Scalable Dense Non-rigid Structure-from-Motion: A Grassmannian Perspective Supplementary Material

Suryansh Kumar[1] Anoop Cherian[1,2,4] Yuchao Dai[1,3] Hongdong Li[1,2]
[1]Australian National University, [2]ACRV Australia, [3]NPU China, [4]MERL Cambridge MA
firstname.lastname@anu.edu.au

Abstract

In this supplementary material, we provide a detailed mathematical derivation to the optimization proposed in [5]. Additionally, we provide additional qualitative results, insights and briefly discussed some practical issues.

1. Detailed Mathematical Derivation

The cost function from Section 4.

$$\underset{S_s,S_t^\sharp,C_s,C_t,J_s,J_t}{\text{minimize}} \ \mathbf{E} = \frac{1}{2}\|W_s - RS_s\|_F^2 + \frac{\beta}{2}\|S_t^\sharp - \mathscr{T}_1(S_s)\|_F^2 + < Y_1, S_t^\sharp - \mathscr{T}_1(S_s) > + \gamma\|S_t^\sharp\|_* + \lambda_1\|\mathcal{T}_\mathbf{s} - \mathcal{T}_\mathbf{s}C_s\|_F^2 + \lambda_3\|J_s\|_* +$$
$$\frac{\beta}{2}\|C_s - J_s\|_F^2 + < Y_2, C_s - J_s > + \lambda_2\|\mathcal{T}_\mathbf{t} - \mathcal{T}_\mathbf{t}C_t\|_F^2 + \lambda_4\|J_t\|_* + \frac{\beta}{2}\|C_t - J_t\|_F^2 + < Y_3, C_t - J_t >,$$
$$\textit{subject to:}$$
$$\mathbf{\Psi_s} = \xi(C_s, S_s, \sigma_\epsilon); \mathbf{\Psi_t} = \xi(C_t, S_t^\sharp, \sigma_\epsilon);$$
$$S_s = \zeta(\mathbf{\Psi_s}, \Sigma_s, V_s, N_s); S_t^\sharp = \zeta(\mathbf{\Psi_t}, \Sigma_t^\sharp, V_t, N_t);$$
$$W_s = \mathscr{T}_2(W_s, S_s). \tag{1}$$

1.1. Background

To make the optimization simpler, let's consider an error term that involves the tensor structure

$$\|E_s\|_F^2 = \|\mathcal{T}_\mathbf{s} - \mathcal{T}_\mathbf{s}C_s\|_F^2. \tag{2}$$

Considering the i^{th} term, and using $\|E_{si}\|_F^2 = trace(E_{si}^T E_{si})$
From our notation definition $\mathcal{T}_\mathbf{s} = \left\{(\Psi_{s1})(\Psi_{s1})^T, (\Psi_{s2})(\Psi_{s2})^T ..., (\Psi_{sK_s})(\Psi_{sK_s})^T\right\}$ and $C_s \in \mathbb{R}^{K_s \times K_s}$

$$\|E_{si}\|_F^2 = trace\Big[\Big((\Psi_{si}\Psi_{si}^T) - \sum_{j=1}^{K_s} c_{ij}(\Psi_{sj}\Psi_{sj}^T)\Big)^T \Big((\Psi_{si}\Psi_{si}^T) - \sum_{j=1}^{K_s} c_{ij}(\Psi_{sj}\Psi_{sj}^T)\Big)\Big]$$
$$\|E_{si}\|_F^2 = trace\big((\Psi_{si}\Psi_{si}^T)^T(\Psi_{si}\Psi_{si}^T)\big) - 2\sum_{j=1}^{K_s} c_{ij} trace\big((\Psi_{si}\Psi_{si}^T)^T(\Psi_{sj}\Psi_{sj}^T)\big) + \sum_{l=1}^{K_s}\sum_{m=1}^{K_s} c_{il}c_{im} trace\big((\Psi_{sl}\Psi_{sl}^T)^T(\Psi_{sm}\Psi_{sm}^T)\big). \tag{3}$$

Now using the trace cyclic property and the orthonormality property of matrices.

$$\|E_{si}\|_F^2 = trace(I_d) - 2\sum_{j=1}^{K_s} c_{ij} trace\big((\Psi_{sj}^T \Psi_{si})(\Psi_{si}^T \Psi_{sj})\big) + \sum_{l=1}^{K_s}\sum_{m=1}^{K_s} c_{il}c_{im} trace\big((\Psi_{sl}^T \Psi_{sm})(\Psi_{sm}^T \Psi_{sl})\big).$$
$$\|E_{si}\|_F^2 = d - 2\sum_{j=1}^{K_s} c_{ij}\Omega_{ij}^s + \sum_{l=1}^{K_s}\sum_{m=1}^{K_s} c_{il}c_{im}\Omega_{lm}^s, \text{where, } \Omega_{ij}^s = trace\big((\Psi_{sj}^T \Psi_{si})(\Psi_{si}^T \Psi_{sj})\big). \tag{4}$$

Here, d stands for the dimension. Notice Ω_{ij}^s has a dimension of $d \times d$ which is easy to handle than the total number of points in a dense datasets. Also, it's simple to verify that Ω_{ij}^s is symmetric.

