Non-Rigid Structure from Motion: Prior-Free Factorization Method Revisited Supplementary Material

Suryansh Kumar
Computer Vision Laboratory, ETH Zürich, Switzerland
sukumar@vision.ee.ethz.ch

Abstract

In this supplementary material, we first provide mathematical derivation to the sub-problems proposed in the paper [9]. For reference, we provide few qualitative comparison of our method in comparison to Dai et.al. approach [4]. Additional experimental results on real and synthetic dense dataset using our algorithm are also supplied. Lastly, we provide some general discussions on our algorithm.

1. Mathematical Derivations

The augmented form of the optimization is as follows:

$$\mathcal{L}_\rho(\mathtt{S}^\sharp, \mathtt{S}) = \mu\|\mathtt{S}^\sharp\|_{\Theta,*} + \frac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_{\mathtt{F}}^2 + \frac{\rho}{2}\|\mathtt{S}^\sharp - g(\mathtt{S})\|_{\mathtt{F}}^2 + < \mathtt{Y}, \mathtt{S}^\sharp - g(\mathtt{S}) > \tag{1}$$

(a) Solution to $\mathtt{S}$: Minimization the Eq:(1) w.r.t 'S' gives the following form

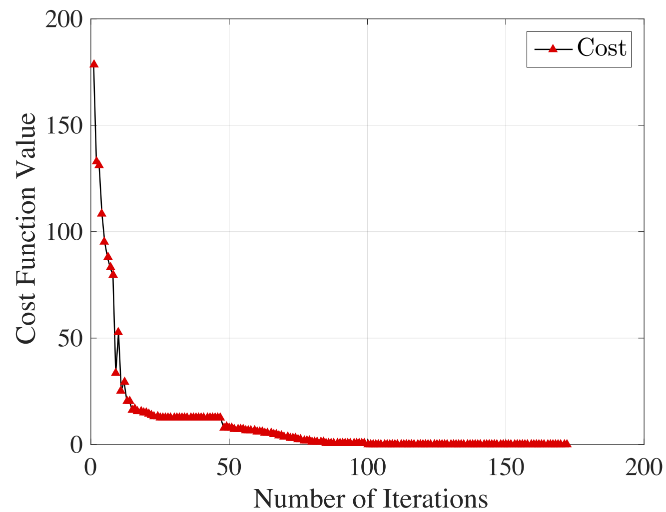
$$\underset{\mathtt{S}}{\text{argmin}}\ \mathcal{L}_\rho(\mathtt{S}) = \frac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_{\mathtt{F}}^2 + \frac{\rho}{2}\|g^{-1}(\mathtt{S}^\sharp) - \mathtt{S}\|_{\mathtt{F}}^2 + < g^{-1}(\mathtt{Y}), g^{-1}(\mathtt{S}^\sharp) - \mathtt{S} >$$
$$\equiv \underset{\mathtt{S}}{\text{argmin}}\frac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_{\mathtt{F}}^2 + \frac{\rho}{2}\Big\|\mathtt{S} - \Big(g^{-1}(\mathtt{S}^\sharp) + \frac{g^{-1}(\mathtt{Y})}{\rho}\Big)\Big\|_{\mathtt{F}}^2 \tag{2}$$

Taking the derivative of Eq:(2) w.r.t $\mathtt{S}$ and equating it to 0 gives

$$(\rho\mathtt{I} + \mathtt{R}^{\mathtt{T}}\mathtt{R})\mathtt{S} = \rho\Big(g^{-1}(\mathtt{S}^\sharp) + \frac{g^{-1}(\mathtt{Y})}{\rho}\Big) + \mathtt{R}^{\mathtt{T}}\mathtt{W} \tag{3}$$

