Unifying Viewgraph Sparsification and Disambiguation of Repeated Structures in Structure-from-Motion Supplementary Material

Lalit Manam[1] and Venu Madhav Govindu[1*]

[1*]Indian Institute of Science, Bengaluru - 560012, INDIA.

*Corresponding author(s). E-mail(s): venug@iisc.ac.in;
Contributing authors: lalitmanam@iisc.ac.in;

In this supplementary material, we provide additional details of the experiment results shown in the main paper. Below, we restate the notation used in the main paper and introduce new notation which will be used further.

- $\mathcal{G}$: Original viewgraph obtained using COLMAP Schonberger & Frahm (2016).
- $\mathcal{G}_{\#in}$: Graph obtained after thresholding number of inliers in $\mathcal{G}$ (threshold set to 150 as in Cai et al. (2023)).
- $\mathcal{G}_{Dopp}$: Graph obtained after applying Doppelgangers Cai et al. (2023) on $\mathcal{G}$.
- $\mathcal{G}_{FG}(m)$: Graph obtained after applying Algo. 1 with a minimum edge score m.
- $\#N_{CR}$: Number of cameras reconstructed.
- t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016).
- RRE: Recall (in %) of camera rotation errors at 5°.
- RTE: Recall (in %) of camera translation errors at 10 meters for datasets with ground truth available, otherwise 10 units, where units is defined by reference.
- MRE: Mean camera rotation error (degrees).
- MTE: Mean camera translation error (meters or units specified by the reference).

A Additional Results on Generic Datasets

In Sec. 4.1 of the main paper, we presented reconstruction results on generic datasets. Here, we provide additional details.

Effect of Minimum Edge Score: In Table S1, we show the impact of m in removing ghost artifacts using COLMAP Schonberger & Frahm (2016). It can be seen that the ghost artifacts are removed for all datasets after applying our method for m between 0.7 and 0.9.

Camera Motion Errors for both Rotations and Translations: In Table S2, we compare camera motions for generic datasets in terms of recall and mean errors. We use the same reference motions and the same values of minimum edge score m suggested in the main paper. Additionally, as stated in the main paper, for computing recall, we only use common cameras reconstructed across different methods. It can be seen that our method has the highest recall for most datasets, both for rotations and translation, suggesting that the reconstruction accuracy is maintained after sparsification of viewgraphs. Moreover, we also report the mean errors of camera motions using our method to check the errors in all the cameras. It can be seen that applying our method results in less than 1° rotation error and less than 2 meters of translation error for most datasets.

Comparison with using only Strong Triples Manam & Govindu (2024)**:** In Table 4 of the main paper, we provided a comparison of the results with our method, which uses both strong and weak triples to that of the method in Manam & Govindu (2024), which uses only strong triples, on generic datasets. Here, we provide more details of the comparison in Table S3. We observe that the mean reprojection errors are similar for both methods, while our method reconstructs more cameras and 3D points (Table 4 of the main paper), resulting in increased reconstruction time.

Visual Results: In Figs. S1, S2, S3, and S4, we provide visual results on generic datasets with varying minimum edge score m. From Figs. S1, S2, and S3, it can be seen that reconstructions obtained with sparsified viewgraphs $\mathcal{G}_{FG}$ are visually similar to those of the original viewgraphs $\mathcal{G}$. For Quad, the cameras are not well connected, due to which our method reconstructs a part of the original reconstruction. Our method avoids ghost artifacts as seen in the reconstructions in Fig. S4.

Choosing a Minimum Edge Score: Based on these results, we recommend using the value of minimum edge score m near 0.7 for viewgraph sparsification on generic datasets, keeping a trade-off between reconstruction quality and time.

B Additional Results on Ambiguous Datasets

In Sec. 4.2 of the main paper, we presented reconstruction results on ambiguous datasets with our method. Here, we study the impact of choosing different values of minimum edge score m. For large-scale datasets, we show for $m = \{0.6, 0.7, 0.8, 0.9\}$ and for medium and small-scale datasets, we show for $m = \{0.3, 0.4, 0.5, 0.6\}$. We also provide results from Doppelgangers Cai et al. (2023) with the same setting discussed in the main paper. (The probability threshold is chosen as $p = 0.8$ for all datasets except for Louvre Wilson & Snavely (2013), where we use $p = 0.9$. Doppelgangers Cai et al. (2023) did not disambiguate some datasets for any probability threshold we checked and thus are marked as not disambiguated.)

Effect of Minimum Edge Score: In Table S4, we show the effectiveness of our method on ambiguous datasets with different values of m. It can be seen that large-scale datasets Wilson & Snavely (2013, 2014) are disambiguated using $m = 0.7$ (which is recommended for generic datasets) except for Sacre Coeur Wilson & Snavely (2013), Seville Wilson & Snavely (2013) and Ellis Island Wilson & Snavely (2014), where higher values of m are required to remove false edges. In general, higher values of m are not recommended since they can lead to oversplit reconstructions, as in the case of Seville Wilson & Snavely (2013) and Piazza del Popolo Wilson & Snavely (2014) for $m = 0.9$. For medium Heinly et al. (2014) and small-scale Yan et al. (2017) datasets, all datasets are disambiguated with $m = 0.5$. But it leads to oversplit reconstructions for small-scale datasets Yan et al. (2017). This is because our method sparsifies the graphs apart from removing false edges, and low redundancy of edges in small-scale datasets leads to disconnected graphs.

Camera Motion Errors for both Rotations and Translations: In Table S5, we provide camera motion errors on ambiguous datasets in terms of recall and mean errors. We follow the same procedure for calculating recall as done in the main paper, and use the same values of minimum edge score m suggested in the main paper. It can be seen that our method performs overall best when compared to disambiguated reconstructions. Moreover, mean errors of rotations and translations are not high, signifying that the reconstructions obtained have similar camera motions compared to Doppelgangers Cai et al. (2023).

Time Taken: In Table S6, we provide reconstruction time taken by COLMAP Schonberger & Frahm (2016) with varying values of minimum edge score m. It can be seen that the reconstruction time decreases significantly with an increase in m when compared to Doppelgangers Cai et al. (2023) for large and medium-scale datasets. This is because our method removes false edges and also sparsifies the graphs. We also provide a comparison of time taken by our method with other disambiguation methods in Table S7. It can be seen that our method takes the least compute time for all datasets, even compared to Manam & Govindu (2024), which uses strong triples. This is because, unlike Manam & Govindu (2024), our method does not require any preprocessing to extract strong triples and is parallelized on edges.

Comparison with using only Strong Triples Manam & Govindu (2024)**:** In Table 8 of the main paper, we provided a comparison of the results with our method, which uses both strong and weak triples to that of the method in Manam & Govindu (2024), which uses only strong triples, on ambiguous datasets. Here, we provide more details of the comparison in Table S8. We observe that the mean reprojection errors are similar for both methods, while our method reconstructs more cameras and 3D points (Table 8 of the main paper), resulting in increased reconstruction time.

Visual Results: In Figs. S5, S6, S7, S8, S9, and S10, we provide visual results on the ambiguous datasets using our method. It can be seen that our method is able to disambiguate all datasets. For large and medium-scale datasets, viewgraphs from our method $\mathcal{G}_{FG}$ lead to visually similar reconstructions compared to Doppelgangers Cai et al. (2023) $\mathcal{G}_{Dopp}$ with $m = 0.7$ for large-scale and $m = 0.5$ for medium-scale datasets. For small-scale datasets, our method disambiguates all datasets but leads to oversplit reconstructions, as discussed in the main paper.

