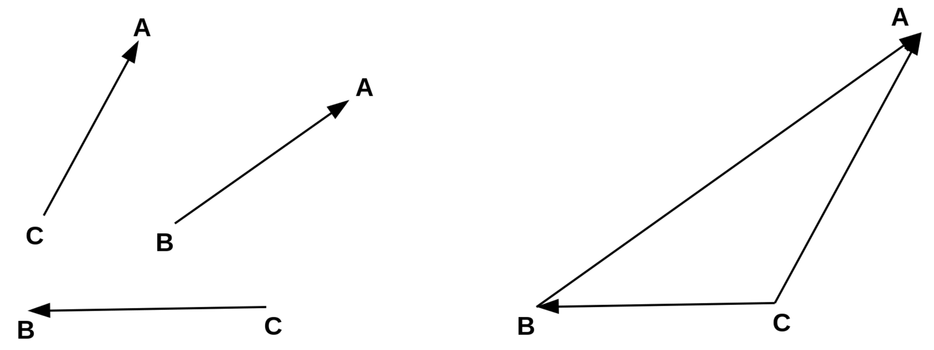
Correspondence Reweighted Translation Averaging

Lalit Manam and Venu Madhav Govindu

Computer Vision Lab, Department of Electrical Engineering, Indian Institute of Science, Bengaluru, INDIA

ECCV TEL AVIV 2022

Translation Averaging

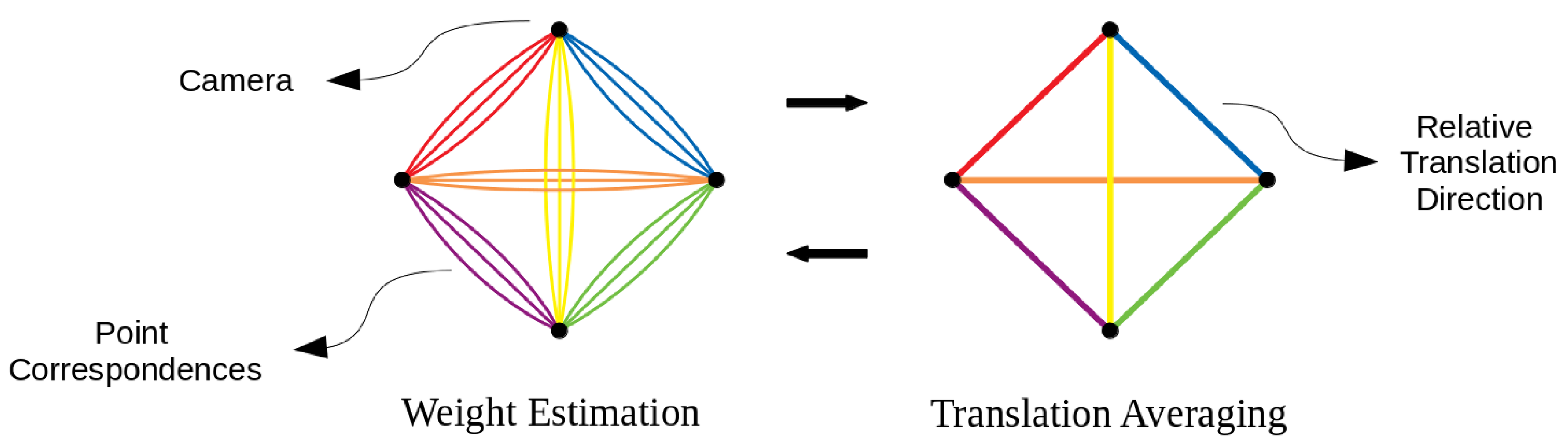


- Solving for camera translations given pairwise relative translation directions.
- Unique upto scale and choice of origin.
- Assumes known rotations.

Motivation

- Translation averaging methods use consistency of input directions.
- Translation directions obtained from Epipolar Geometry are error prone.
- Performance of averaging methods limited by the input quality.
- Improved accuracy of translation averaging should be leveraged to mitigate the errors in input translation directions.

Proposed Framework



- Given an initial set of translation directions, solve for camera translations.
- Obtain globally consistent translation directions from the estimated camera translations.
- Weight point correspondences in every edge based on the globally consistent directions.
- Get new estimate of translation directions based on the weighted point correspondences.
- Solve for camera translations with the new estimate of directions and repeat the process.

