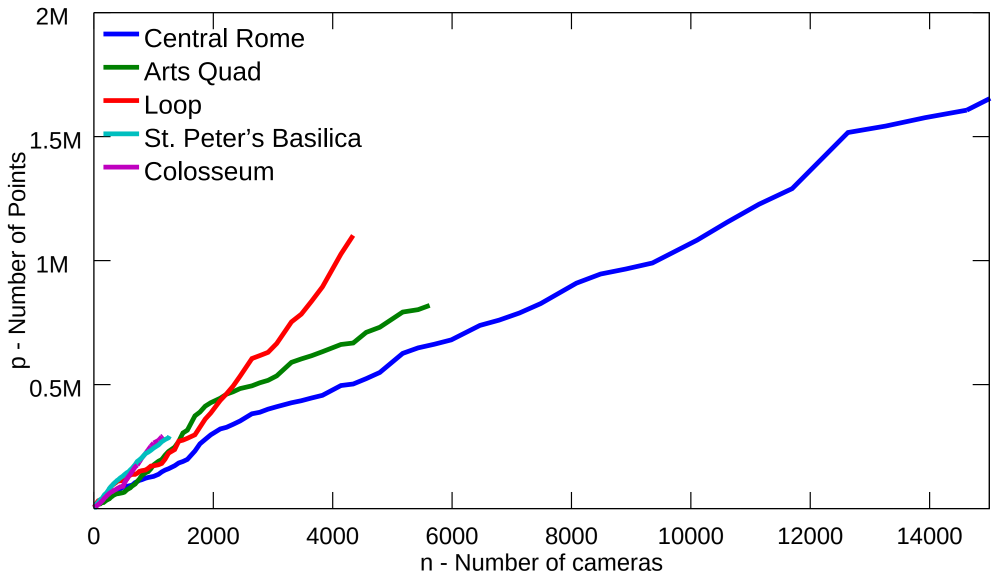


Figure 3: Number of 3D points increases as more cameras are added to a 3D model. The slope of the curve depends on the amount of new structures each camera observes.