Choosing a Minimum Edge Score: Based on these results on ambiguous datasets, we recommend using minimum edge score $m = 0.7$ for large-scale datasets (same as for generic datasets) and $m = 0.5$ for medium and small-scale datasets. The minimum edge score should only be increased for highly ambiguous datasets to avoid superimposed reconstructions in such datasets.

Analysis with GLOMAP Pan et al. (2024)**:** In Table S9, we check the performance of different methods using GLOMAP. We observe that no method performs well across datasets. This suggests that the accuracy of disambiguation methods in terms of detecting false edges should be improved, and global methods like GLOMAP require more robustness to false edges.

Table S1: Analysis of reconstruction artifacts and reprojection errors on generic datasets with our method. ✓/✓*/✗: No ghost artifact/no ghost artifact but over-split reconstruction/contains ghost artifacts. Applying our method avoids ghost artifacts in the reconstructions

Dataset	No Ghost Artifact				
	$\mathcal{G}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$
$m \rightarrow$		0.6	0.7	0.8	0.9
ALM	✓	✓	✓	✓	✓
GMM	✗	✓	✓	✓	✓
MDR	✓	✓	✓	✓	✓
MND	✓	✓	✓	✓	✓
ND	✗	✓	✓	✓	✓
NYC	✓	✓	✓	✓	✓
PIC	✗	✗	✓	✓	✓
ROF	✓	✓	✓	✓	✓
TOL	✓	✓	✓	✓	✓
TFG	✓	✓	✓	✓	✓
USQ	✓	✓	✓	✓	✓
VNC	✗	✓	✓	✓	✓
DVK	✓	✓	✓	✓	✓
ROM	✓	✓	✓	✓	✓
QAD	✓	✓	✓	✓	✓*

Table S2: Camera motion error comparison in terms of recall (%) and mean errors on generic datasets. -: No reconstruction obtained from COLMAP. -ˆ: No camera motion errors computed as COLMAP reference or the compared method contains ghost artifacts. **Bold** indicates best recall value. Applying our method maintains reconstruction accuracy

Dataset	Rotation Recall (RRE) (at 5°)		Translation Recall (RTE) (at 10 units)		Mean Errors ($\mathcal{G}_{FG}$)	
	$\mathcal{G}_{\#in}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{\#in}$	$\mathcal{G}_{FG}$	MRE	MTE
ALM	**100.00**	**100.00**	**99.70**	**99.70**	0.79	1.29
GMM	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ
MDR	**96.77**	**96.77**	96.37	**97.18**	2.90	2.02
MND	**100.00**	**100.00**	99.80	**100.00**	0.31	0.11
NYC	**99.75**	**99.75**	99.50	**100.00**	3.81	2.51
ND	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ
PIC	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ
ROF	99.61	**99.85**	99.61	**99.92**	0.14	0.42
TOL	**100.00**	**100.00**	98.94	**99.47**	0.17	0.90
TFG	99.18	**99.47**	98.21	**99.12**	0.77	1.14
USQ	**99.55**	**99.55**	**98.65**	**98.65**	3.45	4.30
VNC	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ
DVK	99.82	**99.89**	**100.00**	**100.00**	0.09	0.02
ROM	**99.94**	**99.94**	**100.00**	**100.00**	0.02	0.03
QAD	94.74	**96.48**	**100.00**	**100.00**	1.49	0.04

Dataset	Rotation Recall (RRE) (at 5°)			Translation Recall (RTE) (at 10 units)			Mean Errors ($\mathcal{G}_{FG}$)	
	$\mathcal{G}$	$\mathcal{G}_{\#in}$	$\mathcal{G}_{FG}$	$\mathcal{G}$	$\mathcal{G}_{\#in}$	$\mathcal{G}_{FG}$	MRE	MTE
BDG	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.12	0.54
BRM	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.20	0.65
BKP	**100.00**	**100.00**	**100.00**	**99.77**	**99.77**	**99.77**	0.08	0.39
COE	**100.00**	**100.00**	**100.00**	**98.38**	**98.38**	98.18	0.12	1.42
GPB	**100.00**	**100.00**	**100.00**	**99.12**	**99.12**	97.70	0.32	1.62
LMS	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.09	0.06
NDF	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.06	0.09
PNE	**100.00**	**100.00**	**100.00**	99.36	99.36	**99.68**	0.11	0.22
PSM	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ	-ˆ
SCR	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.03	0.36
SAF	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.27	0.12
SPC	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.43	0.46
SPS	**100.00**	**100.00**	**100.00**	**99.66**	99.49	98.63	0.07	1.04
TAJ	**100.00**	**100.00**	**100.00**	99.70	**100.00**	99.70	0.08	1.49
TNJ	**100.00**	**100.00**	**100.00**	**98.21**	**98.21**	96.43	0.18	1.26
TRF	99.86	**100.00**	**100.00**	99.71	99.71	**99.86**	0.06	0.14
CTY	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.78	0.08
DYA	**100.00**	77.78	97.78	**100.00**	**100.00**	**100.00**	3.00	0.23
ELT	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.46	0.05
FCD	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.54	0.09
KKR	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.44	0.02
MDW	**100.00**	-	**100.00**	**100.00**	-	**100.00**	1.16	0.17
OFC	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.31	0.01
PYG	94.74	94.74	**100.00**	**100.00**	**100.00**	**100.00**	0.37	0.11
RLF	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.29	0.02
RLF2	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.37	0.02
TRC	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	0.16	0.01
TRN	**100.00**	64.00	88.00	**100.00**	**100.00**	**100.00**	2.13	0.03

Table S3: Comparison of our method [ST+WT] with Manam & Govindu (2024) [ST] on generic datasets. ST and WT indicate the usage of strong and weak triples, respectively. Using both strong and weak triples (our method) reconstructs more cameras and 3D points (Table 4 in the main paper) compared to using only strong triples Manam & Govindu (2024) with similar reprojection errors, resulting in increased reconstruction time

Dataset	Mean Reprojection Errors (px)				Reconstruction Time (mins) (t_R)			
	$m = 0.6$		$m = 0.7$		$m = 0.6$		$m = 0.7$	
Triples →	ST	ST+WT	ST	ST+WT	ST	ST+WT	ST	ST+WT
ALM	0.62	0.63	0.59	0.62	28	32	25	28
GMM	0.61	0.76	0.74	0.75	144	254	103	211
MDR	0.57	0.60	0.58	0.59	22	56	21	42
MND	0.80	0.85	0.79	0.83	50	109	52	89
ND	0.70	0.72	0.68	0.71	1323	1407	758	1229
NYC	0.71	0.72	0.70	0.71	40	79	40	55
PIC	0.74	0.75	0.72	0.74	1060	1382	1000	1026
ROF	0.73	0.74	0.72	0.74	477	677	453	508
TOL	0.62	0.62	0.61	0.62	44	75	42	66
TFG	0.73	0.74	0.72	0.73	3852	5075	3698	4855
USQ	0.69	0.73	0.69	0.73	80	182	86	147
VNC	0.74	0.74	0.73	0.73	398	647	376	558
DVK	0.73	0.75	0.72	0.74	521	470	428	406
ROM	0.87	0.89	0.85	0.88	279	220	265	217
QAD	0.67	0.68	0.66	0.68	318	271	249	290