Using Equation (4) and $\mathbf{\Omega_s} = (\Omega_{ij}^s)_{i,j=1}^{K_s} \in \mathbb{R}^{K_s \times K_s}$, we can rewrite Equation (2) as follows

$$\begin{aligned}
&\|E_s\|_F^2 = const - 2trace(C_s\mathbf{\Omega_s}) + trace(C_s\mathbf{\Omega_s}C_s^T) \\
&\Rightarrow \|E_s\|_F^2 = const - 2trace(C_sL_sL_s^T) + trace((C_sL_s)(C_sL_s)^T), \text{ where } L_sL_s^T = \text{Cholesky}(\mathbf{\Omega_s}) \\
&\Rightarrow \|E_s\|_F^2 = const + \|L_s - C_sL_s\|_F^2 \{\because \text{constant w.r.t } C_s \text{ will not affect the minimization}\}
\end{aligned} \tag{5}$$

Similarly, other tensor structure can be equivalently represented in the temporal domain.

1.1.1 Overall Optimization

Substituting the above derivation in Equation (1) gives us a simpler representation

$$\begin{aligned}
&\underset{S_s,S_t^\sharp,C_s,C_t,J_s,J_t}{\text{minimize}}\ \mathbf{E} = \frac{1}{2}\|W_s - RS_s\|_F^2 + \frac{\beta}{2}\|S_t^\sharp - \mathscr{T}_1(S_s)\|_F^2 + < Y_1, S_t^\sharp - \mathscr{T}_1(S_s) > + \gamma\|S_t^\sharp\|_* + \lambda_1\|L_s - L_sC_s\|_F^2 + \lambda_3\|J_s\|_* + \\
&\frac{\beta}{2}\|C_s - J_s\|_F^2 + < Y_2, C_s - J_s > + \lambda_2\|L_t - L_tC_t\|_F^2 + \lambda_4\|J_t\|_* + \frac{\beta}{2}\|C_t - J_t\|_F^2 + < Y_3, C_t - J_t > \\
&\textit{subject to:} \\
&\mathbf{\Psi_s} = \xi(C_s, S_s, \sigma_\epsilon); \mathbf{\Psi_t} = \xi(C_t, S_t^\sharp, \sigma_\epsilon); \\
&S_s = \zeta(\mathbf{\Psi_s}, \Sigma_s, V_s, N_s); S_t^\sharp = \zeta(\mathbf{\Psi_t}, \Sigma_t^\sharp, V_t, N_t); \\
&W_s = \mathscr{T}_2(W_s, S_s);
\end{aligned} \tag{6}$$

Solution to S_s

$$\begin{aligned}
&\equiv \underset{S_s}{\text{argmin}} \frac{1}{2}\|W_s - RS_s\|_F^2 + \frac{\beta}{2}\|S_t^\sharp - \mathscr{T}_1(S_s)\|_F^2 + < Y_1, S_t^\sharp - \mathscr{T}_1(S_s) > \\
&\equiv \underset{S_s}{\text{argmin}} \frac{1}{2}\|W_s - RS_s\|_F^2 + \frac{\beta}{2}\|\mathscr{T}_1^{-1}(S_t^\sharp) - S_s\|_F^2 + < \mathscr{T}_1^{-1}(Y_1), \mathscr{T}_1^{-1}(S_t^\sharp) - S_s > \\
&\equiv \underset{S_s}{\text{argmin}} \frac{1}{2}\|W_s - RS_s\|_F^2 + \frac{\beta}{2}\|S_s - \big(\mathscr{T}_1^{-1}(S_t^\sharp) + \frac{\mathscr{T}_1^{-1}(Y_1)}{\beta}\big)\|_F^2.
\end{aligned} \tag{7}$$

The solution to S_s can be derived by differentiating the above term w.r.t S_s and equating it to zero.

$$S_s \equiv (R^TR + \beta I)^{-1}\Big(\beta\big(\mathscr{T}_1^{-1}(S_t^\sharp) + \frac{\mathscr{T}_1^{-1}(Y_1)}{\beta}\big) + R^TW_s\Big)$$

you may use $\mathbf{mldivide}()$ to solve this equation in MATLAB.

