Improving Structure from Motion with Reliable Resectioning Supplementary Material

Rajbir Kataria[1] Joseph DeGol[2] Derek Hoiem[1]

[1]University of Illinois Urbana-Champaign [2]Microsoft
{rk2,dhoiem}@illinois.edu jodegol@microsoft.com

1. Summary

Our supplementary document provides **(1)** Qualitative results for SfM reconstructions and MVS models (in Section 2 and 3); **(2)** Full numerical results for our method compared to other disambiguation methods (in Section 4); **(3)** Full numerical results on the ablation study presented in the main paper (for OpenSfM[1] and COLMAP[6]) (in Section 5); **(4)** The performance of our local resectioning method compared to pruning the tracks graph using ambiguity adjusted matches (in Section 6); **(5)** Timing analysis for our method which includes time to compute ambiguity adjusted matches and the reconstruction time using our method (in Section 7).

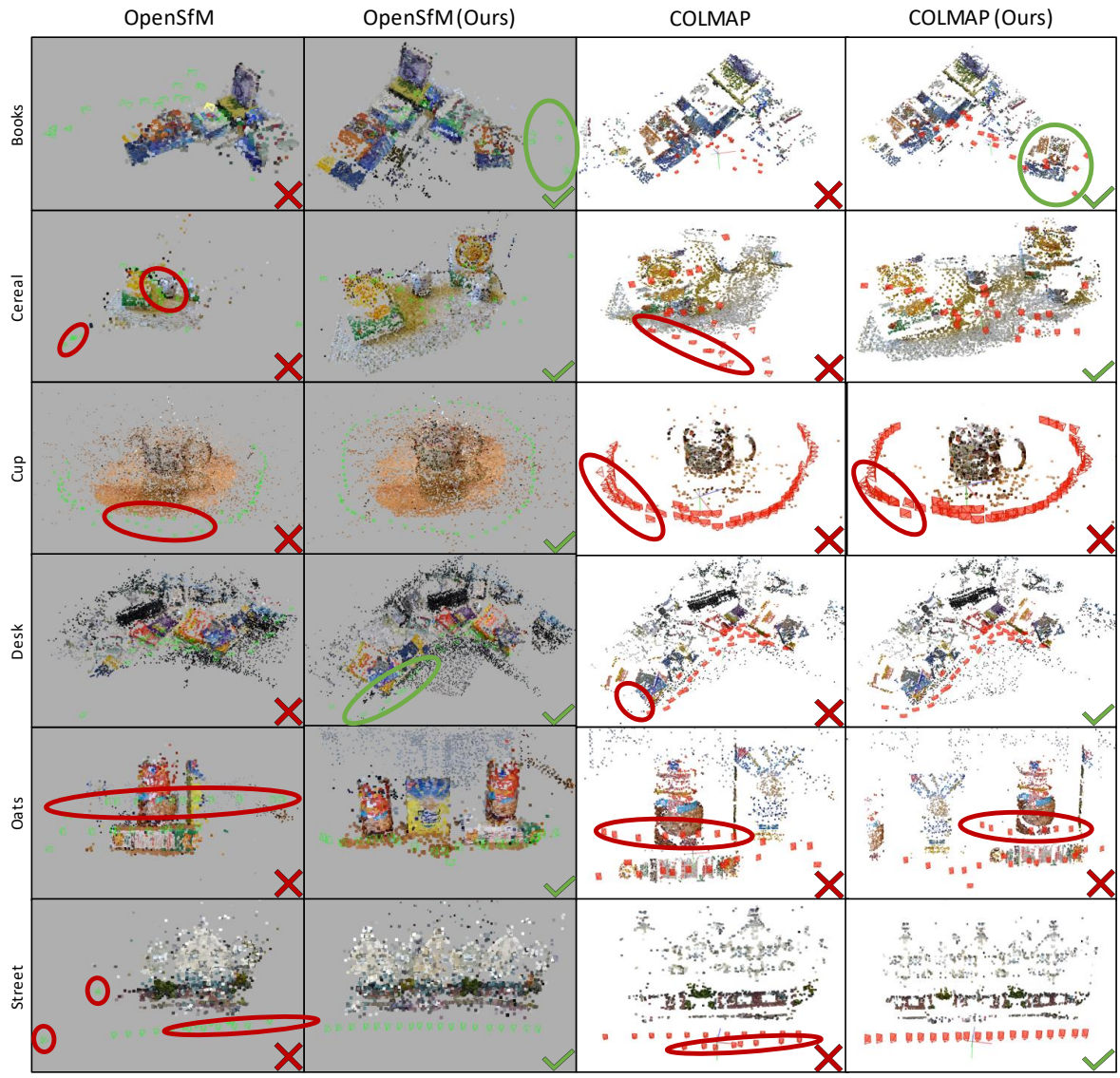

Figure 1. We evaluated the **Duplicate Structures [5]** dataset using our method on both OpenSfM[1] and COLMAP [6] pipelines. We identify obvious mis-registration of images with red markings or show correctly registered images with green markings for each scene. Our method outperforms both base systems by reconstructing 6 additional scenes using OpenSfM [1] and 4 additional scenes using COLMAP [6].