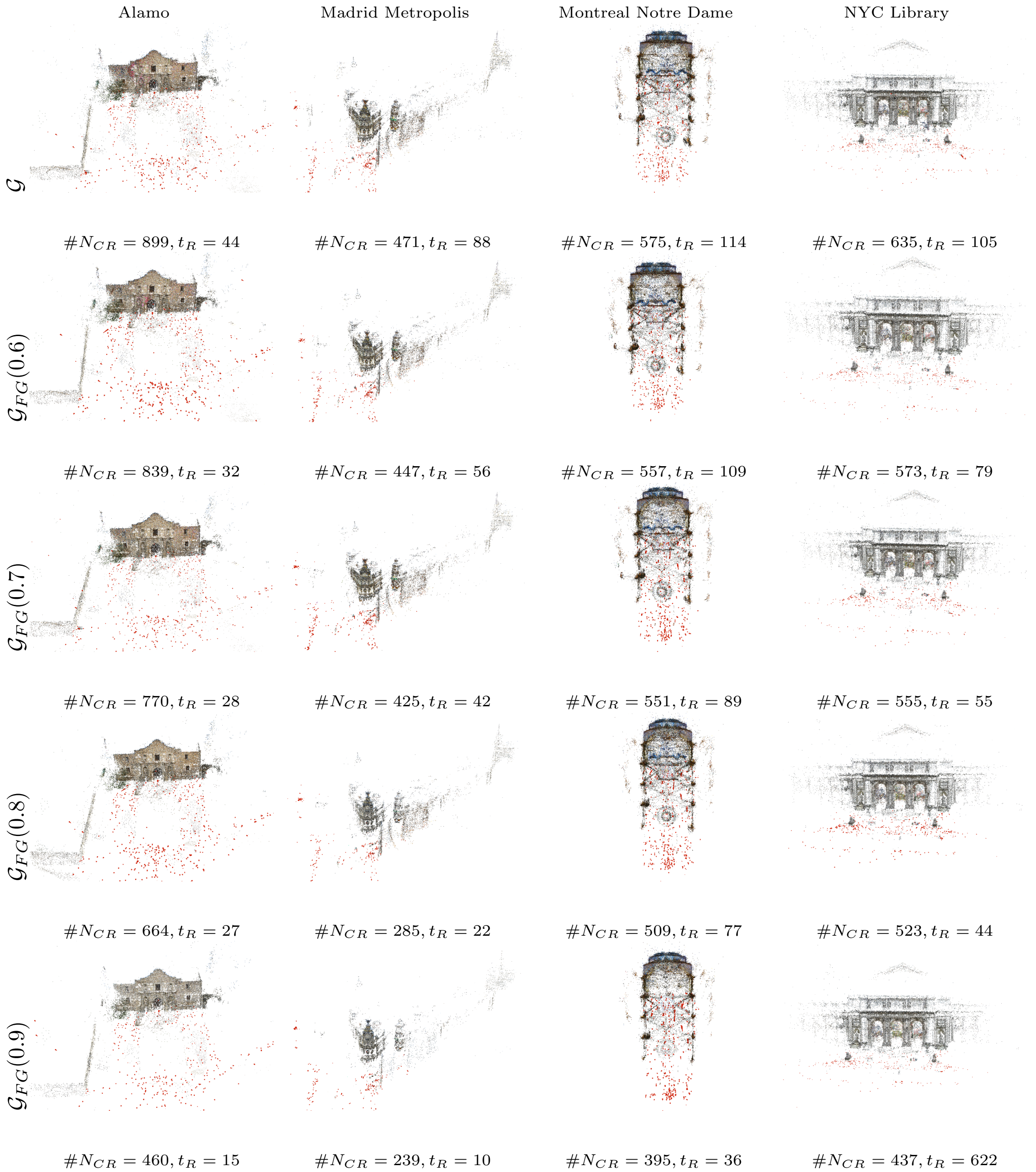

Fig. S1: Reconstructions obtained on generic datasets Wilson & Snavely (2014) using our method. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m): 0.7

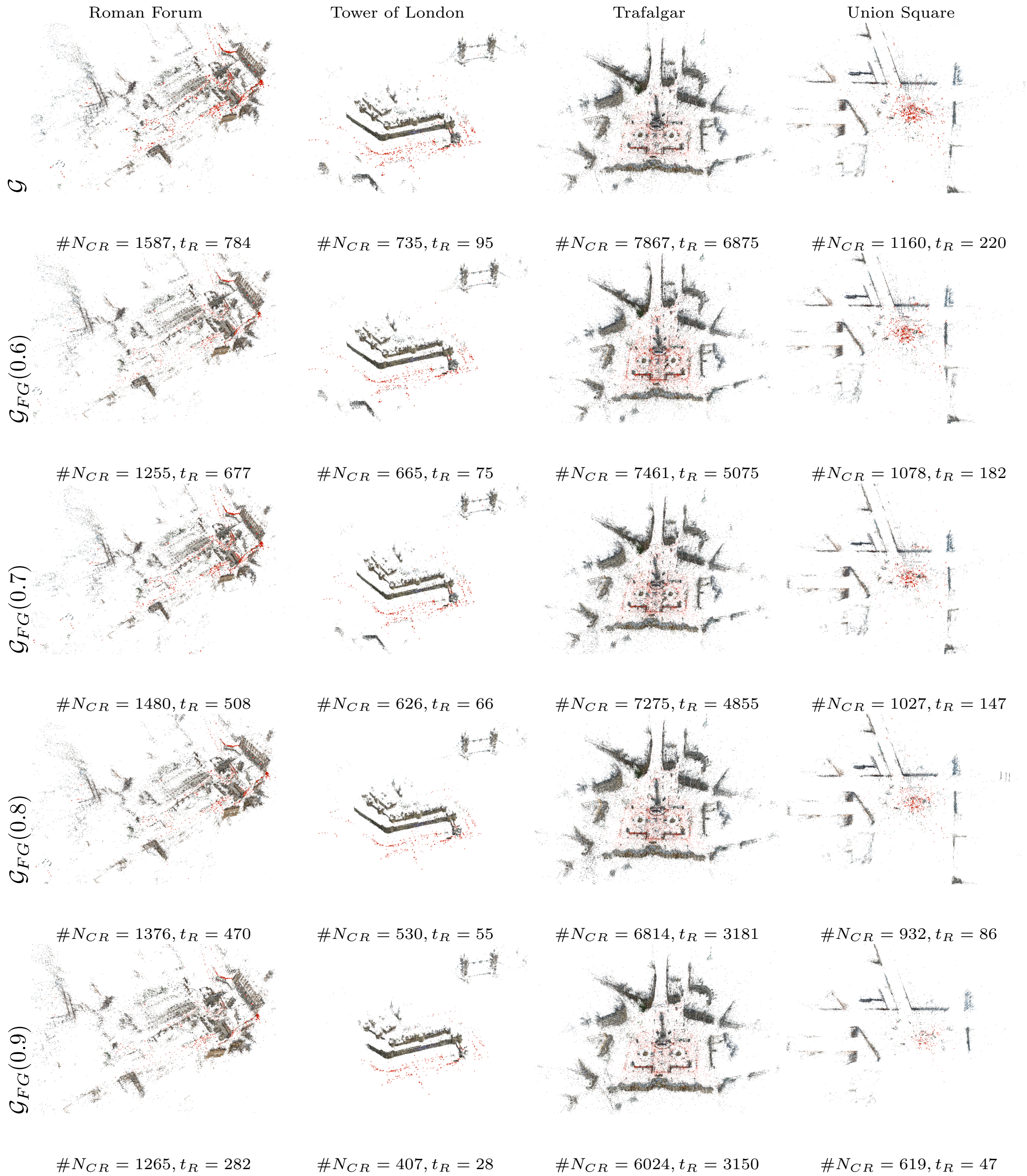

Fig. S2: Reconstructions obtained on generic datasets Wilson & Snavely (2014) using our method. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m): 0.7