Solution to $S_t^\sharp$

$$\begin{aligned}
&\equiv \underset{S_t^\sharp}{\text{argmin}} \gamma\|S_t^\sharp\|_* + \frac{\beta}{2}\|S_t^\sharp - \mathscr{T}_1(S_s)\|_F^2 + < Y_1, S_t^\sharp - \mathscr{T}_1(S_s) > \\
&\equiv \underset{S_t^\sharp}{\text{argmin}} \gamma\|S_t^\sharp\|_* + \frac{\beta}{2}\|S_t^\sharp - \big(\mathscr{T}_1(S_s) - \frac{Y_1}{\beta}\big)\|_F^2
\end{aligned} \tag{8}$$

Let's define the soft-thresholding operation as $\mathcal{S}_\tau[x] = \text{sign}(x)\max(|x| - \tau, 0)$

Then, the optimal solution to $S_t^\sharp$ is given by

$$\begin{aligned} S_t^\sharp &\equiv U_t \mathcal{S}_{\frac{\gamma}{\beta}}(\Sigma_t) V_t, \\ &\text{where, } [U_t, \Sigma_t, V_t] = \text{svd}(\mathscr{T}_1(S_s) - \frac{Y_1}{\beta}) \end{aligned} \tag{9}$$

Solution to C_s

$$\begin{aligned} &\equiv \underset{C_s}{\text{argmin}}\ \lambda_1 \|L_s - L_s C_s\|_F^2 + \frac{\beta}{2}\|C_s - J_s\|_F^2 + < Y_2, C_s - J_s > \\ &\equiv \underset{C_s}{\text{argmin}}\ \lambda_1 \|L_s - L_s C_s\|_F^2 + \frac{\beta}{2}\|C_s - \big(J_s - \frac{Y_2}{\beta}\big)\|_F^2 \end{aligned} \tag{10}$$

The solution to C_s can be derived by differentiating the above term w.r.t C_s and equating it to zero.

$$C_s \equiv \big(2\lambda_1 L_s L_s^T + \beta I_s\big)^{-1}\Big(2\lambda_1 L_s L_s^T + \beta(J_s - \frac{Y_2}{\beta})\Big)$$

Solution to C_t

Similar to the C_s solution derivation, it's solution can be derived as follows:

$$\begin{aligned} &\equiv \underset{C_t}{\text{argmin}}\ \lambda_2 \|L_t - L_t C_t\|_F^2 + \frac{\beta}{2}\|C_t - J_t\|_F^2 + < Y_3, C_t - J_t > \\ &\equiv \underset{C_t}{\text{argmin}}\ \lambda_2 \|L_t - L_t C_t\|_F^2 + \frac{\beta}{2}\|C_t - \big(J_t - \frac{Y_3}{\beta}\big)\|_F^2 \end{aligned} \tag{11}$$

$$C_t \equiv \big(2\lambda_2 L_t L_t^T + \beta I_t\big)^{-1}\Big(2\lambda_2 L_t L_t^T + \beta(J_t - \frac{Y_3}{\beta})\Big)$$

Solution to J_s

$$\begin{aligned} &\equiv \underset{J_s}{\text{argmin}} \lambda_3 \|J_s\|_* + \frac{\beta}{2}\|C_s - J_s\|_F^2 + < Y_2, C_s - J_s > \\ &\equiv \underset{J_s}{\text{argmin}} \lambda_3 \|J_s\|_* + \frac{\beta}{2}\|J_s - \big(C_s + \frac{Y_2}{\beta}\big)\|_F^2 \end{aligned} \tag{12}$$

Similar to Equation 9 derivation, using the soft-thresholding operation, its optimal solution can be obtained as

$$J_s \equiv U_{J_s} \mathcal{S}_{\frac{\lambda_3}{\beta}}(\Sigma_{J_s}) V_{J_s},\ \text{where } [U_{J_s}, \Sigma_{J_s}, V_{J_s}] = \text{svd}(C_s + \frac{Y_2}{\beta}) \tag{13}$$

Solution to J_t

$$\begin{aligned} &\equiv \underset{J_t}{\text{argmin}} \lambda_4 \|J_t\|_* + \frac{\beta}{2}\|C_t - J_t\|_F^2 + < Y_3, C_t - J_t > \\ &\equiv \underset{J_t}{\text{argmin}} \lambda_4 \|J_t\|_* + \frac{\beta}{2}\|J_t - \big(C_t + \frac{Y_3}{\beta}\big)\|_F^2 \end{aligned} \tag{14}$$

$$J_t \equiv U_{J_t} \mathcal{S}_{\frac{\lambda_4}{\beta}}(\Sigma_{J_t}) V_{J_t},\ \text{where } [U_{J_t}, \Sigma_{J_t}, V_{J_t}] = \text{svd}(C_t + \frac{Y_3}{\beta}) \tag{15}$$