Notations

- Underlying viewgraph: $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- Global rotation and translation: $\mathbf{R}_i \in \mathbb{SO}(3), \mathbf{T}_i \in \mathbb{R}^3, \forall i \in \mathcal{V}$
- Relative rotation and translation direction:

$$\mathbf{R}_{ij} = \mathbf{R}_j \mathbf{R}_i^{-1}, \mathbf{t}_{ij} = \frac{\mathbf{R}_j(\mathbf{T}_i - \mathbf{T}_j)}{\|\mathbf{R}_j(\mathbf{T}_i - \mathbf{T}_j)\|_2}, \forall (i,j) \in \mathcal{E} \quad (1)$$

- Relative translation direction in global reference frame:

$$\mathbf{v}_{ij} = -\mathbf{R}_j^{-1}\mathbf{t}_{ij} = \frac{\mathbf{T}_j - \mathbf{T}_i}{\|\mathbf{T}_j - \mathbf{T}_i\|_2}, \forall (i,j) \in \mathcal{E} \quad (2)$$

- k^{th} correspondence in the edge (i,j) : $\mathbf{p}_i^k$ and $\mathbf{q}_j^k$

Method

Epipolar Geometry:

$$\left(\mathbf{q}_j^k\right)^T \left(\mathbf{t}_{ij} \times \mathbf{R}_{ij}\mathbf{p}_i^k\right) = 0 \quad (3)$$

$$\Rightarrow \left(\mathbf{m}_{ij}^k\right)^T \mathbf{v}_{ij} = 0 \text{ (with known rotations)} \quad (4)$$

Our Formulation:

$$\min_{\mathbb{T}} \sum_{(i,j)\in\mathcal{E}} \|\mathbf{W}_{ij}\mathbf{M}_{ij}\mathbf{v}_{ij}(\mathbb{T})\|_2^2 \quad (5)$$

- $\mathbf{v}_{ij}$: dependent on global translations $\mathbb{T}$.
- $\mathbf{W}_{ij}$: diagonal matrix of weights $\mathbf{w}_{ij}^k$ based on global consistency of translation directions.

Algorithm:

Correspondence Reweighted Translation Averaging (CReTA)

1 Initialize global translations $\mathbb{T}$
2 **while** not converged **do**
3 Update weights $\mathbf{w}_{ij}^k \ \forall(\mathbf{p}_i^k, \mathbf{q}_j^k)$
4 Estimate $\{\mathbf{v}_{ij} | \forall(i,j) \in \mathcal{E}\}$
5 Solve Translation Averaging using $\{\mathbf{v}_{ij}\}$
6 **end**

Choice of translation averaging methods:

Revised LUD [2, 4] (compares displacement vectors):

$$\min_{\mathbf{T}_{i,i\in\mathcal{V}}, \lambda_{ij,(i,j)\in\mathcal{E}}} \|\mathbf{T}_j - \mathbf{T}_i - \lambda_{ij}\mathbf{v}_{ij}\|_2 \quad (6)$$

$$\text{s.t.} \sum_{i\in\mathcal{V}} \mathbf{T}_i = \mathbf{0}, \sum_{(i,j)\in\mathcal{E}} \langle \mathbf{T}_j - \mathbf{T}_i, \mathbf{v}_{ij} \rangle = 1, \lambda_{ij} \geq 0, \ \forall (i,j) \in \mathcal{E}$$

BATA [4] (compares directions):

$$\min_{\mathbf{T}_{i,i\in\mathcal{V}}, \gamma_{ij,(i,j)\in\mathcal{E}}} \rho(\|(\mathbf{T}_j - \mathbf{T}_i)\gamma_{ij} - \mathbf{v}_{ij}\|_2) \quad (7)$$

$$\text{s.t.} \sum_{i\in\mathcal{V}} \mathbf{T}_i = \mathbf{0}, \sum_{(i,j)\in\mathcal{E}} \langle \mathbf{T}_j - \mathbf{T}_i, \mathbf{v}_{ij} \rangle = 1, \gamma_{ij} \geq 0, \ \forall (i,j) \in \mathcal{E}$$