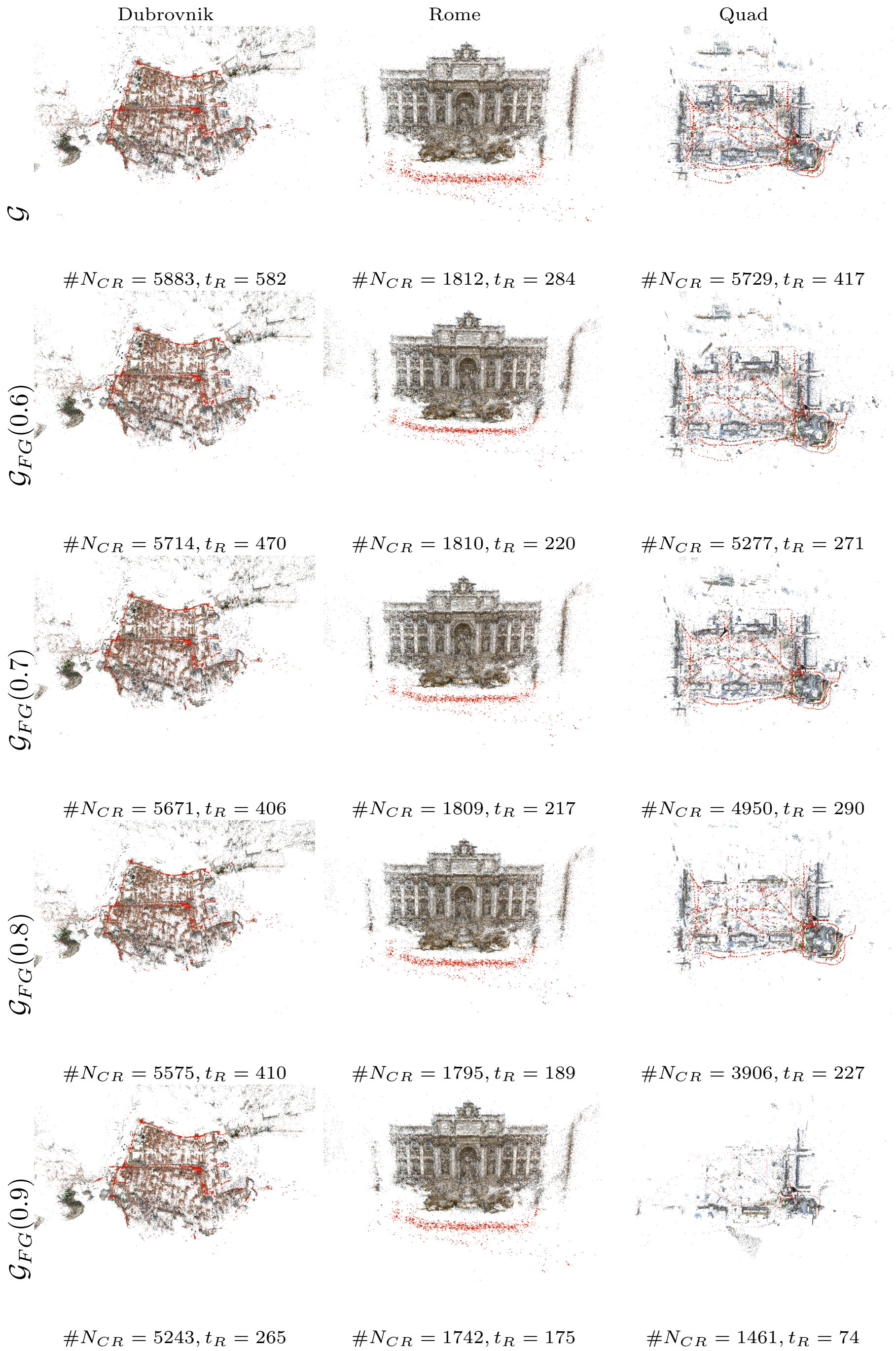

Fig. S3: Reconstructions obtained on generic datasets Crandall et al. (2011); Li et al. (2010) with our method. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m): 0.7

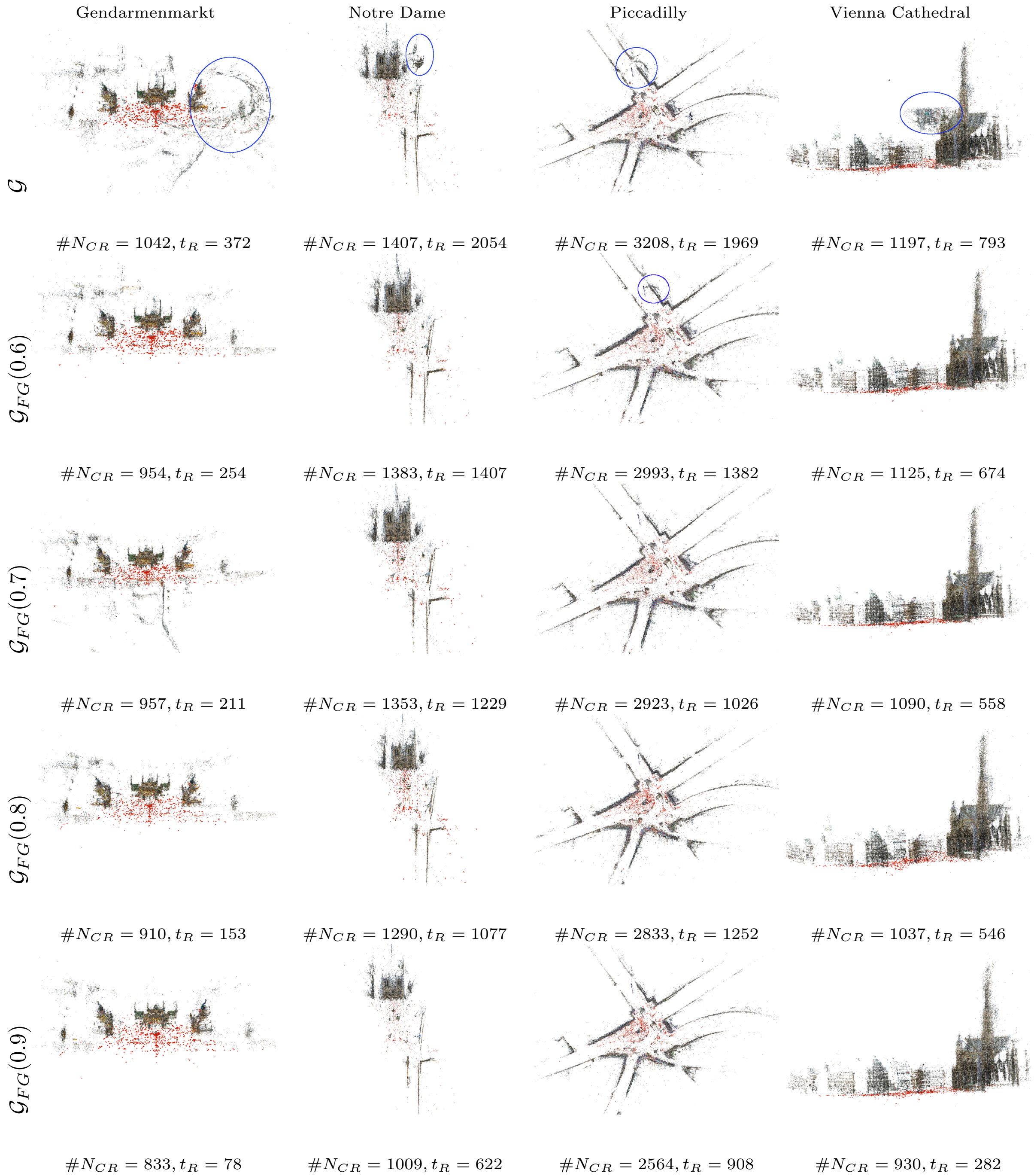

Fig. S4: Reconstructions obtained on generic datasets Wilson & Snavely (2014) using our method. Ghost artifacts are marked in blue. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m): 0.7