1.2. Proof

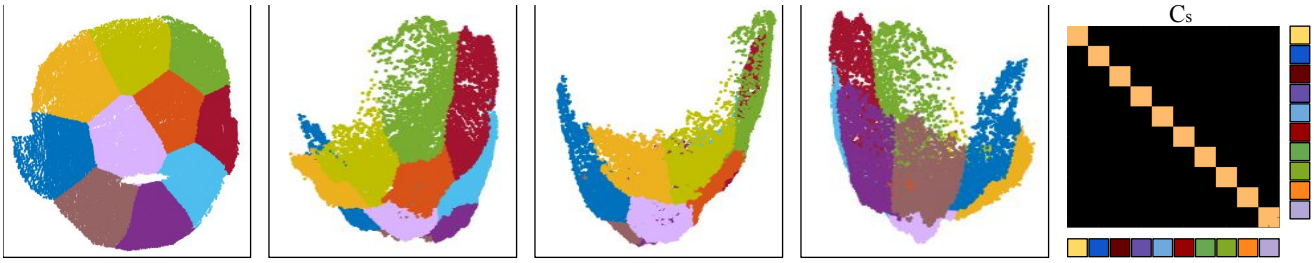
We have stated in the Algorithm table that $\mathbf{\Omega_s} \succeq 0$ The following lemma provides the proof for the same.

Lemma 1.1. *Given a set of orthonormal matrices* $\mathbf{\Psi_s} = \big\{\{\Psi_{si}\}_{i=1}^{K_s} : \forall \Psi_{si} \in \mathbb{R}^{d\times n},\ {\Psi_{si}}^T \Psi_{si} = I\big\}$*, if* $\exists\ \Omega_{ij}^s = trace[\big(\Psi_{sj}^T \Psi_{si}\big)\big(\Psi_{si}^T \Psi_{sj}\big)]$ *such that* $\mathbf{\Omega_s} = (\Omega_{ij}^s)_{i,j=1}^{K_s} \in \mathbb{R}^{K_s \times K_s}$*, then* $\mathbf{\Omega_s} \succeq 0$.

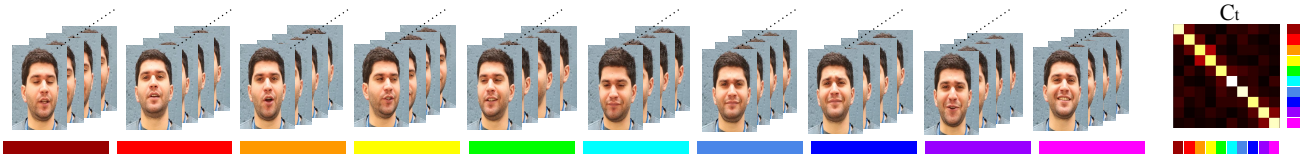
Proof. $Z_i = \Psi_{si}\Psi_{si}^T$ is a $d \times d$ symmetric matrix.

$$\begin{aligned} \text{As per the statement, } \Omega_{ij}^s &= trace[\big(\Psi_{sj}^T \Psi_{si}\big)\big(\Psi_{si}^T \Psi_{sj}\big)] = trace[\big(\Psi_{sj}\Psi_{sj}^T\big)\big(\Psi_{si}\Psi_{sj}^T\big)] \\ &= trace(Z_j Z_i) = trace(Z_j Z_i^T) = trace(Z_i^T Z_j) \\ \mathbf{\Omega_s} &= (\Omega_{ij}^s)_{i,j=1}^{K_s}, \text{then, } \mathbf{\Omega_s} = Z^T Z \quad \{\text{Skipping some elementary steps}\} \\ &\Rightarrow \mathbf{\Omega_s} \succeq 0. \end{aligned}$$

□



(a)



(b)

Figure 1: (a) Grouping of the trajectories based on C_s matrix. We provide four different views to check the fidelity of our result and assumption. (b) Grouping of the shapes based on C_t matrix. Color corresponding to the group block is shown with the color bars (extreme right). This simulation is done on the real face sequence [4] with K_s and K_t = 10.

Similarly, the positive semi-definite proof for $\mathbf{\Omega_t}$ can be derived. Note: In case $\mathbf{\Omega_s} = 0 || \mathbf{\Omega_t} = 0$ while implementing this algorithm, then add δ (a very small positive number) to the diagonal elements of $\mathbf{\Omega_s}$ or $\mathbf{\Omega_t}$ accordingly, to get to an approximate Cholesky factorization. Mathematically, approximate $\mathbf{\Omega_s} = 0 || \mathbf{\Omega_t} = 0$ as $\mathbf{\Omega} \approx \mathbf{\Omega} + \delta \mathbf{I}$ to make it numerically positive definite.

2. Qualitative Results

2.1. