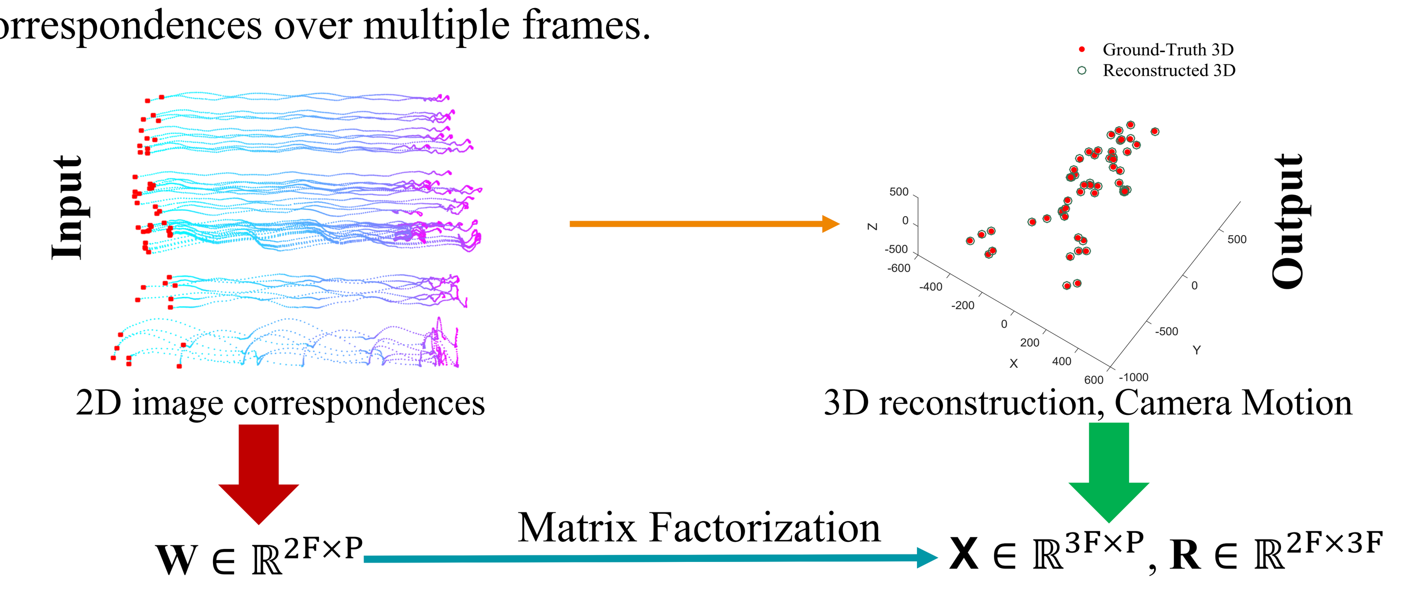
Organic Priors in Non-Rigid Structure from Motion

SURYANSH KUMAR AND LUC VAN GOOL

sukumar@vision.ee.ethz.ch; vangool@vision.ee.ethz.ch

GOAL

- 3D reconstruction of a non-rigidly deforming object using image feature correspondences over multiple frames.



MOTIVATION

- Reinvestigate, improve and enhance the understanding and performance of the classical non-rigid structure from motion factorization.
- Discover novel patterns for solving non-rigid structure from motion with minimal assumptions. Not only for theoretical curiosity but also for practical application.
- Conventional Deep neural network approaches has not shown remarkable progress for this inverse problem.
- Evaluation of progress in non-rigid structure from motion since the seminal matrix factorization work.

ASSUMPTION

- The deforming shape lies in a low-rank space.
- Orthographic camera model.

ORGANIC PRIORS

- **Definition**: Intermediate priors that reside in the factorized matrices, which are the results of natural mathematical steps in NRS*f*M matrix factorization. Such priors carry over within the intermediate factorized matrices, and its existence is independent of camera motion and shape deformation type.
- **For rotation estimation**: column triplets in the correction matrix.
- **For shape estimation**: singular values in the initial pseudo inverse shape.