2. Qualitative Results - SfM Reconstructions

In this section, we present qualitative results for all our evaluation datasets. Each reconstruction is marked with a ✓ or ✗ corresponding to a success or failure as reported in the main paper. Figure 1 shows the results for the Duplicate Structures dataset [5], Figure 2 and Figure 3 show the results for the UIUCTag dataset [2], Figure 4 shows the results for the Tanks and Temples dataset [4], and Figure 5 shows the results on other challenging scenes from Heinly et al.[3].

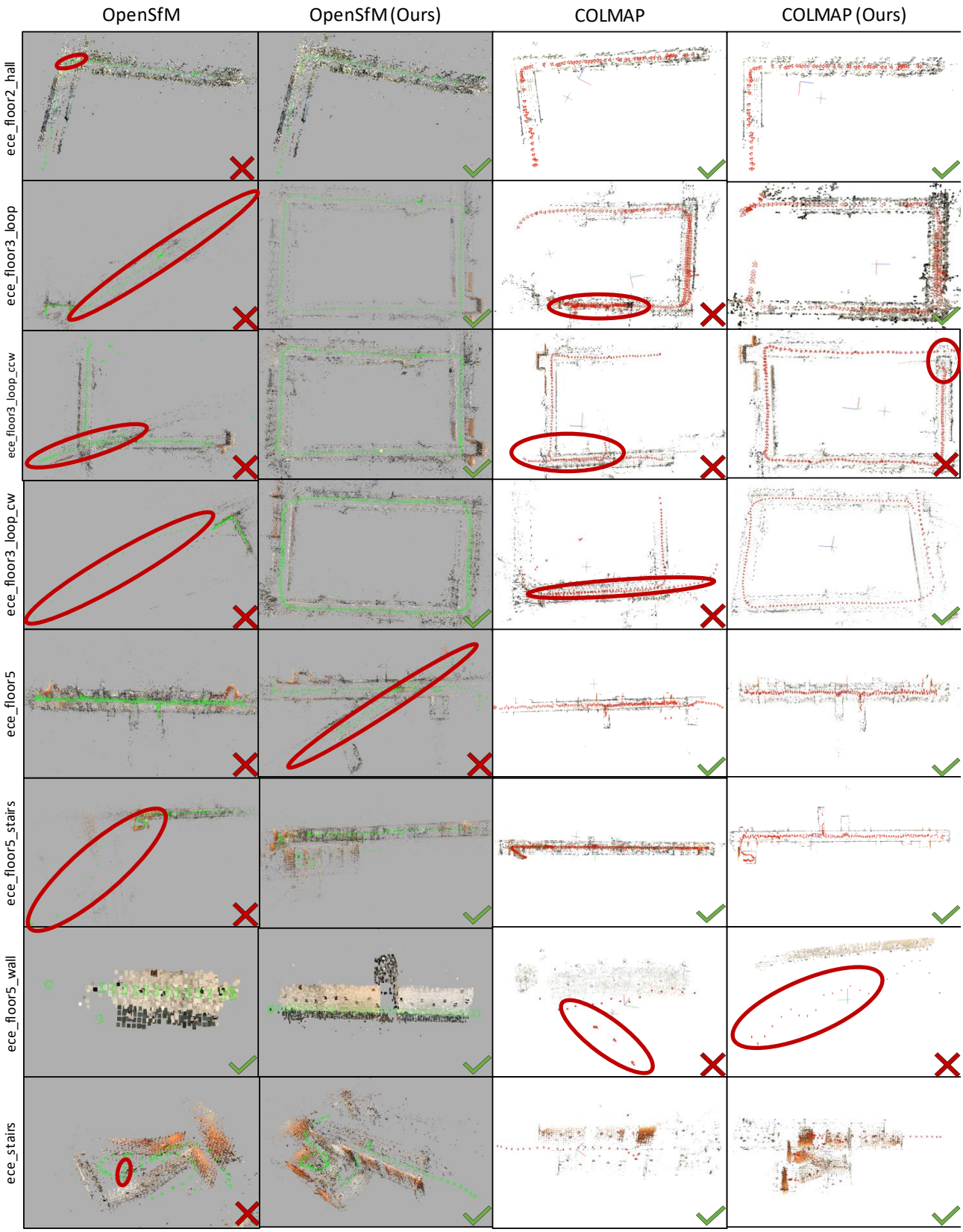

Figure 2. This figure shows the *ECE* scenes from the **UIUCTag [2]** dataset which was evaluated using our method on both OpenSfm[1] and COLMAP [6] pipelines. We identify obvious mis-registration of images with red markings or show correctly registered images with green markings for each scene. Our method outperforms both base systems for *ECE* scenes by reconstructing 6 additional scenes using OpenSfM [1] and 2 additional scenes using COLMAP [6].