(b) Solution to $\mathtt{S}^\sharp$: Minimization the Eq:(1) w.r.t '$\mathtt{S}^\sharp$' gives the following form:

$$\equiv \underset{\mathtt{S}^\sharp}{\text{argmin}}\ \mu\|\mathtt{S}^\sharp\|_{\Theta,*} + \frac{\rho}{2}\|\mathtt{S}^\sharp - g(\mathtt{S})\|_{\mathtt{F}}^2 + < \mathtt{Y}, \mathtt{S}^\sharp - g(\mathtt{S}) >$$
$$\equiv \underset{\mathtt{S}^\sharp}{\text{argmin}}\ \mu\|\mathtt{S}^\sharp\|_{\Theta,*} + \frac{\rho}{2}\Big\|\mathtt{S}^\sharp - \Big(g(\mathtt{S}) - \frac{\mathtt{Y}}{\rho}\Big)\Big\|_{\mathtt{F}}^2 \tag{4}$$



(a)

Figure 1: Convergence Curve

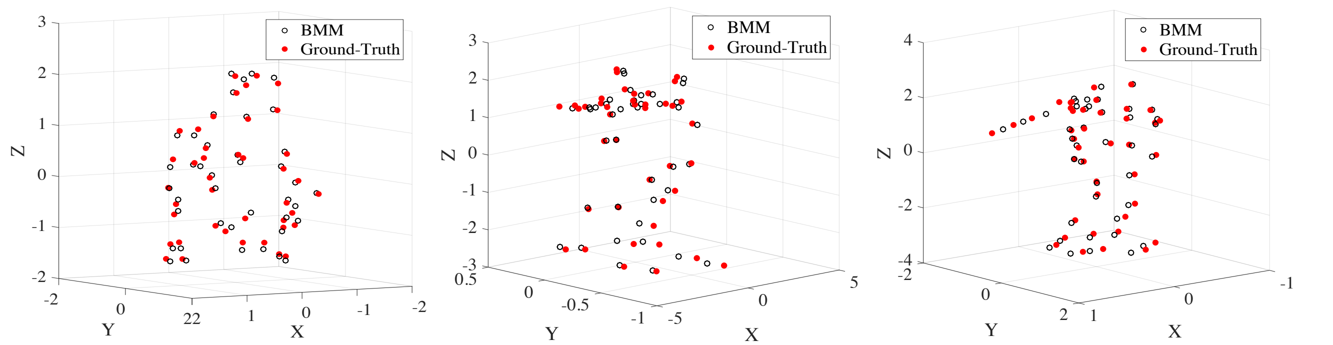
The Eq:(4) is solved by using the thresholding operator $\mathcal{S}_\tau(\sigma) = \text{sign}(\sigma).\text{max}(|\sigma| - \tau, 0)$. Let $[\mathtt{U}, \Sigma, \mathtt{V}]$ be the singular value decomposition of $(g(\mathtt{S}) - \frac{\mathtt{Y}}{\rho})$ then the solution to $\mathtt{S}^\sharp$ is given by $\mathtt{S}^\sharp = \mathtt{U}\mathcal{S}_{\frac{\Theta\mu}{\rho}}(\Sigma)\mathtt{V}$, with Θ as the weight assigned to singular values.

2. Convergence Curve

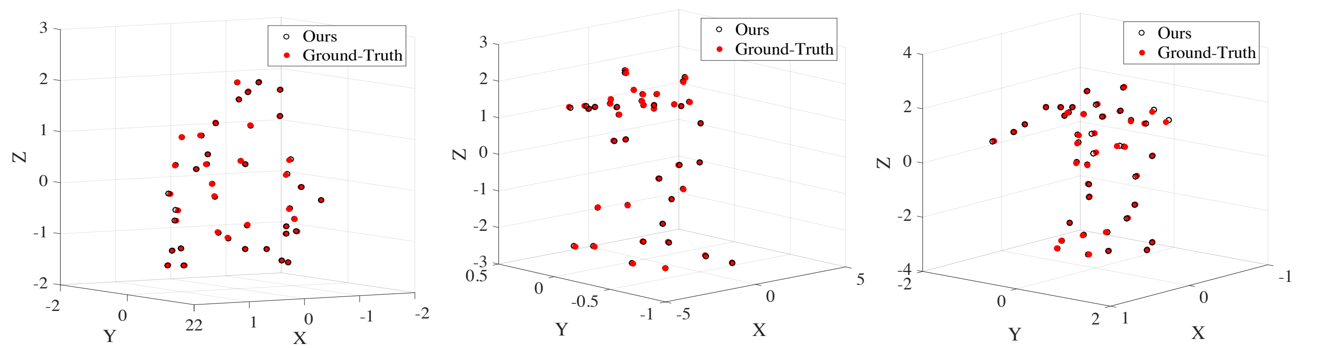
Figure 1(a) show the convergence curve of our proposed optimization for solving non-rigid shape matrix.

