Baseline Desensitizing In Translation Averaging Supplementary Material

Bingbing Zhuang, Loong-Fah Cheong, Gim Hee Lee
National University of Singapore
zhuang.bingbing@u.nus.edu, {eleclf,gimhee.lee}@nus.edu.sg

This supplementary material includes implementation details that help understand the main paper and additional experiments that are not presented in the main paper due to space limitation.

1. Implementation Details

(1) Updating of $\boldsymbol{t}_i$ in Algorithm 1: This is a constrained least squares problem

$$\min_{\boldsymbol{t}_i, i\in V} \quad \sum_{ij\in E} W_{ij}||(\boldsymbol{t}_j - \boldsymbol{t}_i)d_{ij} - \boldsymbol{v}_{ij}||_2^2, \tag{1}$$
$$s.t. \quad \sum_{i\in V} \boldsymbol{t}_i = 0; \sum_{ij\in E} < \boldsymbol{t}_j - \boldsymbol{t}_i, \boldsymbol{v}_{ij} >= 1,$$

which can be solved with Cholesky decomposition after applying Lagrange multipliers [1].

(2) We also provide more explanations for the footnote given in Sec. 4.1 of the main paper. We note that the original NRMSE is not a suitable measurement when the camera locations exhibit a multi-cluster distribution. Specifically, in the case of the cameras forming two separate clusters as investigated in the synthetic experiments, increasing L would cause a concomitant decrease of NRMSE. The main reason is that for a large L, the dominant feature of the camera distributions would be this large inter-cluster distance rather than the specific intra-cluster shapes. In the extreme case of L becoming infinitely large, the shape of the ground truth cameras after normalization would be all the cameras collapsing to the two end points of a unit length pole. This also happens to the estimated solution as long as the inter-cluster distance is sufficiently large, thus by definition the NRMSE would approach 0. Therefore, we instead centralize and normalize each cluster separately for both the ground-truth and estimated locations before computing NRMSE; this avoids the bias from the large inter-cluster distance and better captures the accuracy of the intra-cluster shape of the camera locations.

2. Additional Results

In Fig. 1 and 2, we show the comparison of r_1 and r_2 on the synthesized view graphs with all the six different configurations as discussed in the main paper. Clearly, both r_1 and r_2 from LUD are significantly smaller than the ground truth, especially under those more challenging cases with larger noises. This indicates a significant squashing effect on the shape of the recovered camera locations, thus validating our analysis in the main paper.

In Tab. 1, we report the performance on the 1DSfM dataset [2] with two other robust schemes including Huber and Geman-McClure (both with a 0.1 loss width). Comparing those columns without and with rotation involved (i.e. "R.I. w/o R." versus "R.I. w R.", "w/o Rot." versus 'w Rot.'), we can see that involving the reliable absolute rotation solution generally improves the performance in terms of both accuracy and efficiency, regardless of the choice of the robust functions. Since this strategy does not make the algorithm more complicated, it is always recommended in practice.

For completeness, we provide more examples of qualitative evaluation in Fig. 3. We visualize the point clouds obtained by feeding the initial camera poses to the same BA pipeline as described in the main paper. We also plot the estimated camera locations together with the reference ground truth after registration to visualize the error distribution. Overall, the two angle-based methods generally return better results than the magnitude-based counterparts, which might suffer from bias from specific camera distributions. Specifically, for the Roman For. scene, we observe that all the methods seem to work well except for LUD. Despite its success in reconstructing the two walls of the triumphal arch individually, careful inspection from the top-down view reveals that the relative position of the two walls is significantly distorted in the LUD results. For Tow. London, we observe that Shapefit/kick returns largely distorted results. We can also see that BATA reconstructs the boundary wall more completely (highlighted by the red ellipses). This might be due to the fact that BATA recovers those cameras

on the periphery more accurately (highlighted by the red ellipses). For NYC Library, we supplement the main paper by providing the result from 1DSfM as well. We observe that BATA recovers the two sculptures (especially the right one) most clearly. Referring to the camera distribution, we once again observe the superior capability of BATA in dealing with those peripheral cameras (highlighted by the red ellipses). Finally, we present the results on Alamo, on which all the methods perform very well due to its more uniform distribution of cameras, making the problem more well-posed.