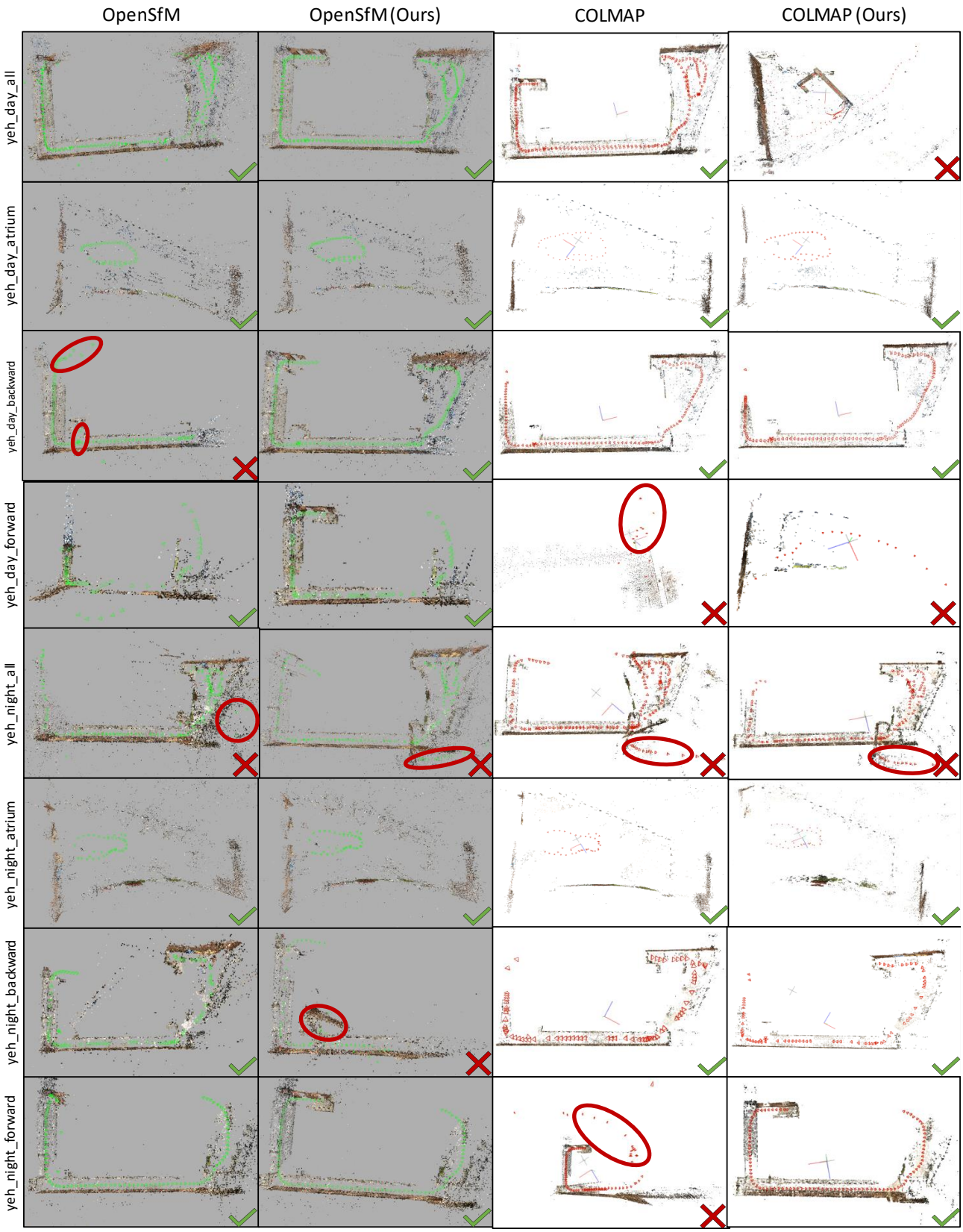

Figure 3. This figure shows the *Yeh* scenes from the **UIUCTag [2]** dataset which was evaluated using our method on both OpenSfm[1] and COLMAP [6] pipelines. We identify obvious mis-registration of images with red markings or show correctly registered images with green markings for each scene.