3. Qualitative Comparison

At last, we provide the visual comparison of our algorithm in comparison to the targeted baseline [4] in Figure 2. The results clearly shows that by simple yet powerful rectification to simple prior free idea, we can achieve a sig-



(a) Pickup (BMM) (b) Yoga (BMM) (c) Stretch (BMM)



(d) Pickup (Ours) (e) Yoga (Ours) (f) Stretch (Ours)

Figure 2: Qualitative comparison of our algorithm with the classical baseline BMM [4] under the same model complexity value (K). The first row and the second row shows the 3D reconstruction using Dai *et al*. and our approach respectively on the benchmark dataset.(Best viewed in color)

nificant boost in the reconstruction quality[1].

4. More Experimental Results

4.1. Results on Dense Datasets

In contrast to Dai *et al*. [4], we also performed experimental analysis on Dense dataset [7]. Table (1) show the performance comparison of our algorithm on synthetic face sequence. The proposed algorithm performs reasonably well even on the dense dataset.

Data	DS [3]	DV [6]	PTA [1]	MP [17]	Ours
Seq.1	0.0636	0.0531	0.1559	0.2572	0.0591
Seq.2	0.0569	0.0457	0.1503	0.0640	0.0478
Seq.3	0.0374	0.0346	0.1252	0.0611	0.0281
Seq.4	0.0428	0.0379	0.1348	0.0762	0.0308

Table 1: Average 3D reconstruction error (e_{3D}) comparison on dense synthetic face sequence[6]. Note: The code for DV [6] is not publicly available, we tabulated its results from DS [3] work. BMM [4] evaluation on this dataset is not available.

[1]Our claims are easy to verify and test using Dai *et al*. [4] publicly available code at http://users.cecs.anu.edu.au/ yuchao/publication.htm