Table S4: Analysis of superimposed reconstructions of ambiguous datasets with our method using COLMAP Schonberger & Frahm (2016). ✓/✓*/✗: Disambiguated/Disambiguated but oversplit/Non-disambiguated reconstructions. Our method removes false edges, resulting in correct reconstructions

Dataset	Disambiguated					
	$\mathcal{G}$	$\mathcal{G}_{Dopp}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$
$m \rightarrow$			0.6	0.7	0.8	0.9
LVE	✓	✓	✓	✓	✓	✓
NDE	✗	✓	✗	✓	✓	✓
SCO	✗	✓	✗	✗	✗	✓
SEV	✗	✓	✗	✗	✓	✓*
ELS	✗	✓	✗	✗	✗	✓
PDP	✗	✓	✓	✓	✓	✓*
YKM	✗	✓	✓	✓	✓	✓
	$\mathcal{G}$	$\mathcal{G}_{Dopp}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$
$m \rightarrow$			0.3	0.4	0.5	0.6
ANC	✗	✓	✗	✗	✓	✓
ADT	✗	✓	✓	✓	✓	✓
BLD	✗	✓	✓	✓	✓	✓
BBN	✗	✓	✓	✓	✓	✓
BRG	✗	✓*	✓*	✓*	✓*	✓*
CSB	✗	✗	✓	✓	✓	✓
IDR	✗	✓	✓	✓	✓	✓*
RAC	✗	✓*	✓*	✓*	✓*	✓*
BOK	✗	✗	✓*	✓*	✓*	✓*
CER	✗	✗	✓*	✓*	✓*	✓*
CUP	✗	✓	✗	✗	✓*	✓*
DSK	✓	✓	✓*	✓*	✓*	✓*
OTS	✗	✗	✓*	✓*	✓*	✓*
STR	✗	✓*	✓	✓	✓*	✓*
TOH	✓	✓	✓	✓	✓	✓

Table S5: Camera motion error comparison in terms of recall (%) and mean errors on ambiguous datasets. -^: No camera motion errors computed as Doppelgangers Cai et al. (2023) reference or the compared reconstruction is not disambiguated. -#: No camera motion errors computed as $\mathcal{G}_{Dopp}$ and $\mathcal{G}_{FG}$ do not have common cameras. **Bold** indicates best recall value. Applying our method has reconstruction accuracy similar to Doppelgangers

Dataset	Rotation Recall (RRE) (at 5°)			Translation Recall (RTE) (at 10 units)			Mean Errors ($\mathcal{G}_{FG}$)	
	$\mathcal{G}$	$\mathcal{G}_{\#in}$	$\mathcal{G}_{FG}$	$\mathcal{G}$	$\mathcal{G}_{\#in}$	$\mathcal{G}_{FG}$	MRE	MTE
LVE	87.60	**100.00**	99.59	**100.00**	**100.00**	**100.00**	0.75	0.18
NDE	-^	-^	**99.89**	-^	-^	**100.00**	0.10	0.01
SCO	-^	-^	**98.60**	-^	-^	**99.95**	1.11	0.56
SEV	-^	-^	**99.55**	-^	-^	**100.00**	0.15	0.07
ELS	-^	**99.29**	98.93	-^	**99.29**	98.93	2.42	1.05
PDP	-^	**100.00**	**100.00**	-^	**100.00**	**100.00**	1.74	1.26
YKM	-^	99.62	**100.00**	-^	99.62	**99.81**	0.09	0.16
ANC	-^	-^	**100.00**	-^	-^	**100.00**	0.05	0.01
ADT	-^	**99.71**	99.12	-^	99.71	**100.00**	9.16	0.51
BLD	-^	**100.00**	**100.00**	-^	**100.00**	**100.00**	0.09	0.02
BBN	-^	-^	**99.18**	-^	-^	**100.00**	2.33	0.04
BRG	-^	**100.00**	**100.00**	-^	**100.00**	**100.00**	0.15	0.10
CSB	-^	-^	-^	-^	-^	-^	-^	-^
IDR	-^	**100.00**	**100.00**	-^	**100.00**	**100.00**	0.33	0.06
RAC	-^	-#	-#	-^	-#	-#	-#	-#
BOK	-^	-^	-^	-^	-^	-^	-^	-^
CER	-^	-^	-^	-^	-^	-^	-^	-^
CUP	-^	-^	**100.00**	-^	-^	**100.00**	1.22	0.08
DSK	-^	-^	**100.00**	-^	-^	**100.00**	0.18	0.01
OTS	-^	-^	-^	-^	-^	-^	-^	-^
STR	-^	-^	**100.00**	-^	-^	**100.00**	0.14	0.01
TOH	99.71	**100.00**	99.71	**100.00**	**100.00**	**100.00**	1.05	0.05

Table S6: Reconstruction time on ambiguous datasets with our method. Reconstructions were obtained using COLMAP Schonberger & Frahm (2016). Applying our method removes false edges and sparsifies the viewgraphs, leading to reduced reconstruction time

Dataset	Reconstuction Time (mins) (t_R)					
	$\mathcal{G}$	$\mathcal{G}_{Dopp}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$
$m \rightarrow$			0.6	0.7	0.8	0.9
LVE	95	79	25	25	23	14
NDE	1558	1160	1434	979	703	663
SCO	385	347	268	234	211	191
SEV	203	87	200	139	89	25
ELS	212	33	165	139	116	29
PDP	178	128	148	117	87	47
YKM	383	113	118	116	106	47
	$\mathcal{G}$	$\mathcal{G}_{Dopp}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$	$\mathcal{G}_{FG}$
$m \rightarrow$			0.3	0.4	0.5	0.6
ANC	32	31	20	18	19	15
ADT	22	17	16	16	16	13
BLD	142	126	82	80	82	79
BBN	40	43	37	37	35	28
BRG	8	9	2	2	2	2
CSB	18	16	4	4	4	3
IDR	9	8	1	1	1	1
RAC	18	6	4	6	3	3
BOK	0.3	0.3	0.1	0.1	0.1	0.1
CER	0.5	0.5	0.1	0.1	0.1	0.1
CUP	0.5	0.3	0.2	0.2	0.2	0.1
DSK	1.0	0.8	0.1	0.1	0.1	0.1
OTS	0.4	0.4	0.1	0.1	0.1	0.1
STR	0.2	0.1	0.1	0.1	0.1	0.1
TOH	78	32	62	65	62	62

Table S7: Time taken for different disambiguation methods on ambiguous datasets. Time taken consists of the time required for specific preprocessing for the algorithms and their filtering times. Our methods take significantly less time compared to other methods

Dataset	Time Taken (sec)				
	Cui & Tan (2015)	Yan et al. (2017)	Cai et al. (2023)	Manam & Govindu (2024)	[Ours]
LVE	1261	259681	16009	48	28
NDE	10038	-	139065	5314	2886
SCO	3154	253093	78242	1269	665
SEV	1057	42490	13479	22	20
ELS	27385	1742	12450	34	21
PDP	356	20370	8925	13	11
YKM	868	82653	12748	19	20
ANC	217	174	11203	96	10
ADT	141	168	7715	15	3
BLD	1095	3161	31719	607	107
BBN	164	158	9011	31	5
BRG	67	33	3657	11	1
CSB	147	98	4360	21	3
IDR	78	48	2535	11	1
RAC	167	106	4381	17	2
BOK	12	6	81	1	1
CER	13	8	104	1	1
CUP	37	7	706	1	1
DSK	16	8	167	1	1
OTS	12	6	100	1	1
STR	8	6	72	1	1
TOH	1144	211	10790	35	7