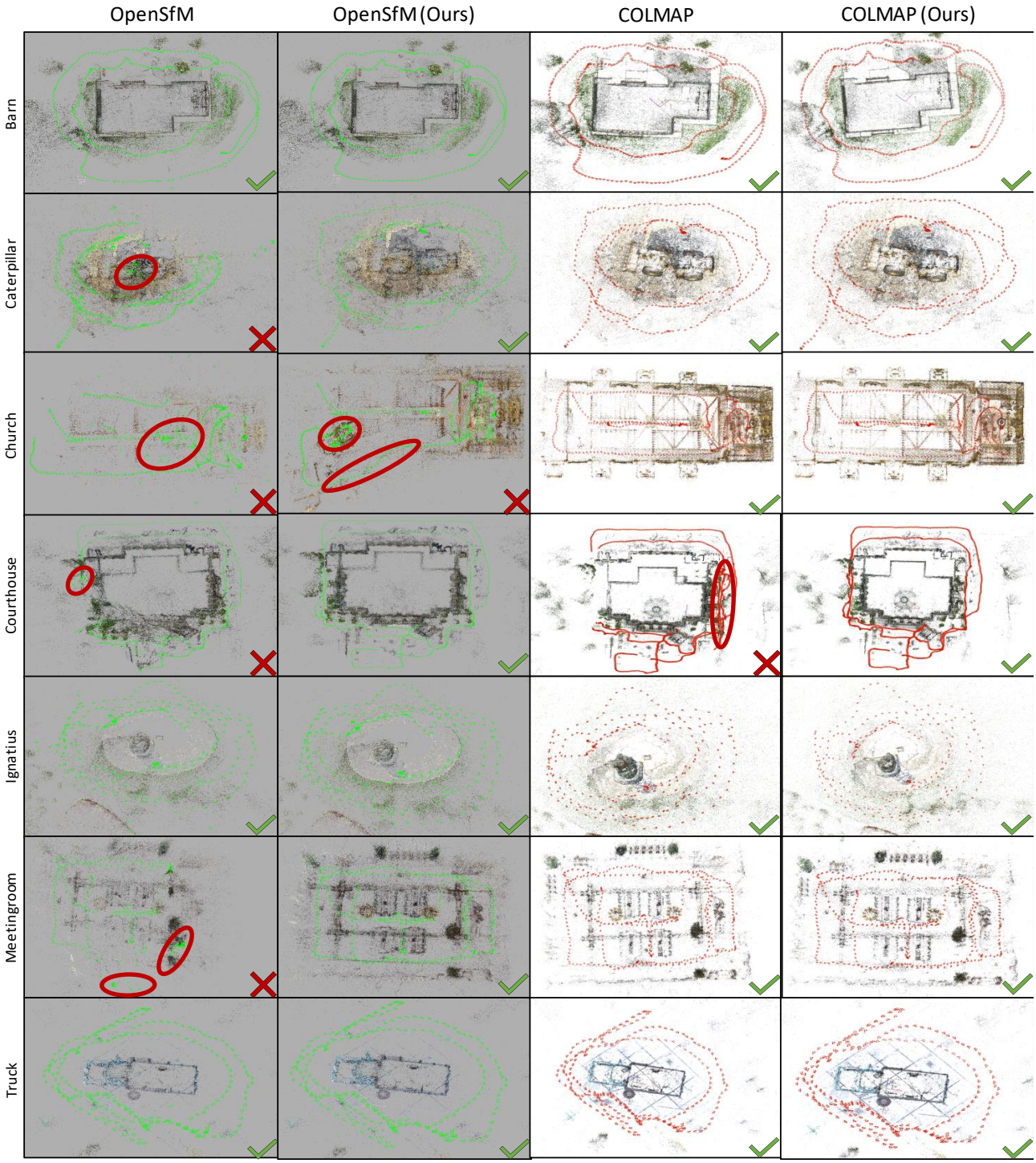

Figure 4. We evaluated the **Tanks and Temples [4]** dataset using our method on both OpenSfm[1] and COLMAP [6] pipelines. We identify obvious mis-registration of images with red markings or show correctly registered images with green markings for each scene. Our method outperforms both base systems by reconstructing 3 additional scenes using OpenSfM [1] and ‘ additional scene using COLMAP [6].