4.2. Results on Missing Datasets

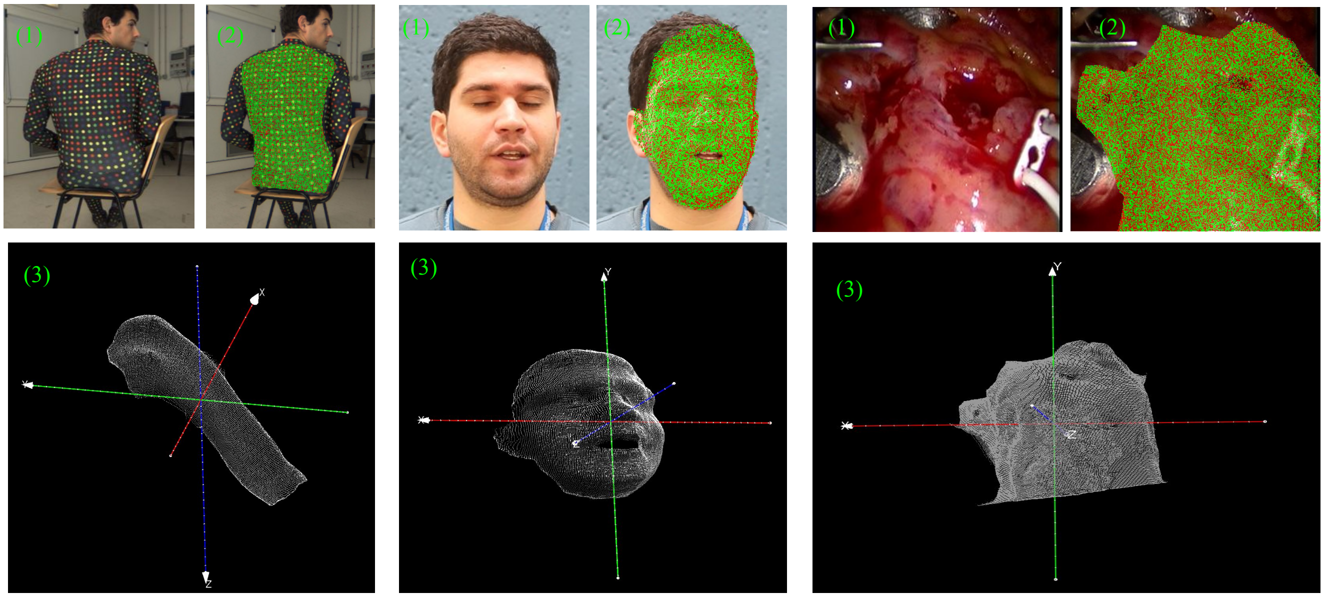
For more rigorous test on the missing dataset, we used Garg *et al*. real dense dataset sequence [6]. This dataset comprises of Face, Back and Heart sequence with 28332, 20561, and 68295 feature points tracked over 120, 150, and 80 images. Figure (3) show the qualitative results on the missing data for the available categories. The percentage of missing trajectories used for the experiments for Back, Face and Heart sequence are 29.87%, 41.17% and 43.93% respectively.

4.3. Timing details of the method

Run time of our algorithm for a typical sparse setting say 50 points, 300 frames is 39.46s in comparison to BMM [4] which is 34.24s.

5. Ablation Test

An ablation test is performed to show the contribution of smooth motion assumption and weighted nuclear norm minimization to 3D reconstruction accuracy. Table (2) provides the statistics, which clearly show the contribution of



(a) Back Sequence (29.87 %) (b) Face Sequence (41.17%) (c) Heart Sequence (43.93%)

Figure 3: Qualitative results on Garg *et al*. [6] real dense sequence. (1) Input image sequence (2) The green and red dot show the complete and missing trajectory respectively. (3) The qualitative results on the Back, Face and Heart sequence with 29.87%, 41.17% and 43.93% missing data sequence respectively.

our approach to improve the prior-free approach.

Data	D.R. + NN	D.R. + WNN	O.R + NN	O.R + WNN
Drink	0.0266	0.0119	0.0266	0.0119
Pickup	0.1731	0.0622	0.1517	0.0198
Yoga	0.1150	0.0129	0.1150	0.0129
Stretch	0.1034	0.0547	0.0910	0.0144

Table 2: Ablation study to show the contribution of both the step. D.R stands for Dai *et al*. rotation [4], O.R stands for Our rotation. NN and WNN refers to Nuclear and Weighted Nuclear Norm based optimization respectively to estimate shape.

6. Discussion

Note: The term «regularity» in the section(2) paragraph "plausible rectification" to the solution of rotation, in the main paper, is used in a loose sense. Kindly, ignore this if it's not mathematically precise to use it to convey the intuition.