CONTRIBUTIONS

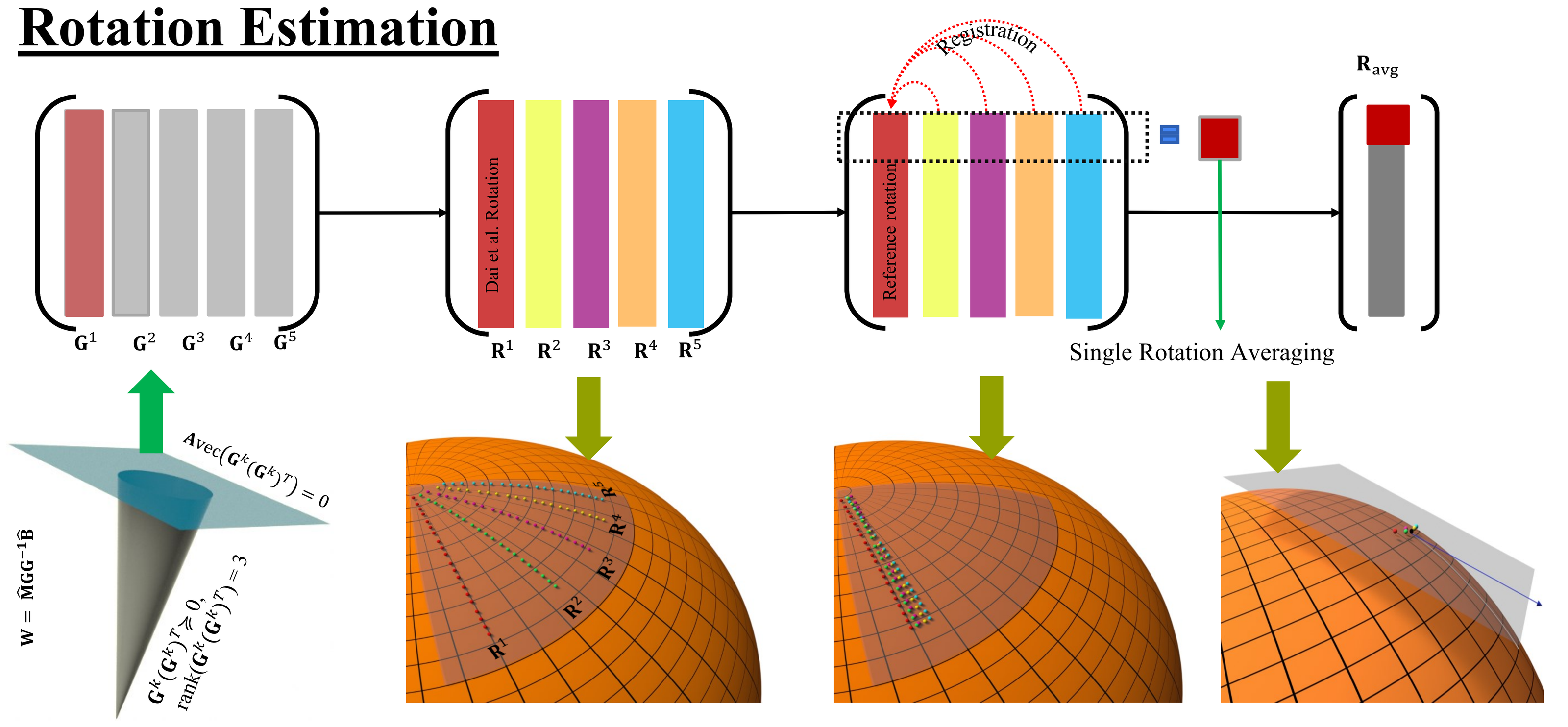
- A methodical approach for solving NRS*f*M that provides outstanding results using simple matrix factorization under low-rank shape assumption.
- An approach that endorses the use of organic prior rather than extraneous priors or assumptions. Our method introduce single rotation averaging to estimate better rotation while being free from smooth camera motion heuristics.
- A different setup for low-rank shape optimization is proposed. We present a blend of partial sum minimization of singular values theory and weighted nuclear norm optimization for shape recovery. We observed that the proposed optimization better exploits the introduced organic shape priors and yields shape reconstruction superior to other NRS*f*M matrix factorization methods.