Next, we present the full results of the investigation, described in the main paper, of how performance varies against the sparsity of the view graph. In Fig. 4, we plot the median error, 90th pencentile error and ratio of cameras with large error ($>$20m) against the ratio of edges removed, on all the fourteen scenes. As we can see, although 1DSfM achieves lower median errors in some scenes (e.g. Roman For.), BATA generally obtains much lower 90th-percentile error and much smaller number of bad positions. This indicates that no matter what the inherent difficulties of the view graph are, BATA is more resilient to these difficulties than 1DSfM in that it recovers those difficult camera positions much better, especially when the view graph becomes increasingly sparser and more cameras become sparsely connected.

References

[1] S. Boyd and L. Vandenberghe. Vectors, matrices, and least squares. 1

[2] K. Wilson and N. Snavely. Robust global translations with 1dsfm. In *ECCV*, 2014. 1

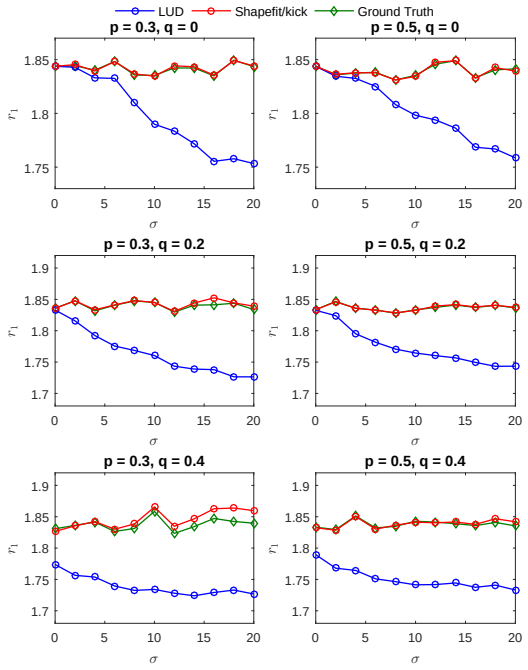

Figure 1: Comparison of r_1 under different view graph setup (p, q) and noise level σ.

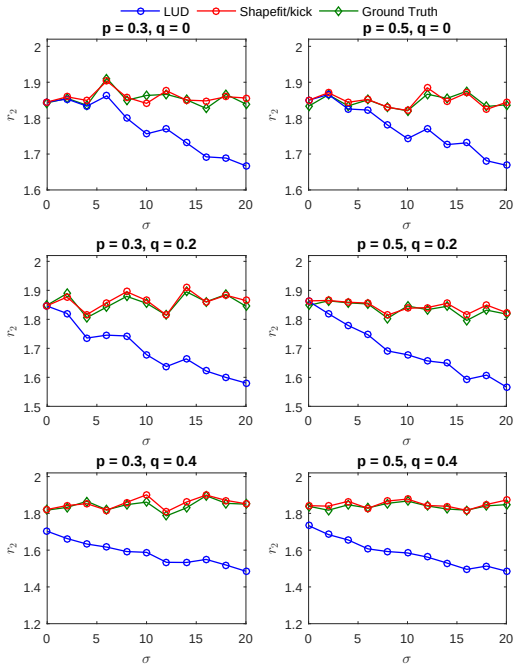

Figure 2: Comparison of r_2 under different view graph setup (p, q) and noise level σ.