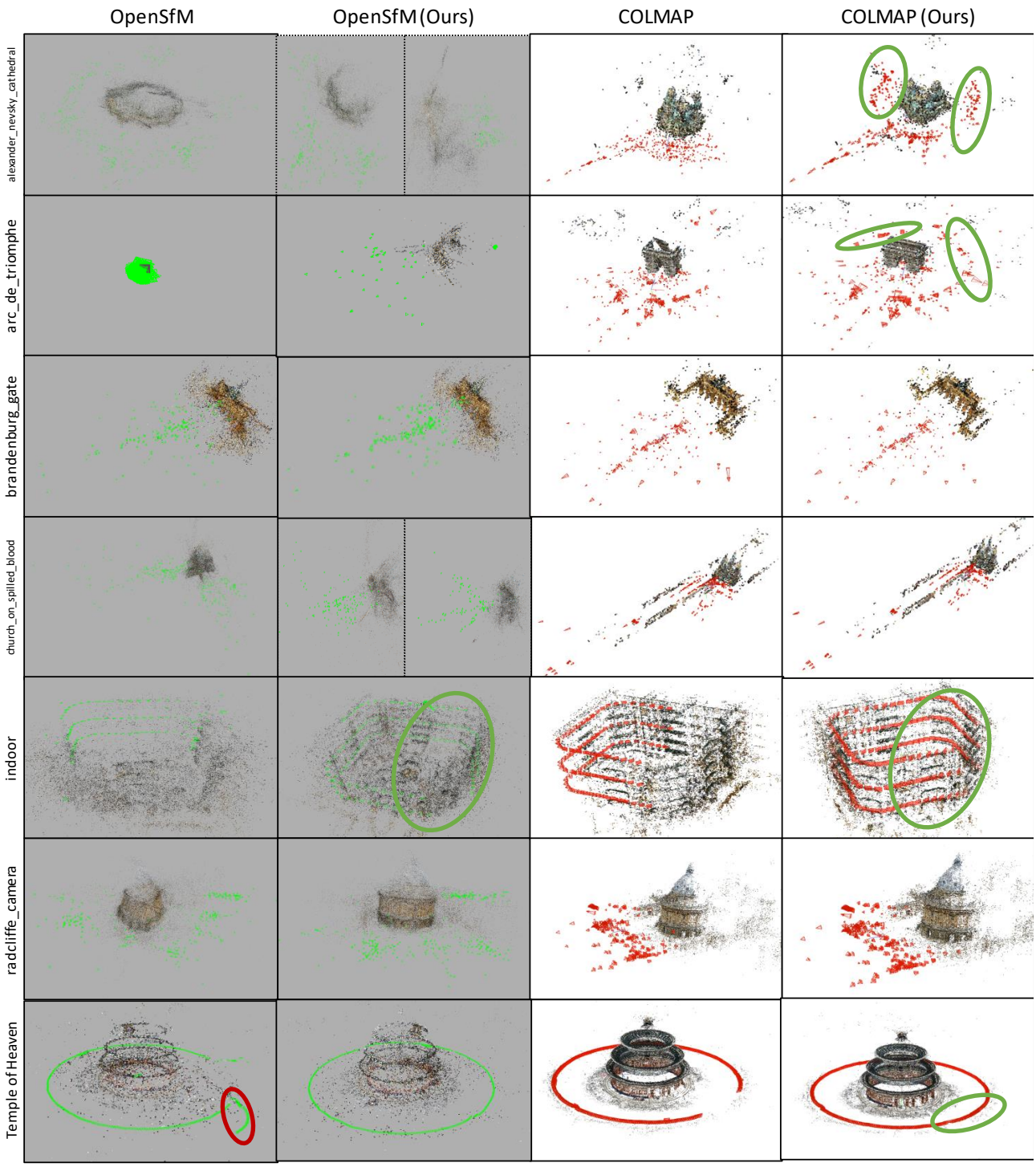

Figure 5. We evaluated several additional challenging internet datasets from Heinly et al.[3] using our method on both OpenSfm[1] and COLMAP [6] pipelines. We identify obvious mis-registration of images with red markings or show correctly registered images with green markings for each scene. We do not label these reconstructions as successful or failures as there is no ground truth or a discernible capture pattern for most of the reconstructions (with the exception of "indoor" and "Temple of Heaven"). For "arc_de_triomphe" and "church_on_spilled_blood", our method in OpenSfM yields multiple reconstructions, which are separated with dotted lines.

Figure 6. Qualitative results for dense models produced by (from left to right) OpenSfM [1], our method integrated in OpenSfM, COLMAP [6], our method integrated in COLMAP, and the ground-truth. The OpenSfM [1] baseline failed to produce a model for 4/7 scenes.

3. Qualitative Results - MVS

Figure 6 shows qualitative results of the dense models generated from the reconstructions by OpenSfM [1], COLMAP [6], and our method. For the OpenSfM [1] pipeline, the baseline system was unable to produce models for *Caterpillar*, *Church*, *Courthouse*, *Meetingroom* while our method produced a dense model for all scenes (though "Church" is erroneous). For the COLMAP [6] pipeline, the base system and our method produced comparable results for all scenes except "Courthouse". Our method produces a better model as seen by the correct reconstruction of the dome.

3.1. Qualitative Results - Barn

The improvements obtained using our method matched our qualitative inspection of SfM output for all scenes except "Barn", where the models produced by baseline and our method look nearly identical, but our method has lower precision and recall, likely due to a slight misregistration in part of the model that is difficult to perceive. Figure 7 shows several images of the models from different viewpoints along with their alignment to ground-truth.

