[Supplementary Material]
HARA: A Hierarchical Approach for Robust Rotation Averaging

Seong Hun Lee Javier Civera
I3A, University of Zaragoza, Spain
{seonghunlee, jcivera}@unizar.es

1. Ablation study

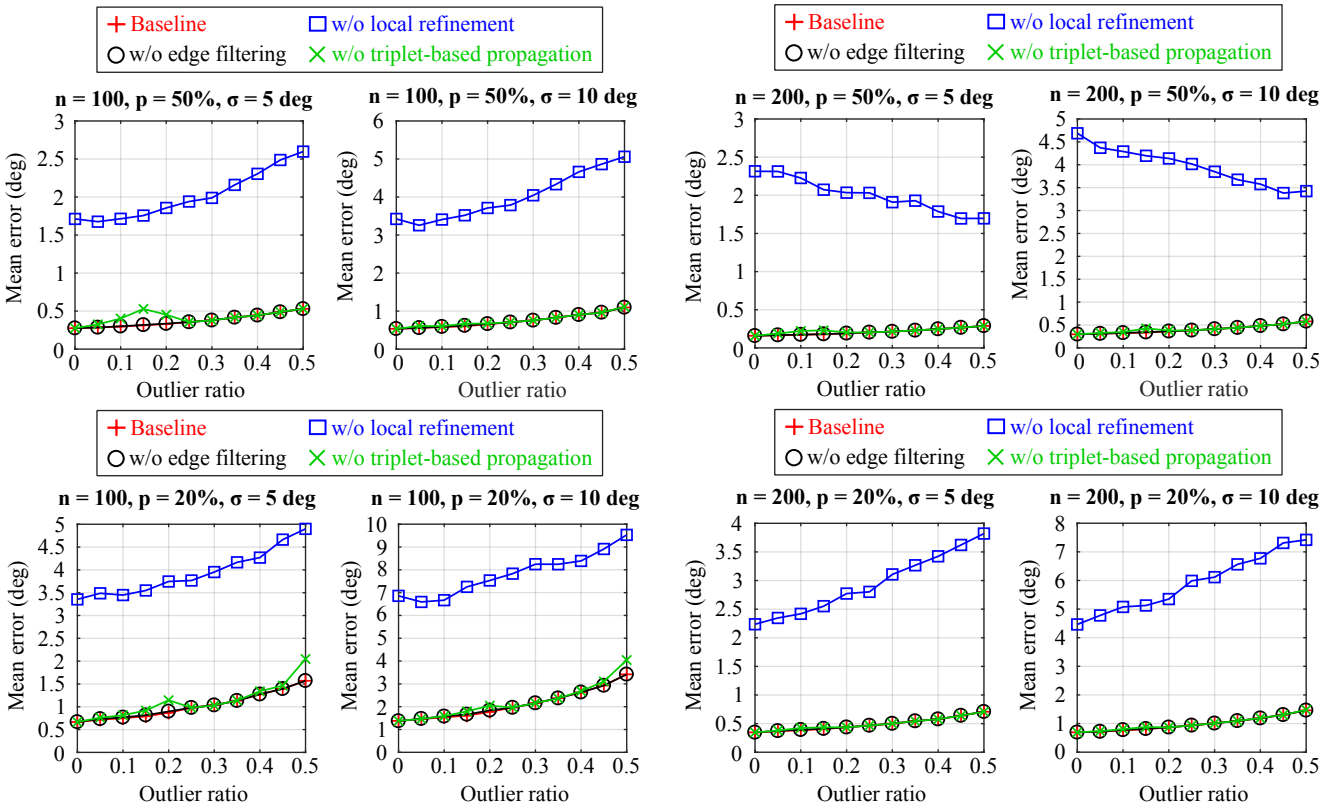
We perform an ablation study to see how each component of HARA contributes to the final accuracy. We compare three different variations of HARA against the baseline version:

1. HARA without the local refinement (Section 4.4),
2. HARA without the edge filtering (Section 4.3),
3. HARA without the triplet-based propagation (Alg. 2.7–2.25): That is, the initial solution is obtained via a series of voting + single rotation averaging only.

We use the same datasets as in the main paper (see our main paper for the experimental setup). Fig. 1 and Table 1 present the results on the synthetic and the real datasets, respectively. These results clearly show that the best performance can be achieved by utilizing all the components.