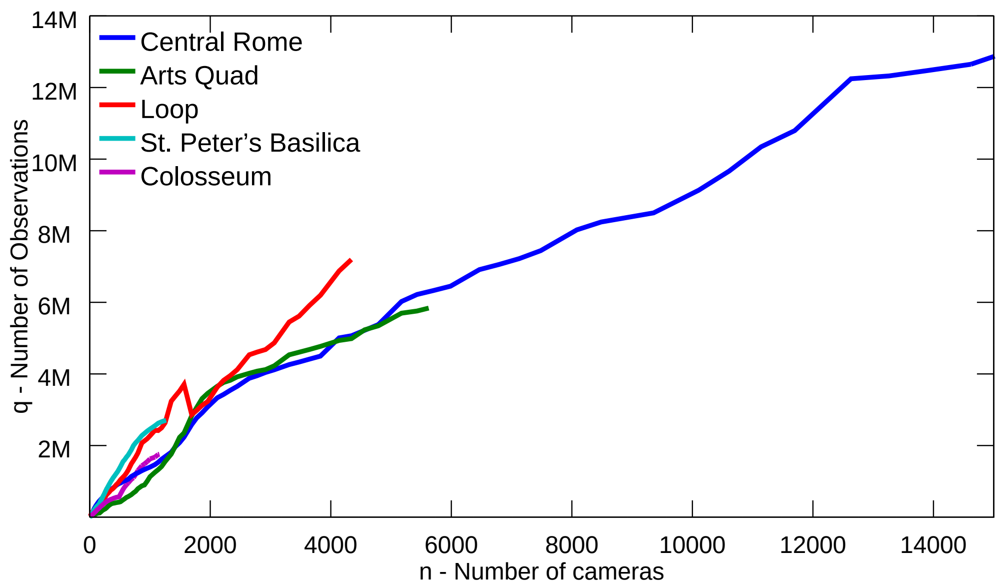
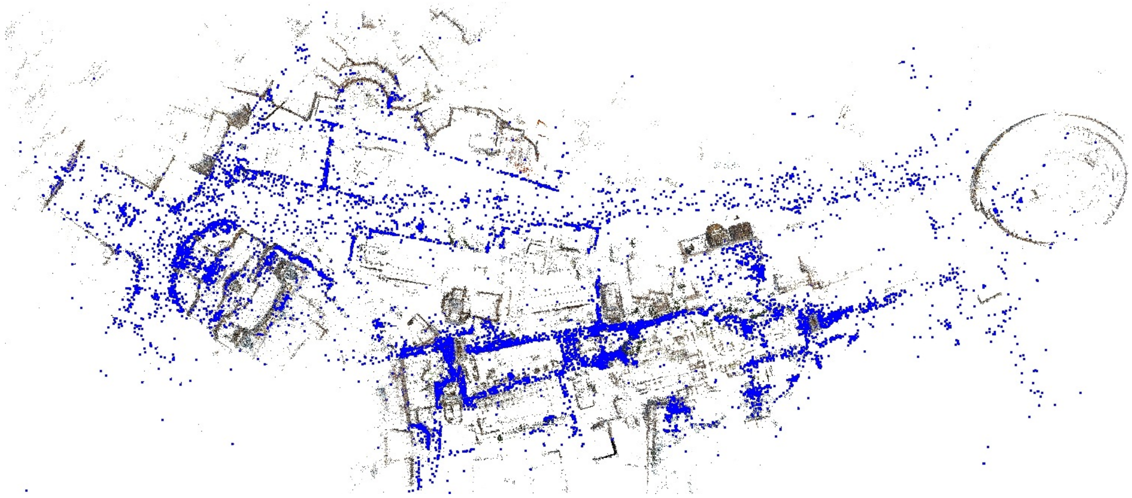
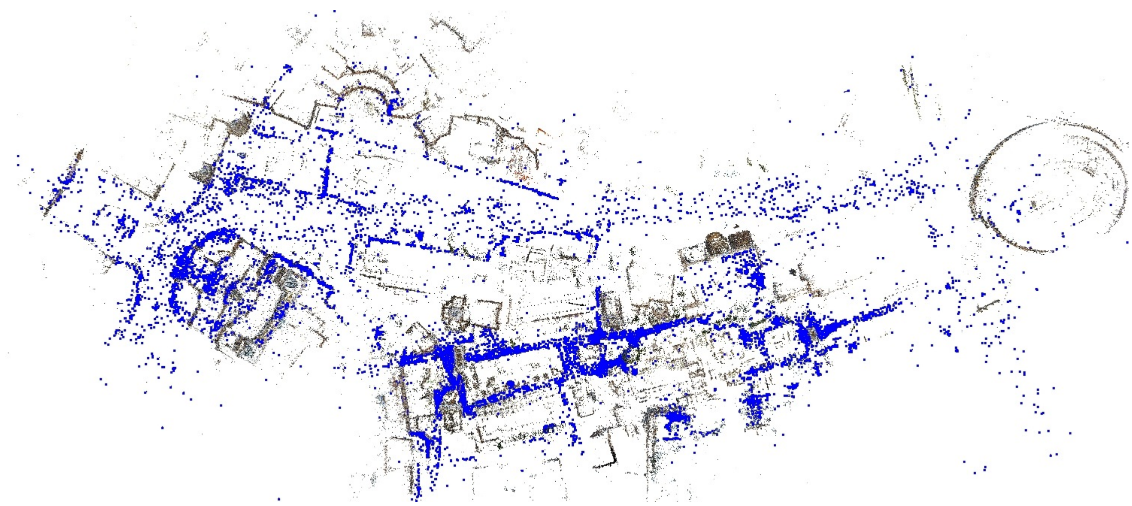