Data	BATA-Huber														BATA-Geman-McClure											
	R.I. w/o R.			R.I. w R.			Con. Init.		w/o Rot.			w Rot.			R.I. w/o R.			R.I. w R.			Con. Init.		w/o Rot.			w Rot.		
Name	$\tilde{e}$	$\bar{e}$	#iter	$\tilde{e}$	$\bar{e}$	#iter	$\tilde{e}$	$\bar{e}$	$\tilde{e}$	$\bar{e}$	#iter	$\tilde{e}$	$\bar{e}$	#iter	$\tilde{e}$	$\bar{e}$	#iter	$\tilde{e}$	$\bar{e}$	#iter	$\tilde{e}$	$\bar{e}$	$\tilde{e}$	$\bar{e}$	#iter	$\tilde{e}$	$\bar{e}$	#iter
Piccadilly	2.1	19.5	95	1.4	11.8	51	3.0	5.0	1.2	7.8	100	1.1	5.8	52	9.9	14.9	100	1.0	6.3	100	3.0	5.0	1.0	3.8	100	0.9	3.6	86
Union Sq.	8.3	20.1	100	7.8	19.7	100	6.2	11.9	4.9	14.3	100	4.7	13.1	100	10.0	17.3	100	7.3	14.4	100	6.2	11.9	4.6	11.7	100	4.1	11.4	100
Roman For.	2.3	30.3	100	2.1	29.9	89	9.4	20.8	1.6	10.9	78	1.8	14.0	55	4.5	29.5	100	2.5	20.9	100	9.4	20.8	2.0	9.8	100	2.1	17.2	100
Vienna Cath.	3.2	31.6	82	3.3	26.0	53	6.1	13.1	2.1	10.8	63	2.0	12.5	41	3.2	24.4	100	2.2	16.2	100	6.1	13.1	1.9	9.5	100	1.9	9.7	76
Piazza Pop.	2.4	18.3	99	1.9	12.8	82	1.4	6.5	2.0	7.0	72	2.0	6.3	38	1.9	14.6	100	3.3	9.4	100	1.4	6.5	2.7	5.7	100	2.8	5.4	80
NYC Library	1.1	8.9	84	1.0	9.3	54	1.1	3.3	0.8	3.3	51	0.7	2.9	37	1.6	10.3	100	0.8	12.7	100	1.1	3.3	0.6	2.9	100	0.5	2.7	91
Alamo	0.6	10.4	53	0.6	9.0	35	1.8	3.9	0.7	3.8	51	0.7	3.6	38	0.6	8.1	100	0.5	6.4	90	1.8	3.9	0.6	3.1	74	0.8	3.5	44
Metropolis	4.4	69.8	87	3.1	56.6	59	4.5	15.7	3.2	20.4	81	2.7	14.1	45	10.4	35.9	100	2.1	18.2	100	4.4	15.7	1.9	13.1	99	1.7	11.5	82
Yorkminster	1.5	74.6	100	1.4	45.0	92	4.4	12.8	1.3	9.0	77	1.2	8.8	62	1.2	17.7	100	0.9	11.3	100	4.4	12.8	1.1	7.6	100	0.9	7.6	100
Montreal N.D.	0.9	3.6	61	0.9	3.0	51	1.0	1.7	0.4	0.8	39	0.4	0.8	28	1.7	6.4	100	1.3	3.3	98	1.0	1.7	0.4	1.0	93	0.4	0.9	75
Tow. London	2.6	34.7	100	2.5	28.2	80	5.1	22.9	2.3	19.9	81	2.2	21.5	53	4.2	35.8	100	2.2	18.3	100	5.1	22.9	2.0	15.0	100	2.0	17.0	100
Ellis Island	1.6	25.0	65	1.6	20.8	43	2.2	9.7	1.5	12.1	44	1.5	16.4	50	1.7	19.2	100	1.3	10.6	80	2.2	9.7	1.1	7.9	100	1.2	7.7	62
Notre Dame	0.4	9.2	77	0.2	6.4	54	3.1	4.1	0.3	2.2	64	0.3	2.0	41	0.5	5.8	100	0.3	3.0	100	3.1	4.1	0.2	1.3	100	0.2	1.5	100
Trafalgar	7.0	45.7	91	5.8	41.8	73	8.8	14.7	4.5	13.1	65	4.1	12.1	43	10.1	2e3	100	3.7	19.9	100	8.8	14.7	3.8	11.2	100	3.1	11.2	98

Table 1: Additional results from BATA with Huber and Geman-McClure as the robust schemes.