Q. Why the assumption of «smooth» deformation of an object over frames is reasonable in solving NRSfM?
In many real world scenario's the transition of a non-rigidly moving object from one state to another over frames is not arbitrary but is well ordered or regular in terms of rigidity. Such assumption successfully captures the general notion about the global behavior of a deforming surface, at the same time maintains the local attribute of the surface. Therefore, to assume smooth motion is a reasonable choice and works well for most non-rigidly moving object [18].
Q. In some applications, we have more prior knowledge about the shape in addition to its low-rank matrix assumption, for example: exact rank of the clean shape matrix. In such cases, one may choose to minimize partial sum minimization of singular values optimization *i.e.*,

$$\underset{\mathtt{S}^\sharp,\mathtt{S}}{\mathtt{minimize}}\ \mu|\mathtt{rank}(\mathtt{S}^\sharp) - \mathtt{T}| + \frac{1}{2}\|\mathtt{W} - \mathtt{RS}\|_{\mathtt{F}}^2 \qquad (5)$$

where, $\mathtt{T}$ is the target rank of the shape matrix. However, such an optimization needs an introduction to new operator known as PSVT [16] to optimize the problem. Nevertheless, PSVT can be regarded as special case of solving the weighted nuclear norm minimization [2, 5]. Therefore, the point is, depending on the application, the proposed approach can be modified or changed, hence, its flexible.

References

[1] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Nonrigid structure from motion in trajectory space. In *Advances in neural information processing systems*, pages 41–48, 2009.

[2] K. Chen, H. Dong, and K.-S. Chan. Reduced rank regression via adaptive nuclear norm penalization. *Biometrika*, 100(4):901–920, 2013.

[3] Y. Dai, H. Deng, and M. He. Dense non-rigid structure-from-motion made easy-a spatial-temporal smoothness based solution. *arXiv preprint arXiv:1706.08629*, 2017.

[4] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. *International Journal of Computer Vision*, 107(2):101–122, 2014.

[5] S. Gaïffas and G. Lecué. Weighted algorithms for compressed sensing and matrix completion. *arXiv preprint arXiv:1107.1638*, 2011.

[6] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1272–1279, 2013.

[7] R. Garg, A. Roussos, and L. Agapito. A variational approach to video registration with subspace constraints. *International journal of computer vision*, 104(3):286–314, 2013.

[8] S. Kumar. Jumping manifolds: Geometry aware dense non-rigid structure from motion. In *IEEE, CVPR*, pages 5346–5355, 2019.

[9] S. Kumar. Non-rigid structure from motion: Prior-free factorization method revisited. *arXiv preprint arXiv:1902.10274*, 2019.

[10] S. Kumar, A. Cherian, Y. Dai, and H. Li. Scalable dense non-rigid structure-from-motion: A grassmannian perspective. In *IEEE, CVPR*, pages 254–263, 2018.

[11] S. Kumar, Y. Dai, and H. Li. Multi-body non-rigid structure-from-motion. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 148–156. IEEE, 2016.

[12] S. Kumar, Y. Dai, and H. Li. Monocular dense 3d reconstruction of a complex dynamic scene from two perspective frames. In *IEEE, ICCV*, pages 4649–4657, Oct 2017.

[13] S. Kumar, Y. Dai, and H. Li. Spatio-temporal union of subspaces for multi-body non-rigid structure-from-motion. *Pattern Recognition*, 71:428–443, May 2017.

[14] S. Kumar, Y. Dai, and H. Li. Superpixel soup: Monocular dense 3d reconstruction of a complex dynamic scene. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 2019.

[15] S. Kumar, R. S. Ghorakavi, Y. Dai, and H. Li. Dense depth estimation of a complex dynamic scene without explicit 3d motion estimation. *arXiv preprint arXiv:1902.03791*, 2019.

[16] T.-H. Oh, Y.-W. Tai, J.-C. Bazin, H. Kim, and I. S. Kweon. Partial sum minimization of singular values in robust pca: Algorithm and applications. *IEEE transactions on pattern analysis and machine intelligence*, 38(4):744–758, 2016.

[17] M. Paladini, A. Del Bue, J. Xavier, L. Agapito, M. Stovsic, and M. Dodig. Optimal metric projections for deformable and articulated structure-from-motion. *International journal of computer vision*, 96(2):252–276, 2012.

[18] V. Rabaud and S. Belongie. Re-thinking non-rigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.