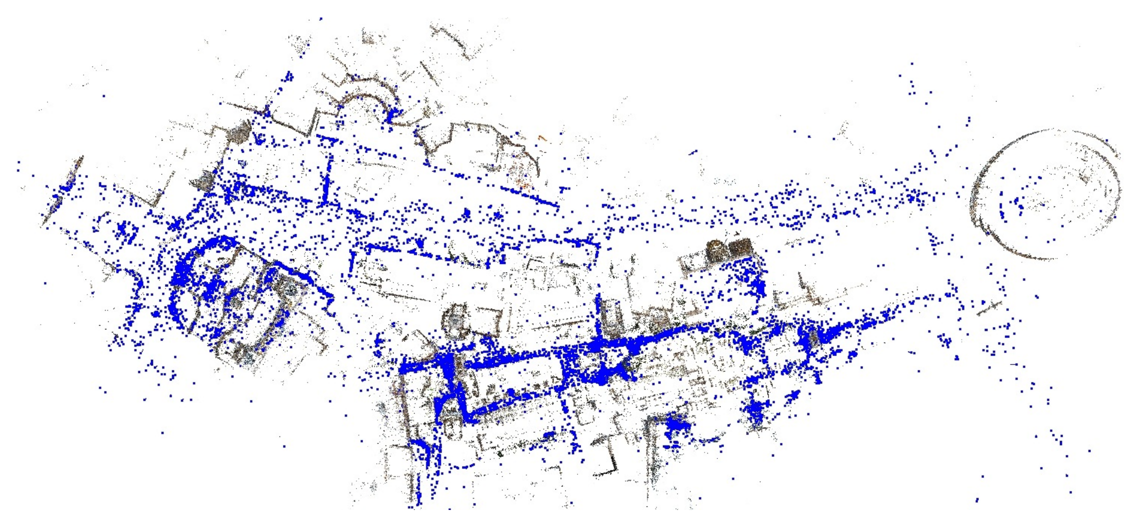
Figure 4: Number of observations. The slope of the curve depends on the number of feature matches found for each image. Due the greedy nature of incremental SfM, the speed of the increase of observations normally gets lower towards the end of the reconstruction. The small reduction of observations in the Loop reconstruction is due to the filtering step after RT. The loop reconstruction has a small issue with two closeby duplicated structures (traffic signs) in the scene.



(a) $r = 5\%, r' = 25\%$, run partial BA every image. 15065 cameras are reconstructed.

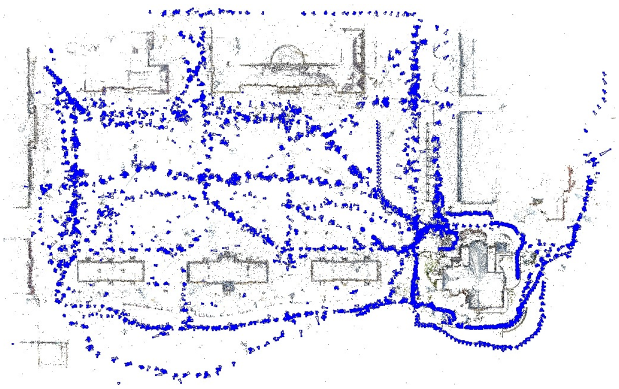


(b) $r = 25\%, r' = 50\%$, run partial BA every image. 15113 cameras are reconstructed

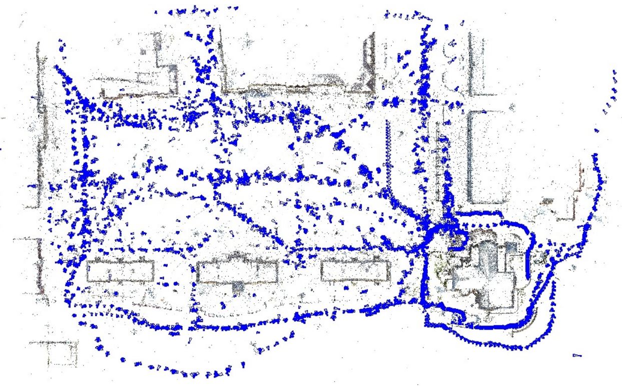


(c) $r = 5\%, r' = 25\%$, run partial BA every 3 images. 14998 cameras are reconstructed.

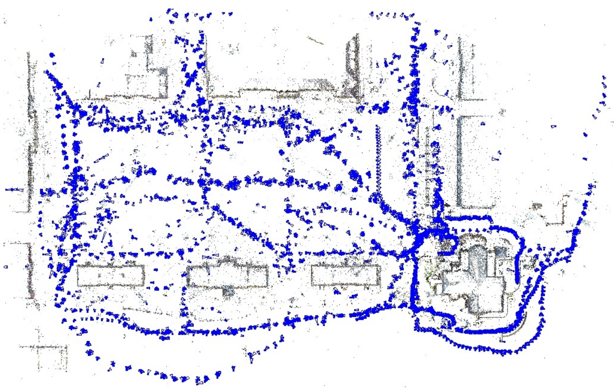
Figure 5: Top views of the Central Rome reconstructions. Under the various settings, the main structure of the reconstructions is stable despite small differences in some regions.