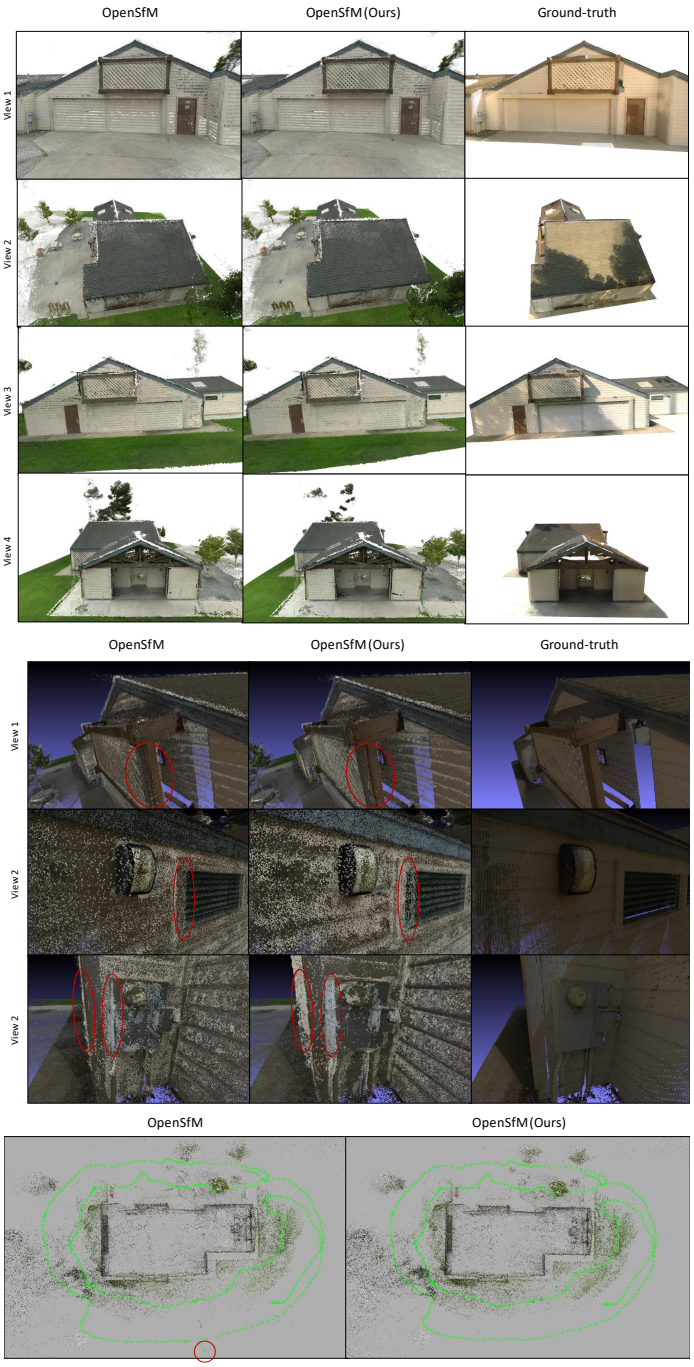

Figure 7. **Left**: Generated models of "Barn" from base system of OpenSfM [1] and our method followed up the ground-truth model. **Top-right**: Alignments of the generated models to the ground-truth along with the ground-truth from the same viewpoint (misalignment marked in red. **Bottom-right**: SfM reconstructions from base OpenSfM [1] and our method (misregistration marked by red circle).

	Total	[7]		[8]		Ours	
		%R	%O	%R	%O	%R	%O
Books	21	0	0	71	58	100	85
Cereal	25	0	0	56	50	100	82
Cup	64	✗	✗	✗	✗	100	74
Desk	31	0	0	100	91	100	91
Oats	23	0	0	100	86	100	74
Street	19	0	0	100	83	100	63
ece_floor2_hall	74	95	55	✗	✗	96	68
ece_floor3_loop	362	✗	✗	✗	✗	100	72
ece_floor3_loop_ccw	192	✗	✗	48	19	99	75
ece_floor3_loop_cw	170	✗	✗	99	33	100	77
ece_floor5	239	43	27	✗	✗	✗	✗
ece_floor5_stairs	328	✗	✗	✗	✗	94	64
ece_floor5_wall	39	0	0	10	2	97	90
ece_stairs	89	62	47	38	12	100	90
yeh_day_all	252	66	39	60	25	100	82
yeh_day_atrium	37	0	0	41	12	100	69
yeh_day_backward	120	68	40	52	18	100	88
yeh_day_forward	63	86	66	37	8	98	86
yeh_night_all	170	✗	✗	54	21	✗	✗
yeh_night_atrium	41	0	0	✗	✗	100	85
yeh_night_backward	79	0	0	46	13	✗	✗
yeh_night_forward	96	✗	✗	61	25	100	88
Barn	410	100	76	100	97	100	84
Caterpillar	383	✗	✗	100	93	100	81
Church	507	✗	✗	✗	✗	✗	✗
Courthouse	1106	✗	✗	0	0	100	76
Ignatius	263	100	62	100	87	100	82
Meetingroom	371	✗	✗	100	91	100	73
Truck	251	100	59	100	84	100	72

Table 1. In this table we compare our method to Wilson et al.[7] and Yan et al.[8]. **%R** and **%O** indicate the percentage of images and observations reconstructed and "✗" indicates an unsuccessful reconstruction. Our method is able to successfully reconstruct more scenes across varying levels of ambiguities.