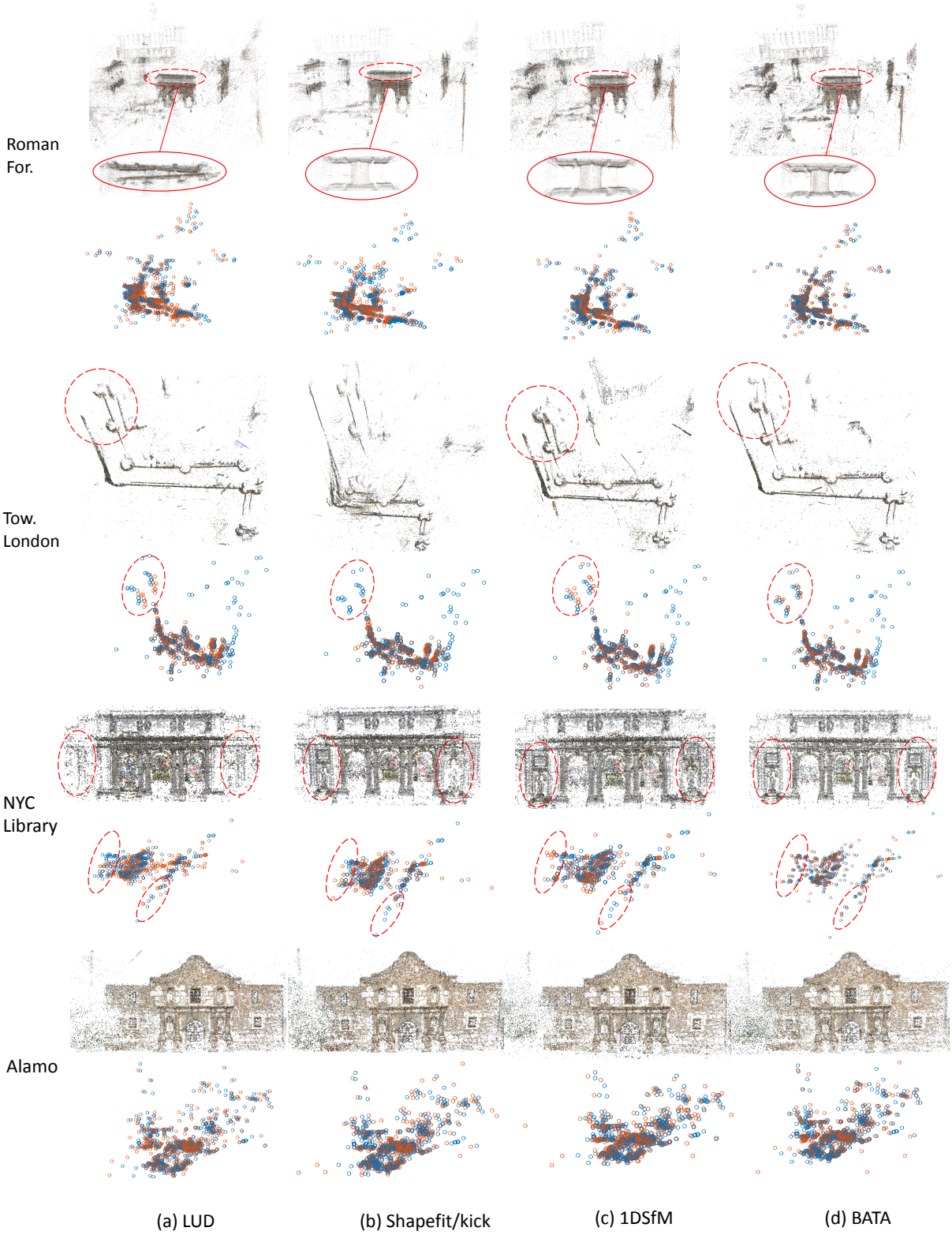

Figure 3: Qualitative evaluation on four different scenes from different methods. For each scene, the point clouds after BA is presented in the first row. In the second row, the estimated camera locations (orange circles) are plotted together with the reference ground truth (blue circles) to visualize the errors. Note that those cameras with large errors (>50m) are not plotted so that subtle position difference for the rest of the cameras can be better visualized.