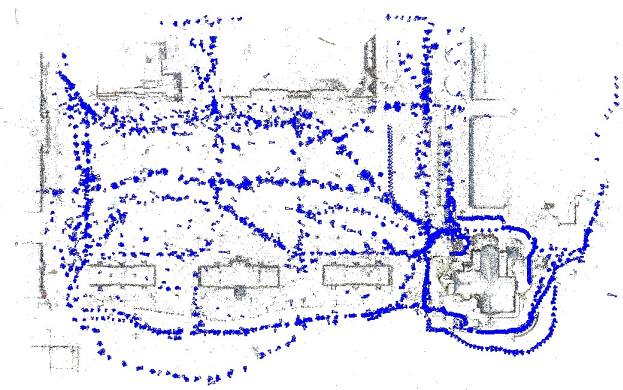
(a) $r = 5\%, r' = 25\%$;
Run partial BA every image.
5624 cameras are reconstructed.



(b) $r = 25\%, r' = 50\%$;
Run partial BA every image.
5598 cameras are reconstructed

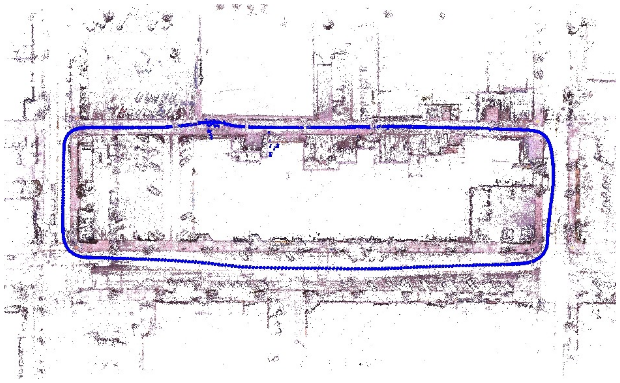


(c) $r = 5\%, r' = 25\%$;
Run partial BA every 3 images.
5461 cameras are reconstructed.

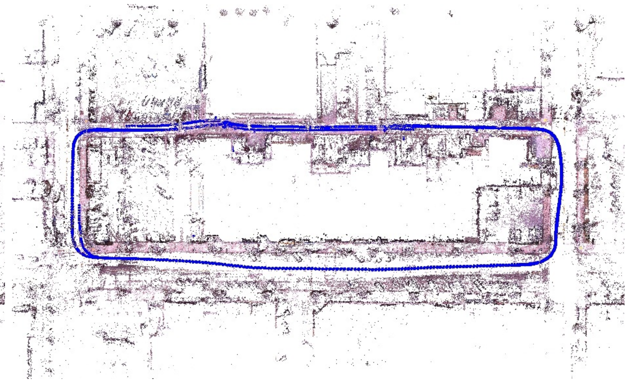


(d) $r = 5\%, r' = 25\%$;
Run partial BA every images.
Use preemptive matching $h = 100, t_h = 2$;
5393 cameras are reconstructed.

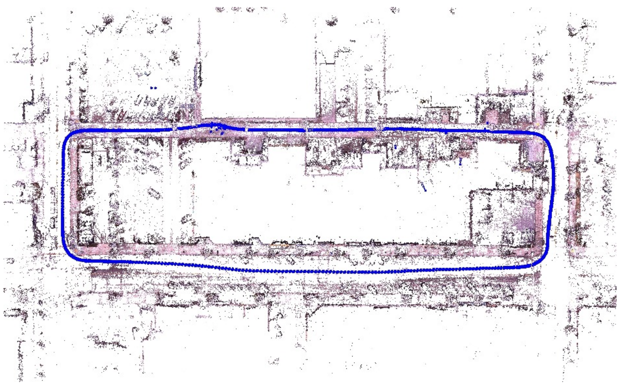
Figure 6: Top views of the Arts Quad reconstructions. Although the main structure of the reconstructions are stable, the reconstruction starts to get errors and ghost structures when we reduce the frequency of full BA and RT.