4. Comparison to Disambiguation Methods

Table 1 shows numerical results for Wilson et al.[7], Yan et al.[8] and our full method. As discussed in the main paper, our method is able to successfully reconstruct more scenes from the Duplicate structures[5], UIUCTag[2], and TanksAndTemples[4] datasets. All of the methods start with the same tracks graph and reconstruct using the same reconstruction pipeline (OpenSfM[1])

	OOS		OOS w/o AAM		OOS w/o LRO		OOS w/o LPE		OCM		OCM w/o AAM	
	% R	% O	% R	% O	% R	% O	% R	% O	% R	% O	% R	% O
Books	100	85	✗	✗	100	85	✗	✗	100	80	✗	✗
Cereal	100	82	✗	✗	100	73	✗	✗	100	68	✗	✗
Cup	100	74	100	68	✗	✗	✗	✗	✗	✗	✗	✗
Desk	100	91	100	62	100	91	✗	✗	100	78	✗	✗
Oats	100	74	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Street	100	63	100	68	✗	✗	✗	✗	100	55	100	55
ece_floor2_hall	96	68	97	68	✗	✗	96	59	95	36	✗	✗
ece_floor3_loop	100	72	100	69	100	70	✗	✗	83	35	✗	✗
ece_floor3_loop_ccw	99	75	99	74	✗	✗	✗	✗	✗	✗	98	48
ece_floor3_loop_cw	100	77	100	76	100	76	✗	✗	100	51	100	52
ece_floor5	✗	✗	95	63	97	70	✗	✗	87	39	91	39
ece_floor5_stairs	94	64	98	74	✗	✗	✗	✗	80	33	82	34
ece_floor5_wall	97	90	97	90	90	78	49	31	✗	✗	✗	✗
ece_stairs	100	90	100	90	100	91	80	53	100	52	100	52
yeh_day_all	100	82	100	82	100	82	✗	✗	✗	✗	✗	✗
yeh_day_atrium	100	69	100	69	97	70	100	69	97	44	97	40
yeh_day_backward	100	88	100	88	100	88	✗	✗	90	55	92	55
yeh_day_forward	98	86	98	86	98	86	51	40	27	20	✗	✗
yeh_night_all	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
yeh_night_atrium	100	85	100	84	100	84	100	67	93	47	93	48
yeh_night_backward	✗	✗	100	87	100	86	✗	✗	90	46	✗	✗
yeh_night_forward	100	88	100	88	100	88	✗	✗	100	56	✗	✗
Barn	100	84	100	84	100	84	100	64	100	67	100	67
Caterpillar	100	81	100	81	100	81	✗	✗	100	67	100	67
Church	✗	✗	✗	✗	100	75	✗	✗	100	78	100	78
Courthouse	100	76	100	77	100	76	✗	✗	100	75	100	75
Ignatius	100	82	100	82	100	81	✗	✗	100	72	100	72
Meetingroom	100	73	100	74	100	73	100	45	100	61	✗	✗
Truck	100	72	100	72	100	72	100	22	100	67	100	67

Table 2. In this table, we show how our method (in OpenSfM) performs without ambiguity adjusted matches (OOS w/o AAM), our resectioning order (OOS w/o LRO) and our pose estimation method (OOS w/o LPE). We also compare our full method in COLMAP (OCM) to our method without ambiguity adjusted matches (OCM w/o AAM). **% R** and **% O** indicate the percentage of images and observations reconstructed and "✗" indicates an unsuccessful reconstruction. Our final method is able to successfully reconstruct the entire duplicate structures dataset and perform on par or better than the other versions of our method on the remaining datasets.

5. Ablation Study

Table 2 presents the full numerical results of our ablation study. We compare our full method (OOS), implemented in OpenSfM[1], to **(1)** Our method without ambiguity adjusted matches (OOS w/o AAM); **(2)** Our method without local resectioning order (OOS w/o LRO); **(3)** Our method without local pose estimation (OOS w/o LPE). We also compare our full method in COLMAP[6] (OCM) to our method without ambiguity adjusted matches (OCM w/o AAM).