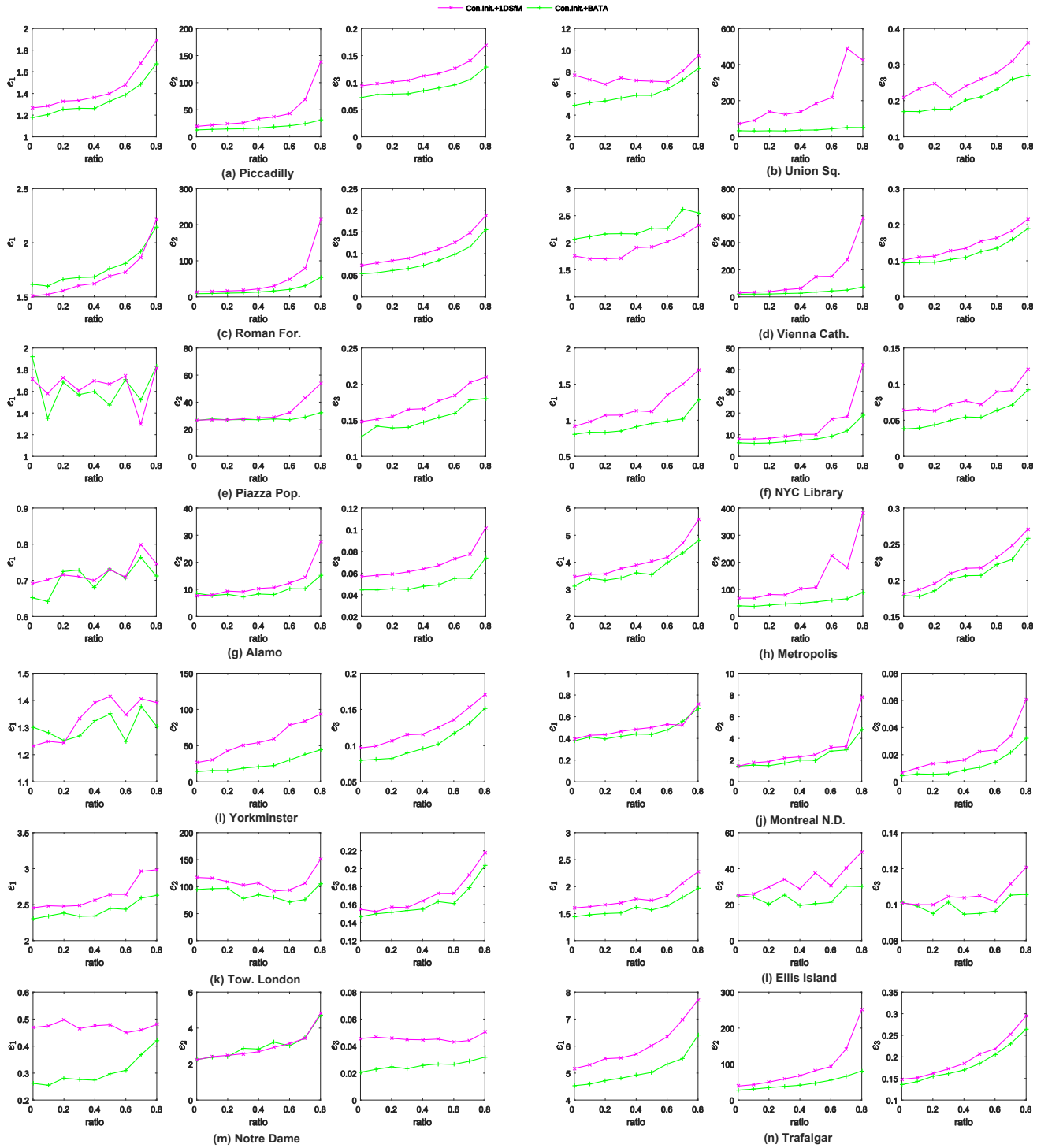

Figure 4: Different error quantities plotted against the ratio of the observed edges removed. e_1, e_2 and e_3 respectively denote the median error, 90th pencentile error, and ratio of cameras with large error ($>$20m).