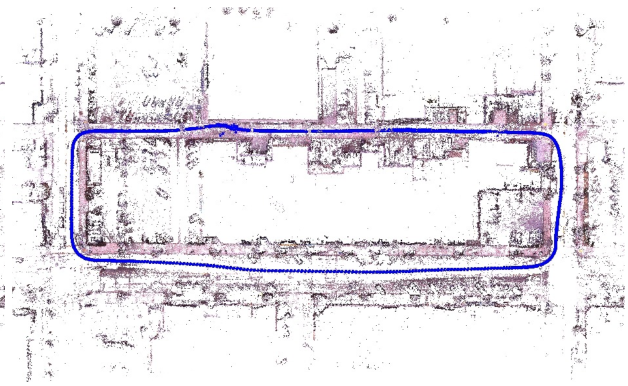
(a) $r = 5\%, r' = 25\%$;
Run partial BA every image.
4342 cameras are reconstructed.



(b) $r = 25\%, r' = 50\%$;
Run partial BA every image.
4342 cameras are reconstructed



(c) $r = 5\%, r' = 25\%$;
Run partial BA every 3 images.
4341 cameras are reconstructed.

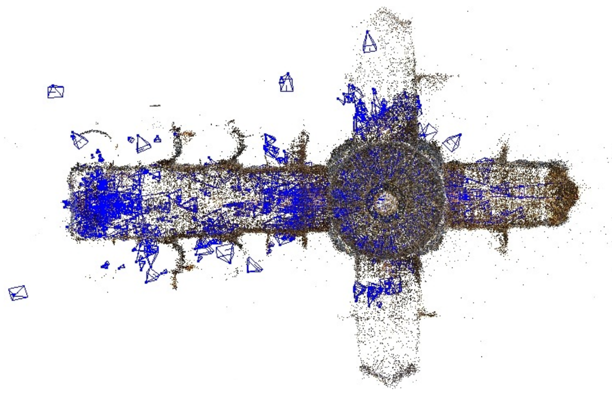


(d) $r = 5\%, r' = 25\%$;
Run partial BA every images.
Use preemptive matching $h = 100, t_h = 8$;
4342 cameras are reconstructed.

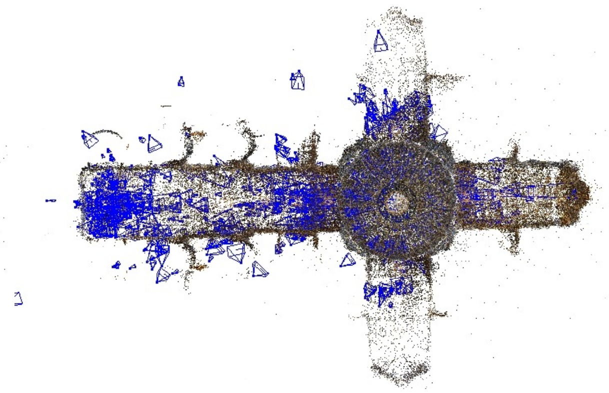
(e) $r = 5\%, r' = 25\%$.

(f) $r = 5\%$, without RT.

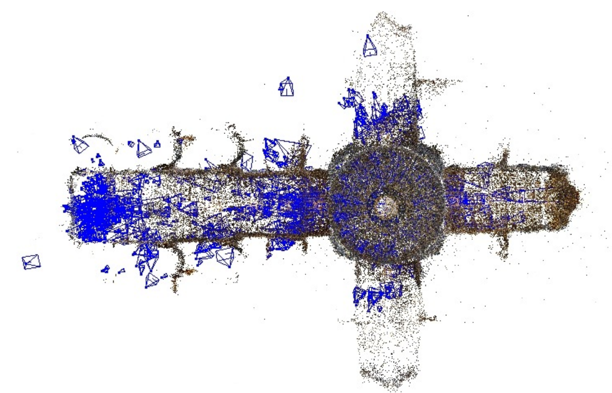
Figure 7: Top views of the Loop reconstructions. The few bad cameras in this reconstruction are caused by two closeby duplicated structures (traffic signs). As show in (b) and (f), there are loop-closing problems when the RT steps are loose or not used.