PROPOSED APPROACH

Master Equation

$\mathbf{W} = \mathbf{R}\mathbf{X}$; $\quad \widehat{\mathbf{M}}\widehat{\mathbf{B}} = \mathbf{factorize}(\mathbf{W})$ where, $\widehat{\mathbf{M}} \in \mathbb{R}^{2F\times 3K}$, $\widehat{\mathbf{B}} \in \mathbb{R}^{3K\times P}$

Rotation Estimation



(a) Steps for rotation Estimation.

- Recover $\mathbf{G}^{k} \in \mathbb{R}^{3K\times 3}$ using $\mathbf{svd}(\mathbf{Q}_k)$ and its corresponding $\mathbf{R}^{k} \in \mathbb{R}^{2F\times 3}$.
- Map $\mathbf{R}^{k} \in \mathbb{R}^{2F\times 3}$ to $\mathbf{R}^{k} \in \mathbb{R}^{3F\times 3}$ via cross product (correcting for sign).
- Register per frame rotation assuming $\mathbf{R}^{1}$ as the reference.
- Perform per frame single rotation averaging « Weiszfeld approach » to estimate rotation.
- Arrange the recovered rotation $\mathbf{R}^{avg} \in \mathbb{R}^{3F\times 3}$ into block diagonal structure $\mathbf{R} \in \mathbb{R}^{2F\times 3F}$.

Structure Estimation

$$\underset{\mathbf{X},\mathbf{X}^{\#}}{\text{minimize}} \frac{1}{2}||\mathbf{W} - \mathbf{R}\mathbf{X}||_{\mathrm{F}}^{2} + \mu||\mathbf{X}^{\#}||_{*,\theta} \qquad \text{subject to: } \mathbf{X}^{\#} = \boldsymbol{\phi}(\mathbf{X})$$

where, $\theta_i = \dfrac{\xi}{\sigma_i(\mathbf{X}^{\#}_{init}) + \gamma}$; ξ, γ are small positive constant; $\mathbf{X}_{init} = \text{pinv}(\mathbf{R})\mathbf{W}$

« *Further Observation*: First singular value of recovered $\mathbf{X}^{\#}_{init}$ is generally good. » Preserve it during minimization.

Combination WNN [2] and PSVT minimization.

$$\underset{\mathbf{X},\mathbf{X}^{\#}}{\text{minimize}} \frac{1}{2}||\mathbf{W} - \mathbf{R}\mathbf{X}||_{\mathrm{F}}^{2} + \mu||\mathbf{X}^{\#}||_{r=1,\theta} \qquad \text{subject to: } \mathbf{X}^{\#} = \boldsymbol{\phi}(\mathbf{X})$$

KEY REFERENCES

1. Y. Dai, H. Li, and M. He, "A simple prior-free method for non-rigid structure-from-motion factorization" IEEE CVPR 2012, IJCV 107(2):101–122, 2014. (BMM).
2. S.Kumar, "Non-rigid structure from motion: Prior-free factorization method revisited" IEEE, WACV 2020, pages 51-60 (R-BMM).
3. I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Non-rigid structure from motion in trajectory space", NIPS 2009, pages 41–48 (PTA).
4. Iglesias, J.P., Olsson, C., Valtonen Örnhag, M.: "Accurate optimization of weighted nuclear norm for non-rigid structure from motion" ECCV 2020, pages 21–37 (AOW).
5. Jensen, S.H.N., Doest, M.E.B., Aanæs, H., Del Bue, A. "A benchmark and evaluation of non-rigid structure from motion" International Journal of Computer Vision (IJCV) 129(4), 882–899 (2021).