	Total	OOS		OpenSfM w/ AAMT	
		% R	% O	% R	% O
Books	21	100	85	100	96
Cereal	25	100	82	96	94
Cup	64	100	74	100	80
Desk	31	100	91	✗	✗
Oats	23	100	74	100	83
Street	19	100	63	100	85
ece_floor2_hall	74	96	68	41	32
ece_floor3_loop	362	100	72	48	42
ece_floor3_loop_ccw	192	99	75	80	72
ece_floor3_loop_cw	170	100	77	100	89
ece_floor5	239	✗	✗	28	25
ece_floor5_stairs	328	94	64	34	32
ece_floor5_wall	39	97	90	100	97
ece_stairs	89	100	90	100	89
yeh_day_all	252	100	82	57	49
yeh_day_atrium	37	100	69	95	63
yeh_day_backward	120	100	88	100	91
yeh_day_forward	63	98	86	51	42
yeh_night_all	170	✗	✗	51	44
yeh_night_atrium	41	100	85	98	82
yeh_night_backward	79	✗	✗	47	39
yeh_night_forward	96	100	88	100	94
Barn	410	100	84	100	97
Caterpillar	383	100	81	100	86
Church	507	✗	✗	✗	✗
Courthouse	1106	100	76	56	65
Ignatius	263	100	82	97	78
Meetingroom	371	100	73	100	94
Truck	251	100	72	100	78

Table 3. In this table we show the results of our investigations into pruning the tracks graph using our similarity measure (*OpenSfM w/ AAMT*) instead of using our local resectioning strategy (*OOS*). **% R** and **% O** indicate the percentage of images and observations reconstructed and "✗" indicates an unsuccessful reconstruction. Our final method has more complete reconstructions overall compared to pruning the tracks graph.

6. Similarity Measures

Our final method uses ambiguity adjusted matches (AAM) as a measure of similarity in our local resectioning strategy, however we explored one other variant during our experimentation. This involved pruning the tracks graph based on AAM and using the default resectioning strategy instead of our method. Table 3 shows that pruning the tracks graph based on ambiguity adjusted matches gives 17/29 successful reconstructions and 11/29 partial reconstructions. These results supports our claim that thresholding the tracks graph may not be optimal for every scene and may require careful tuning of the thresholding value to give complete reconstructions.

	Total	OpenSfM			Ours		
		AAM Calc Time	Recon Time	Total Time	AAM Calc Time	Recon Time	Total Time
yeh_day_all	252	-	450.54	**450.54**	11.93	587.97	599.9
yeh_day_atrium	37	-	46.59	46.59	0.74	29.27	**30.01**
yeh_night_atrium	41	-	72.32	72.32	1.17	57.62	**58.78**
Barn	410	-	1429.58	1429.58	43.96	959.39	**1003.34**
Ignatius	263	-	631.76	631.76	31.04	412.16	**443.2**
Truck	251	-	505.8	505.8	25.13	434.3	**459.43**

Table 4. In this table we compare the reconstruction run time of the original OpenSfM[1] to the total run time of our method (sum of calculation time for ambiguity adjusted matches and the reconstruction time using our resectioning strategy). We only consider scenes where the original OpenSfM[1] and our method produced successful reconstructions and all the times are reported in seconds. Our method results in a speedup for 5/6 reconstructions.

7. Timing Analysis

To compute the overhead added by our method, we add the time to calculate the ambiguity adjusted matches to the reconstruction time (using our resectioning strategy) and compare it to the reconstruction time of original OpenSfM[1]. The calculation time for ambiguity adjusted matches is less than 60 seconds for these scenes. Our method results in a lower total time for 5/6 successful reconstructions, yielding a mean speedup of 25% (median of 33%). The speedup is primarily due to less iteration steps required by Bundle Adjustment (BA) to reach the solution resulting from a better initial pose estimation.

References

[1] OpenSfM. https://github.com/mapillary/OpenSfM. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12

[2] J. DeGol, T. Bretl, and D. Hoiem. Improved structure from motion using fiducial marker matching. In *ECCV*, 2018. 2, 3, 4, 9

[3] J. Heinly, E. Dunn, and J.-M. Frahm. Correcting for Duplicate Scene Structure in Sparse 3D Reconstruction. In *European Conference on Computer Vision (ECCV)*, 2014. 2, 6

[4] A. Knapitsch, J. Park, Q.-Y. Zhou, and V. Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), 2017. 2, 5, 9

[5] R. Roberts, S. N. Sinha, R. Szeliski, and D. Steedly. Structure from motion for scenes with large duplicate structures. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2011)*, June 2011. 2, 9

[6] J. L. Schönberger and J.-M. Frahm. Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1, 2, 3, 4, 5, 6, 7, 10

[7] K. Wilson and N. Snavely. Network principles for sfm: Disambiguating repeated structures with local context. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2013. 9

[8] Q. Yan, L. Yang, L. Zhang, and C. Xiao. Distinguishing the indistinguishable: Exploring structural ambiguities via geodesic context. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 9