Table S8: Comparison of our method [ST+WT] with Manam & Govindu (2024) [ST] on ambiguous datasets. ST and WT indicate the usage of strong and weak triples, respectively. '-': No reconstruction was obtained. Using both strong and weak triples (our method) disambiguates all datasets, reconstructs more cameras and 3D points (Tables 8 and 9 in the main paper) compared to using only strong triples Manam & Govindu (2024) with similar reprojection errors, resulting in increased reconstruction time

Dataset	Mean Reprojection Errors (px)				Reconstruction Time (mins) (t_R)			
Triples →	ST	ST+WT	ST	ST+WT	ST	ST+WT	ST	ST+WT
	$m = 0.7$		$m = 0.9$		$m = 0.7$		$m = 0.9$	
LVE	0.58	0.60	0.60	0.58	33	25	52	14
NDE	0.71	0.72	0.67	0.69	1023	979	494	663
SCO	0.68	0.69	0.61	0.66	159	234	156	191
SEV	0.65	0.67	-	0.64	13	139	-	25
ELS	0.74	0.75	0.74	0.75	80	139	16	29
PDP	0.66	0.67	0.65	0.69	92	117	16	47
YKM	0.76	0.76	0.42	0.74	82	116	7	47
	$m = 0.4$		$m = 0.5$		$m = 0.4$		$m = 0.5$	
ANC	0.61	0.64	0.60	0.64	14	18	13	19
ADT	0.64	0.65	0.63	0.65	17	16	11	16
BLD	0.67	0.69	0.66	0.68	95	80	96	82
BBN	0.60	0.62	0.59	0.62	33	37	33	35
BRG	0.72	0.73	0.73	0.72	2	2	2	2
CSB	0.52	0.54	0.51	0.53	4	4	4	4
IDR	0.40	0.40	0.39	0.40	1	1	1	1
RAC	0.58	0.62	0.57	0.61	4	6	3	3
BOK	0.31	0.31	0.31	0.31	0.1	0.1	0.1	0.1
CER	0.27	0.27	0.27	0.27	0.1	0.1	0.1	0.1
CUP	0.38	0.38	0.37	0.37	0.2	0.2	0.2	0.2
DSK	0.40	0.39	0.40	0.39	0.2	0.1	0.2	0.1
OTS	0.25	0.25	0.24	0.24	0.1	0.1	0.1	0.1
STR	0.35	0.35	0.33	0.33	0.1	0.1	0.1	0.1
TOH	0.66	0.65	0.67	0.65	51	65	49	62

Table S9: Analysis of superimposed reconstructions of ambiguous datasets with our method using GLOMAP Pan et al. (2024). ✓/✓*/✗: Disambiguated/Disambiguated but oversplit/Non-disambiguated reconstructions

Dataset	Disambiguated			
	$\mathcal{G}$	$\mathcal{G}_{\#in}$	$\mathcal{G}_{Dopp}$	$\mathcal{G}_{FG}$
ANC	✗	✗	✓	✓
ADT	✗	✗	✓	✗
BLD	✗	✗	✗	✗
BBN	✗	✗	✓	✗
BRG	✗	✗	✓	✓
CSB	✗	✗	✗	✓
IDR	✗	✗	✗	✗
RAC	✗	✗	✗	✓
BOK	✗	✗	✗	✓*
CER	✗	✗	✗	-
CUP	✗	✗	✗	✗
DSK	✗	✗	✓	✗
OTS	✗	✗	✗	✓
STR	✗	✗	✗	✗
TOH	✗	✗	✓	✗

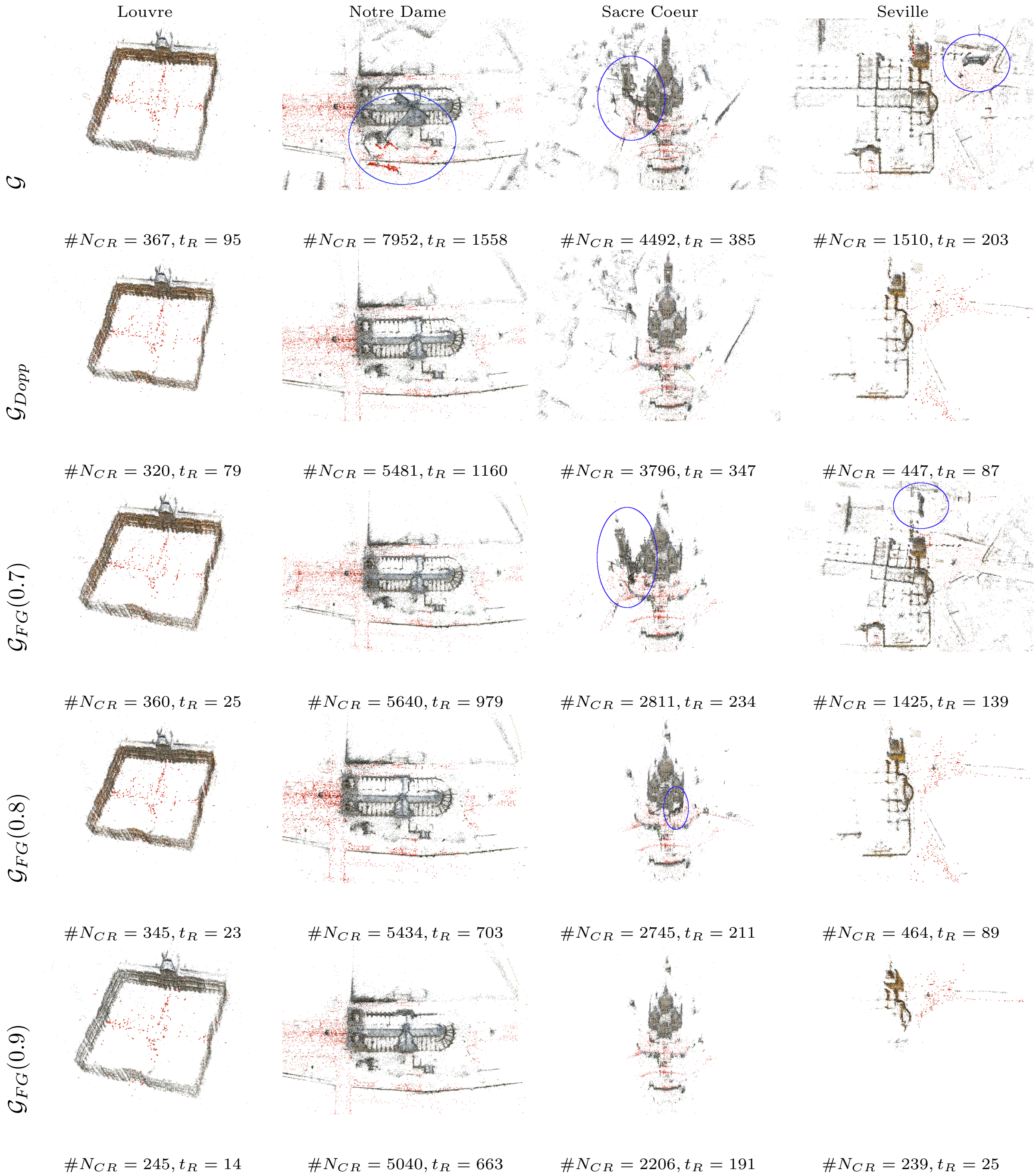

Fig. S5: Reconstructions obtained on ambiguous datasets Wilson & Snavely (2013) with our method. Superimposed parts of the reconstructions are marked in blue. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m) for large-scale datasets: 0.7; for highly ambiguous datasets: $m > 0.8$