Results

Dataset	$\|\mathcal{V}\|$	$\|\mathcal{E}\|$	LUD [2]		ShapeFit [1]		BATA [4]		CReTA-RLUD (Ours)		CReTA-BATA (Ours)	
			μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$
Alamo (ALM)	586	81437	2.7	**0.5**	**0.9**	**0.5**	2.0	0.6	2.0	**0.5**	2.0	**0.5**
Ellis Island (ELS)	229	14728	6.9	3.6	12	**1.9**	6.7	3.2	**6.0**	2.9	6.2	3.3
Gendarmenkmart (GMM)	686	27145	31.2	11.3	-	-	31.3	11.4	**31.0**	11.2	31.5	**11.1**
Madrid Metropolis (MDR)	325	11995	8.4	1.9	81	6.0	6.9	**1.6**	7.6	1.7	**6.1**	**1.6**
Montreal Notre Dame (MND)	461	45737	0.9	0.5	1.7	0.8	0.8	0.5	0.9	0.5	**0.7**	**0.4**
Notre Dame (ND)	552	80647	1.2	0.3	1.5	**0.2**	**1.0**	**0.2**	2.1	1.1	**1.0**	**0.2**
NYC Library (NYC)	337	14365	2.2	0.8	162	1.4	2.1	1.7	**2.0**	**0.7**	**2.0**	**0.7**
Piazza del Popolo (PDP)	334	20974	3.8	2.8	5.9	3.6	**3.4**	**2.0**	4.5	3.4	3.8	2.5
Piccadilly (PIC)	2362	201600	**2.8**	1.3	15	1.2	3.2	**1.1**	**2.8**	**1.1**	3.0	**1.1**
Roman Forum (ROF)	1069	54207	11.9	3.3	25	4.3	8.3	2.0	13.3	3.3	**7.7**	**1.7**
Tower of London (TOL)	474	19252	14.9	3.2	164	**2.3**	9.3	3.0	13.3	3.1	**9.0**	3.0
Trafalgar (TFG)	4900	542480	8.4	5.3	-	-	7.9	4.2	8.0	4.4	**7.5**	**4.1**
Union Square (USQ)	825	19899	10.6	6.1	47	8.9	**10.2**	5.6	10.6	5.4	10.5	**4.9**
Vienna Cathedral (VNC)	826	82793	**5.1**	2.1	11	**1.9**	12.0	2.1	6.5	2.2	6.7	**1.9**
Yorkminster (YKM)	430	22692	7.6	1.8	-	-	**5.1**	**1.3**	7.0	1.6	6.0	2.0

Table: Camera translation errors (in meters) on 1DSfM [3] datasets

Results

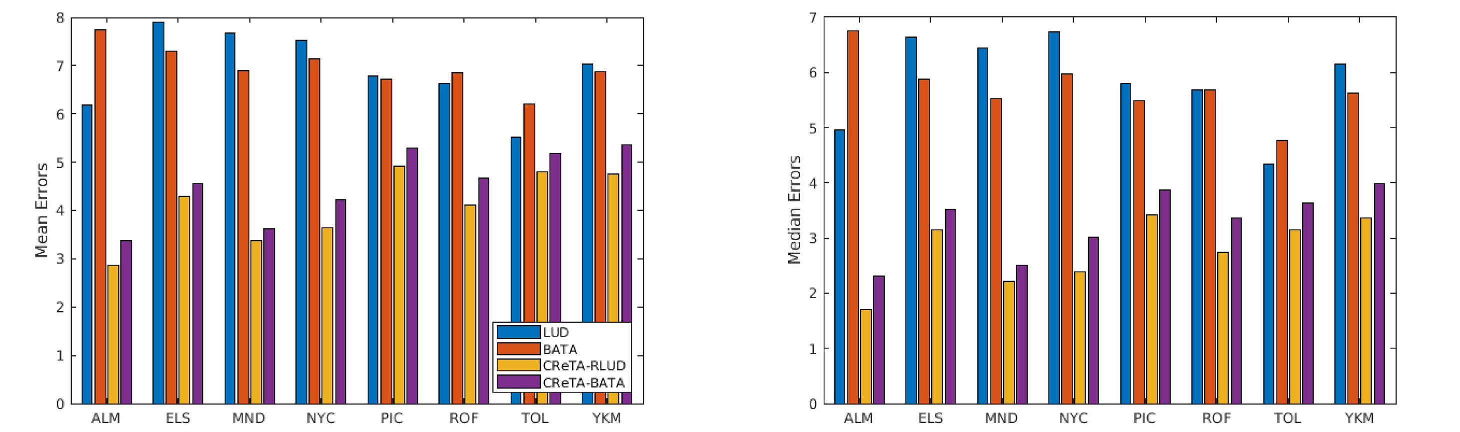


Figure: Reprojection errors (in pixels) after triangulation on 1DSfM datasets using the translation solution obtained from dataset provided initialization