STATISTICAL EVALUATION

Dataset↓ / Method→	MP [42]	PTA [3]	CSF1 [18]	CSF2 [20]	KSTA [19]	PND [36]	CNS [37]	BMM [13]	R-BMM [28]	Ours
Drink	0.0443	0.0250	0.0223	0.0223	0.0156	**0.0037**	0.0431	0.0266	0.0119	0.0071 ($K = 12$)
Pickup	0.0667	0.2369	0.2301	0.2277	0.2322	0.0372	0.1281	0.1731	0.0198	**0.0152** ($K = 12$)
Yoga	0.2331	0.1624	0.1467	0.1464	0.1476	0.0140	0.1845	0.1150	0.0129	**0.0122** ($K = 10$)
Stretch	0.2585	0.1088	0.0710	0.0685	0.0674	0.0156	0.0939	0.1034	0.0144	**0.0124** ($K = 11$)
Dance	0.2639	0.2960	0.2705	0.1983	0.2504	0.1454	**0.0759**	0.1864	0.1491	0.1209 ($K = 4$)
Face	0.0357	0.0436	0.0363	0.0314	0.0339	0.0165	0.0248	0.0303	0.0179	**0.0145** ($K = 7$)
Walking	0.5607	0.3951	0.1863	0.1035	0.1029	0.0465	**0.0396**	0.1298	0.0882	0.0816 ($K = 8$)
Shark	0.1571	0.1804	**0.0081**	0.0444	0.0160	0.0135	0.0832	0.2311	0.0551	0.0551 ($K = 3$)

(a) e_{3D} Performance Comparison on Motion Capture Dataset [3].

Method Type →	Sparse NRS*f*M Methods						Dense NRS*f*M Methods					
Dataset	MP [42]	PTA [3]	CSF1 [18]	CSF2 [20]	BMM [13]	Ours	DV [15]	SMSR [5]	CMDR [17]	GM [29]	ND [47]	Ours
Face Seq.1	0.0926	0.1559	0.5325	0.4677	0.4263	**0.0624**	0.0531	0.1893	-	**0.0443**	-	0.0624
Face Seq.2	0.0819	0.1503	0.9266	0.7909	0.6062	**0.0451**	0.0457	0.2133	-	**0.0381**	-	0.0451
Face Seq.3	0.1057	0.1252	0.5274	0.5474	0.0784	**0.0279**	0.0346	0.1345	0.0373	0.0294	0.0450	**0.0279**
Face Seq.4	0.0717	0.1348	0.5392	0.5292	0.0918	**0.0419**	0.0379	0.0984	0.0369	**0.0309**	0.0490	0.0419

(b) e_{3D} Performance Comparison on Garg et al. dense face sequence.

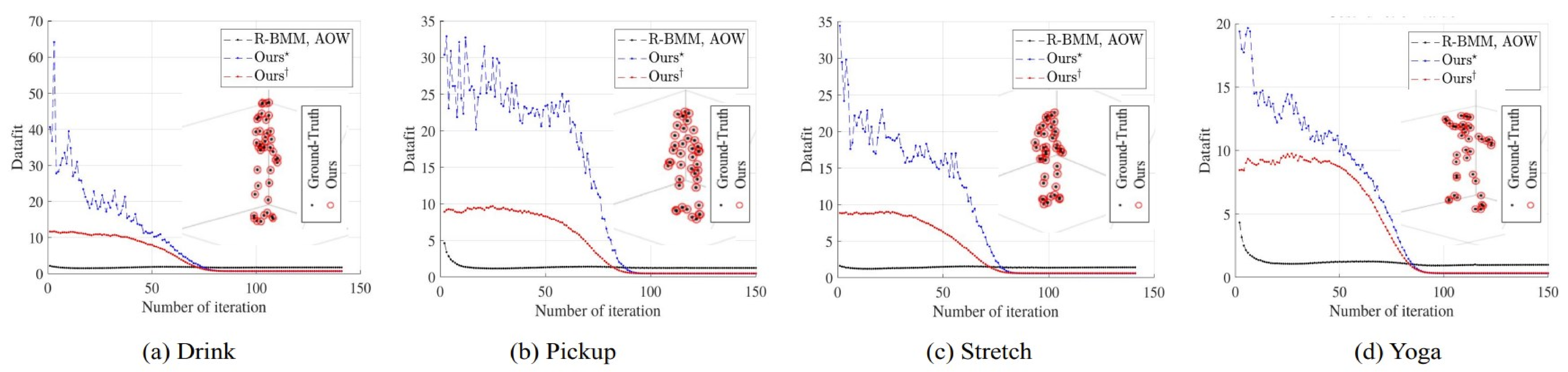
Data	BMM [13]	R-BMM [28]	AOW [24]	BP [41]	Ours
Articul.	18.49	16.00	15.03	16.10	**12.18** ($K = 8$)
Balloon	10.39	7.84	8.05	8.29	**6.29** ($K = 5$)
Paper	8.94	10.69	10.45	**6.70**	8.86 ($K = 2$)
Stretch	10.02	7.53	9.01	7.66	**6.36** ($K = 6$)
Tearing	14.23	16.34	16.20	11.26	**10.91** ($K = 6$)