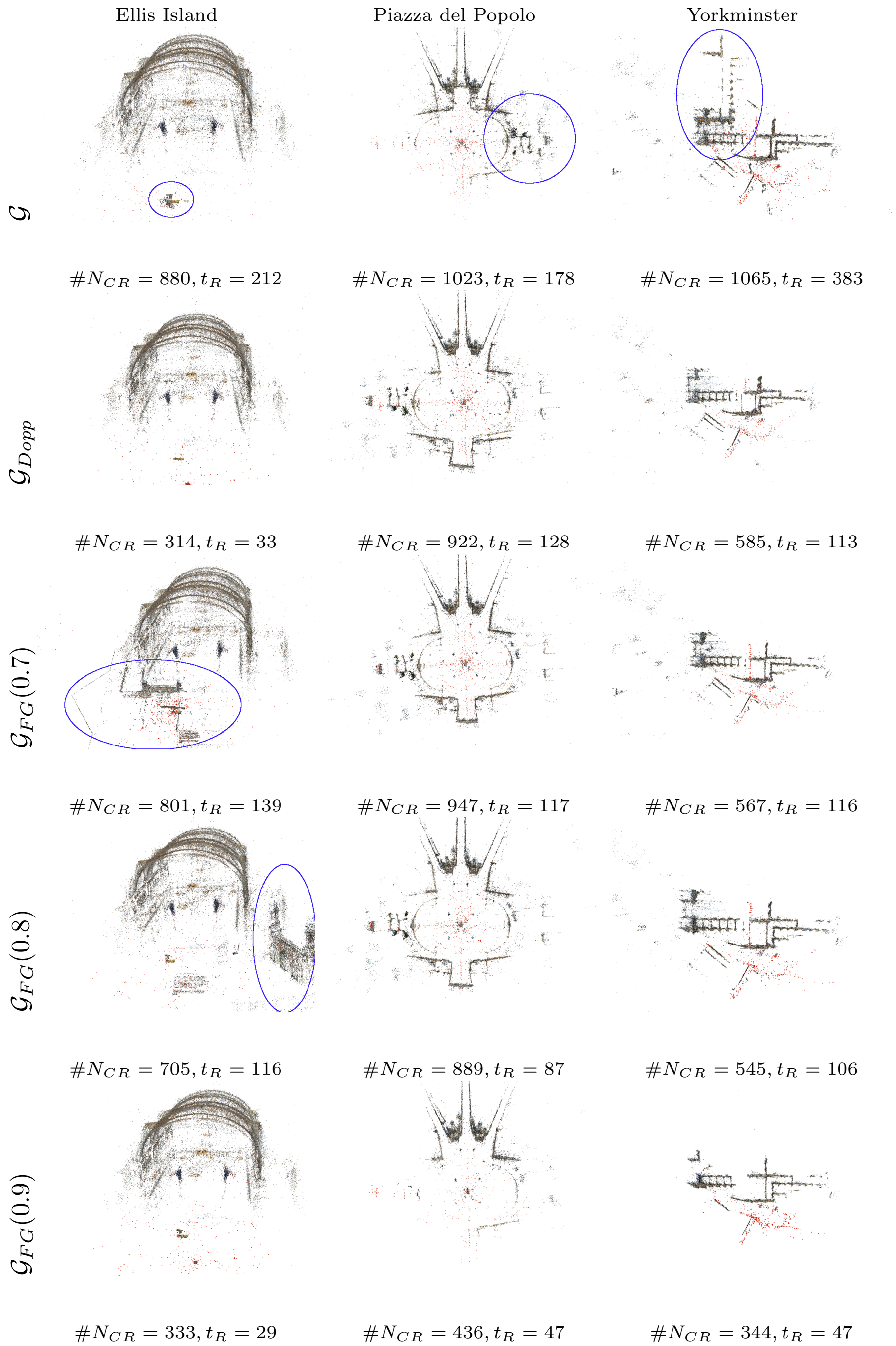

Fig. S6: Reconstructions obtained on ambiguous datasets Wilson & Snavely (2014) with our method. Superimposed parts of the reconstructions are marked in blue. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m) for large-scale datasets: 0.7; for highly ambiguous datasets: $m > 0.8$

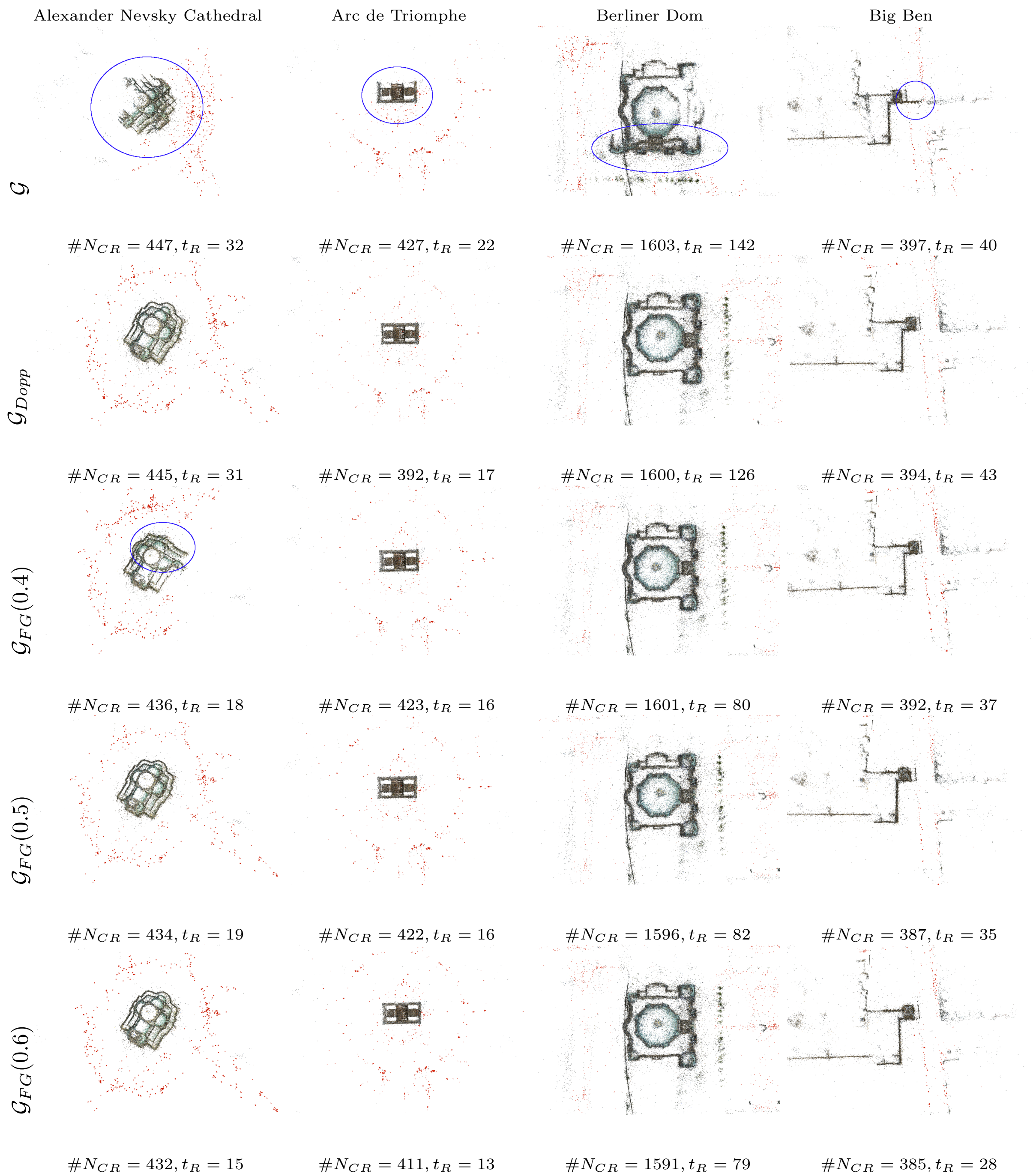

Fig. S7: Reconstructions obtained on ambiguous datasets Heinly et al. (2014) with our method. Superimposed parts of the reconstructions are marked in blue. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m) for medium-scale datasets: 0.5