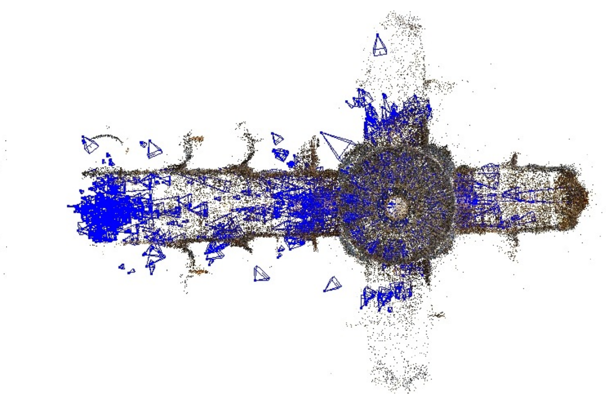
(a) $r = 5\%, r' = 25\%$;
Run partial BA every image.
1267 cameras are reconstructed.



(b) $r = 25\%, r' = 50\%$;
Run partial BA every image.
1267 cameras are reconstructed

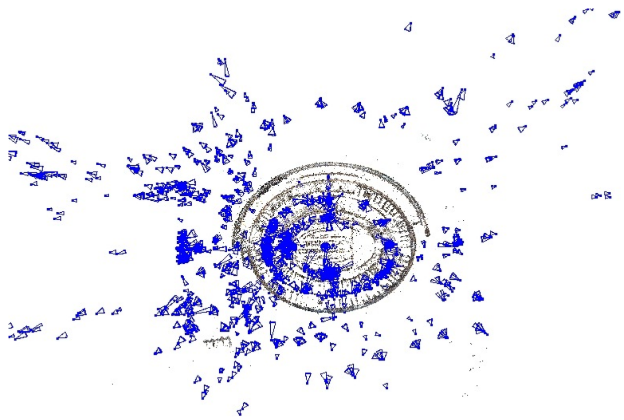


(c) $r = 5\%, r' = 25\%$;
Run partial BA every 3 images.
1262 cameras are reconstructed.

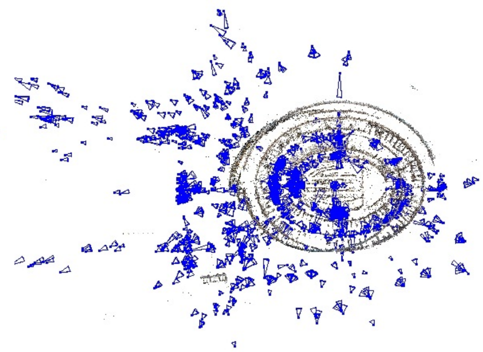


(d) $r = 5\%, r' = 25\%$;
Run partial BA every images.
Use preemptive matching $h = 100, t_h = 4$;
1211 cameras are reconstructed.

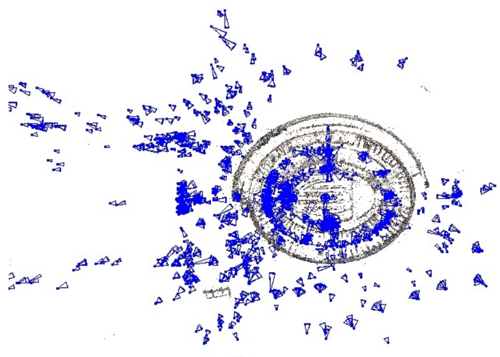
Figure 8: Top views of the St. Peter's Basilica reconstructions.



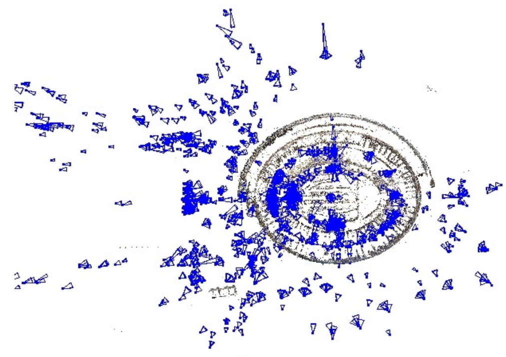
(a) $r = 5\%, r' = 25\%$;
Run partial BA every image.
1157 cameras are reconstructed.



(b) $r = 25\%, r' = 50\%$;
Run partial BA every image.
1087 cameras are reconstructed



(c) $r = 5\%, r' = 25\%$;
Run partial BA every 3 images.
1091 cameras are reconstructed.



(d) $r = 5\%, r' = 25\%$;
Run partial BA every images.
Use preemptive matching $h = 100, t_h = 2$;
1071 cameras are reconstructed.

Figure 9: Top views of the Colosseum reconstructions. The number of images are relative small with respect to the large variety of viewpoints, so some weak links are sensitive to the parameters.