2. Informative Q&A from the review-and-rebuttal process

1. *What happens to cameras that do not belong to any triplet?*
   → It will eventually be estimated via single rotation averaging (line 26 of Alg. 2), so the final number of estimated rotations is the same.
2. *How difficult is it to tune the support threshold (s)?*
   → As s is adaptive, only s_{init} needs to be tuned. Inevitably, there is a trade-off between robustness (large s_{init}) and speed (small s_{init}). For most datasets we tested, the sweet spot was around $s_{\text{init}} = 10$, and larger values only made a small difference.
3. *How much does [40] affect the accuracy?*
   → Since the majority of cameras are added by checking the triplet support than by single rotation averaging [40], it does not affect the accuracy significantly. Using the single rotation averaging method in [33] would have produced similar results, but we chose [40] as an additional safety measure against outliers.
4. *Why exactly is the proposed method better than [13] in Table 2?*
   → The fundamental reason is that, while [13] relies only on the number of inlier matches for initialization, we use both the number of inlier matches AND the hierarchy of triplet supports. This enables us to initialize our solution with the most accurate edges first.
5. *What are the preprocessing required other than triplet sampling? Is this included in the reported runtime?*
   → The reported runtime includes the triplet sampling (described at the end of Section 4.2, corresponding to line 3 of Alg. 2). No other preprocessing is required.
6. *In line 11 of Alg. 1 and line 15 of Alg. 2, what if there are multiple nodes with he same number of non-family neighbors?*
   → We simply choose the first one in the list, as it does not really matter.
7. *Why set such low thresholds for 2D-2D correspondences?*
   → This is because the 1DSfM datasets contain edges with very few (< 5) valid correspondences. We were also surprised by this fact, and we are suspecting that a mistake had been made in the dataset itself. In any case, you can think of our low threshold as a safety measure to filter remaining outliers in the edges.
8. *What is the typical value for each ϵ_i?*
   → For 1DSfM datasets, on average, $\epsilon_1 \approx 0.001$, $\epsilon_2 \approx 0.003$, $\epsilon_3 \approx 0.007$.



(a) 100 rotations (b) 200 rotations

Figure 1: Ablation study on the synthetic dataset.

Datasets			HARA (baseline)			w/o local refinement (Sect. 4.4)			w/o edge filtering (Sect. 4.3)			w/o triplet-based (Alg. 2.7–2.25)		
Name	#views	%edges	θ_1	θ_2	Time	θ_1	θ_2	Time	θ_1	θ_2	Time	θ_1	θ_2	Time
ALM	627	49.5%	**3.5**	**11.5**	41s	4.4	12.6	26s	4.1	12.4	51s	3.9	12.3	30s
ELS	247	66.8%	**2.1**	**7.4**	6s	2.6	7.8	5s	2.5	7.7	9s	3.0	11.6	4s
GDM	742	17.5%	43.8	72.5	21s	44.1	72.0	20s	**37.7**	**62.4**	33s	46.6	71.7	12s
MDR	394	30.7%	**4.8**	**14.5**	14s	6.8	15.2	13s	6.5	16.4	13s	7.7	20.2	5s
MND	474	46.8%	**1.1**	**2.1**	28s	1.6	2.8	26s	1.5	7.4	23s	1.4	7.5	11s
ND1	553	68.1%	**1.6**	**6.3**	48s	2.3	6.7	38s	3.2	12.4	55s	3.7	15.8	31s
NYC	376	29.3%	**2.9**	7.7	10s	3.3	8.0	9s	3.0	**7.0**	8s	3.3	8.6	6s
PDP	354	39.5%	**3.4**	**7.4**	7s	3.5	7.5	6s	4.0	8.0	10s	3.6	8.5	5s
PIC	2508	10.2%	**4.4**	**13.1**	279s	5.7	13.7	140s	5.5	14.5	437s	5.9	18.4	220s
ROF	1134	10.9%	**2.7**	8.5	31s	3.3	9.1	26s	3.0	8.6	30s	**2.7**	**7.9**	15s
TOL	508	18.5%	4.3	10.0	8s	4.6	10.3	8s	**4.0**	**9.2**	14s	4.7	11.6	4s
TFG	5433	4.6%	**3.5**	10.7	924s	5.5	11.6	325s	3.6	10.0	1049s	3.6	**9.7**	1014s
USQ	930	5.9%	**6.0**	12.3	8s	7.1	14.1	7s	7.3	14.7	11s	6.1	**11.4**	5s
VNC	918	24.6%	**6.1**	**18.1**	52s	6.6	18.6	40s	8.0	26.3	56s	8.2	28.3	32s
YKM	458	26.5%	**3.0**	6.9	17s	3.1	**6.5**	16s	3.5	8.4	14s	3.6	9.6	5s
ND2	715	25.3%	1.3	5.5	23s	1.7	5.5	19s	**1.1**	**3.5**	31s	1.3	5.3	17s
ACP	463	10.7%	**1.2**	**1.7**	6s	2.0	2.4	6s	**1.2**	**1.7**	4s	**1.2**	**1.7**	3s
ARQ	5530	1.5%	**3.6**	6.8	136s	5.1	8.1	104s	3.7	**6.6**	169s	4.4	11.2	98s
SNF	7866	0.3%	**3.6**	**4.2**	35s	4.8	6.7	32s	**3.6**	**4.2**	44s	**3.6**	**4.2**	17s

Table 1: Ablation study on the real datasets **without** the knowledge of the 2D-2D correspondences.