(c) Robust RMSE on NRS*f*M Challenge Dataset (Jensen et al. [5])

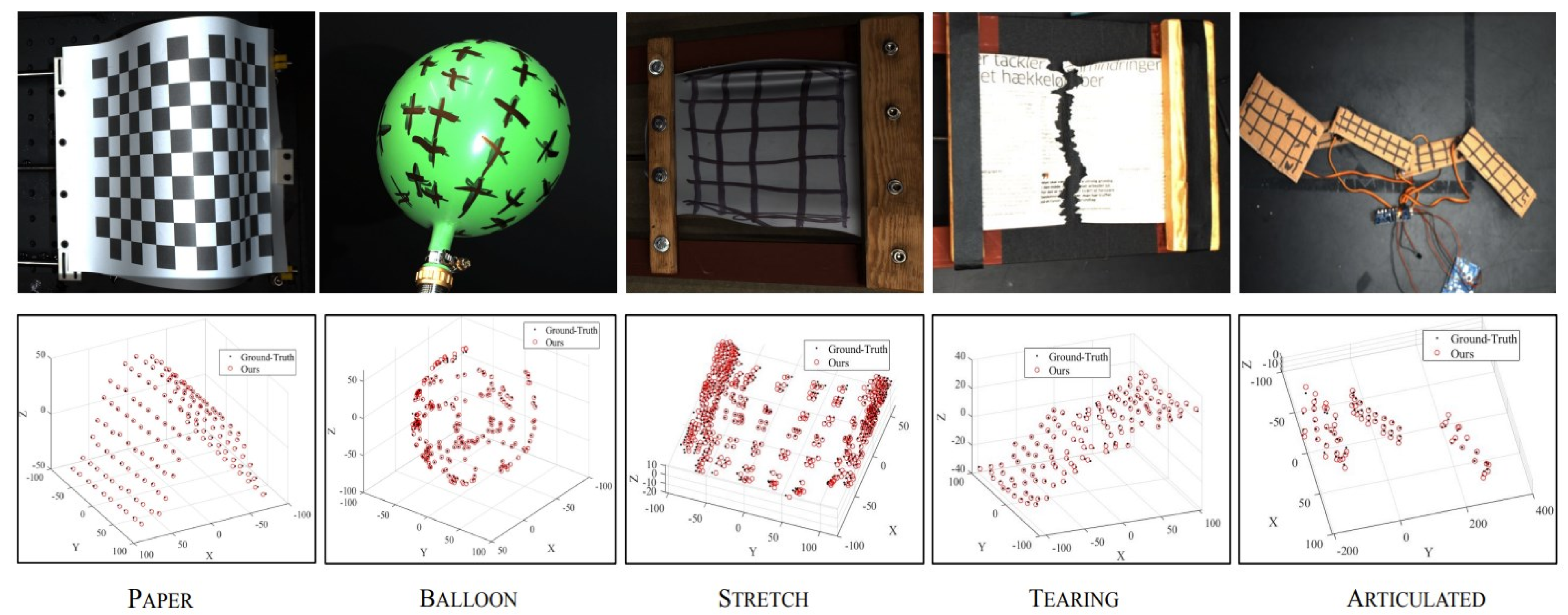
$$e_{3D} = \frac{1}{F}\sum_{i=1}^{F}\frac{\|\mathbf{X}_i^{est} - \mathbf{X}_i^{gt}\|_{\mathcal{F}}}{\|\mathbf{X}_i^{gt}\|_{\mathcal{F}}} \qquad \text{Robust RMSE: } m(\mathbf{X}^{gt}, \mathbf{X}^{est}) = \sqrt{\frac{1}{FP}\sum_{i,j}^{F,P} t(\mathbf{X}_{i,j}^{gt}, \mathbf{X}_{i,j}^{est})}$$

*Note: Reference number present in the above tables can be traced from the paper's main draft references.

QUALITATIVE RESULTS



(a) Qualitative Results on Motion Capture Dataset [3].



(b) Qualitative Results on NRS*f*M Challenge Dataset (Jensen et al. [5])