Fig. S8: Reconstructions obtained on ambiguous datasets Heinly et al. (2014) with our method. Superimposed parts of the reconstructions are marked in blue. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m) for medium-scale datasets: 0.5

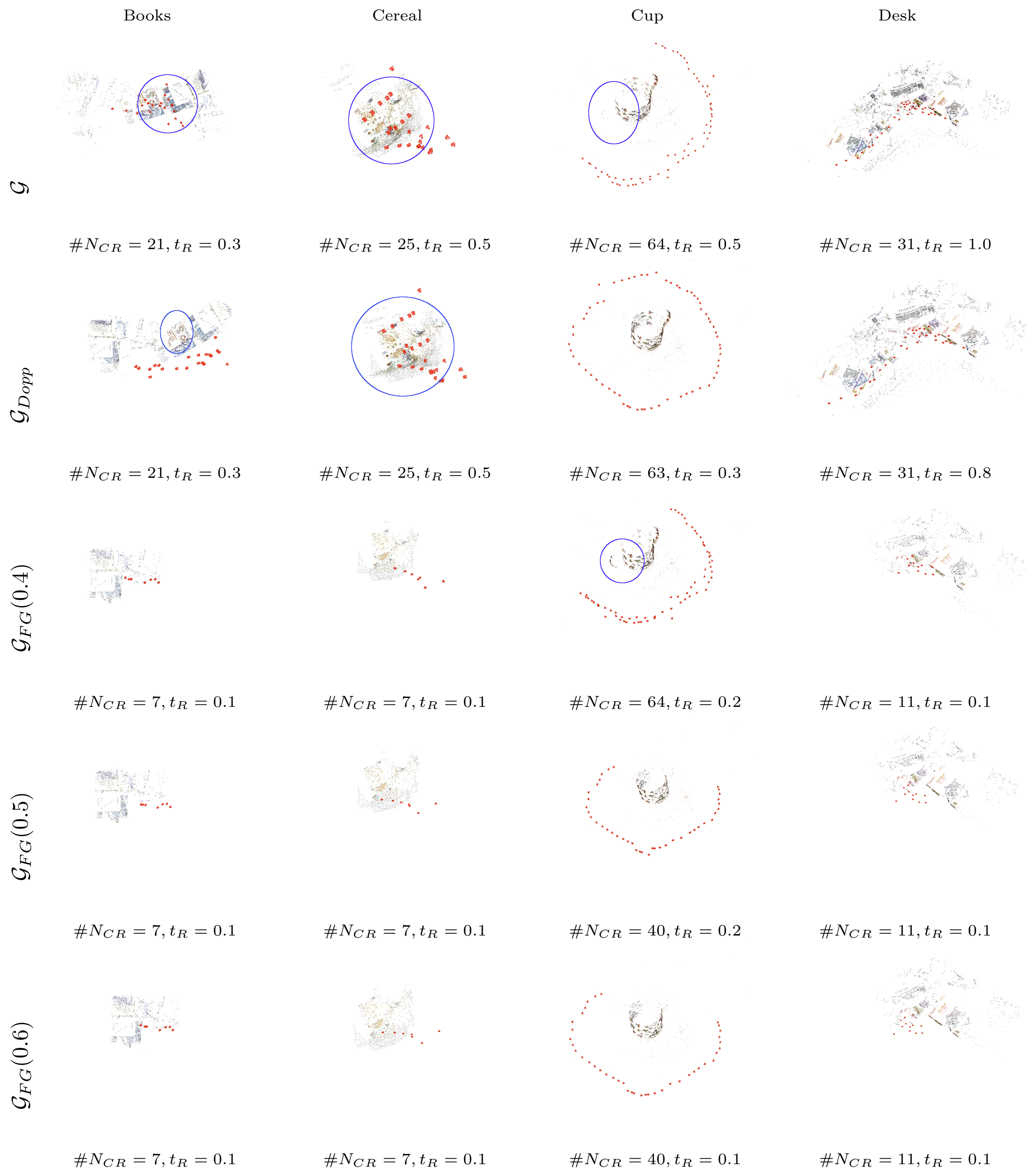

Fig. S9: Reconstructions obtained on ambiguous datasets Yan et al. (2017) with our method. Superimposed parts of the reconstructions are marked in blue. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m) for small-scale datasets: 0.5

Fig. S10: Reconstructions obtained on ambiguous datasets Yan et al. (2017) with our method. Superimposed parts of the reconstructions are marked in blue. $\#N_{CR}$: Number of cameras reconstructed. t_R: Reconstruction time using COLMAP Schonberger & Frahm (2016) in mins. Recommended minimum edge score (m) for small-scale datasets: 0.5

References

Cai, R., Tung, J., Wang, Q., Averbuch-Elor, H., Hariharan, B., Snavely, N. (2023). Doppelgangers: Learning to disambiguate images of similar structures. *International conference on computer vision* (pp. 34–44). IEEE.

Crandall, D., Owens, A., Snavely, N., Huttenlocher, D.P. (2011). Discrete-continuous optimization for large-scale Structure-from-Motion. *Conference on computer vision and pattern recognition* (pp. 3001–3008).

Cui, Z., & Tan, P. (2015). Global Structure-from-Motion by similarity averaging. *International conference on computer vision* (pp. 864–872). IEEE.

Heinly, J., Dunn, E., Frahm, J.-M. (2014). Correcting for duplicate scene structure in sparse 3D reconstruction. *European conference on computer vision* (pp. 780–795).

Li, Y., Snavely, N., Huttenlocher, D.P. (2010). Location recognition using prioritized feature matching. *European conference on computer vision* (pp. 791–804).

Manam, L., & Govindu, V.M. (2024). Leveraging camera triplets for efficient and accurate structure-from-motion. *Conference on computer vision and pattern recognition* (pp. 4959–4968).

Pan, L., Baráth, D., Pollefeys, M., Schönberger, J.L. (2024). Global Structure-from-Motion revisited. *European conference on computer vision* (pp. 58–77).

Schonberger, J.L., & Frahm, J.-M. (2016). Structure-from-Motion revisited. *Conference on computer vision and pattern recognition* (pp. 4104–4113).

Wilson, K., & Snavely, N. (2013). Network principles for SfM: Disambiguating repeated structures with local context. *International conference on computer vision* (pp. 513–520). IEEE.

Wilson, K., & Snavely, N. (2014). Robust global translations with 1DSfM. *European conference on computer vision* (pp. 61–75).

Yan, Q., Yang, L., Zhang, L., Xiao, C. (2017). Distinguishing the indistinguishable: Exploring structural ambiguities via geodesic context. *Conference on computer vision and pattern recognition* (pp. 3836–3844).