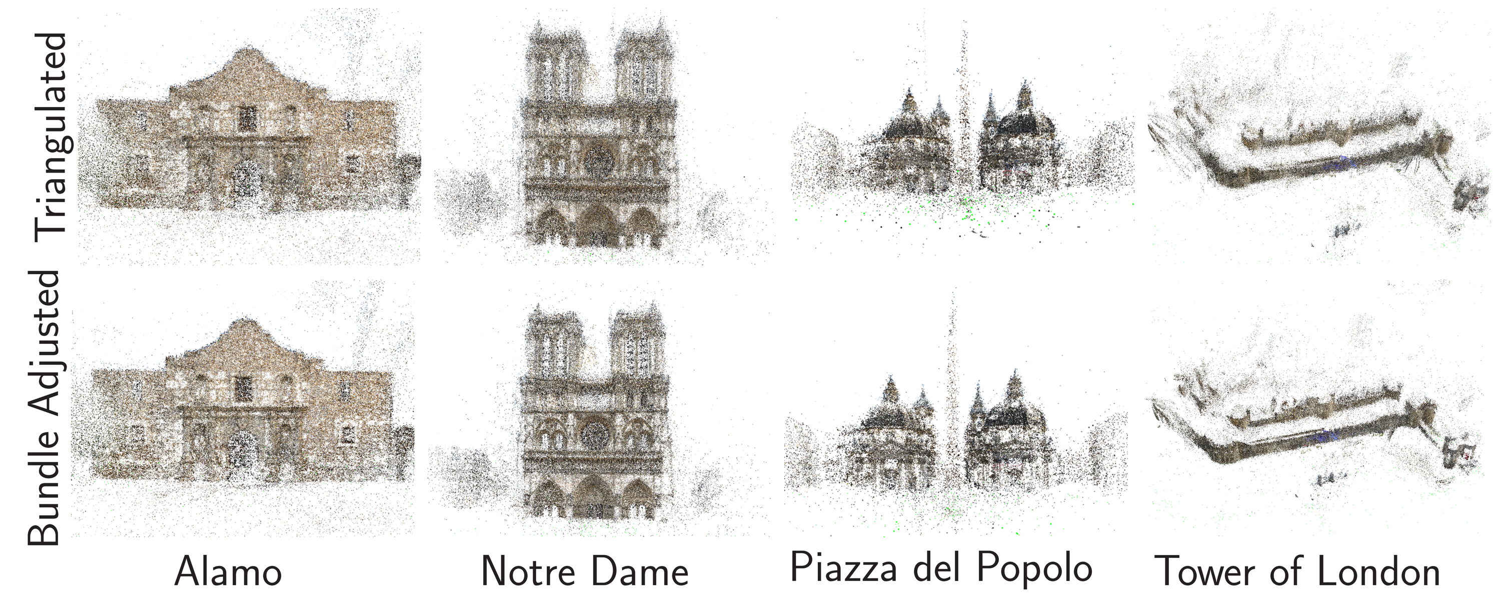


Figure: Reconstructions obtained with triangulation using our CReTA-RLUD translation estimate (first row) compared to bundle adjustment (second row)

Conclusion

We propose CReTA, a framework for incorporating point correspondences in translation averaging problem.

- Iteratively refines global translations by reweighting point correspondences using global consistency.
- Modular framework, can be applied to any existing translation averaging method.
- Improves the reprojection errors with triangulation without any use of bundle adjustment.

Acknowledgments

Lalit Manam is supported by a Prime Minister's Research Fellowship, Government of India. This research was supported in part by a Core Research Grant from Science and Engineering Research Board, Department of Science and Technology, Government of India.

References

[1] Goldstein, T., Hand, P., Lee, C., Voroninski, V., Soatto, S.: Shapefit and shapekick for robust, scalable structure from motion. In: European Conference on Computer Vision. pp. 289–304. Springer (2016)

[2] Ozyesil, O., Singer, A.: Robust camera location estimation by convex programming. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2674–2683 (2015)

[3] Wilson, K., Snavely, N.: Robust global translations with 1dsfm. In: European Conference on Computer Vision. pp. 61–75. Springer (2014)

[4] Zhuang, B., Cheong, L.F., Lee, G.H.: Baseline desensitizing in translation averaging. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4539–4547 (2018)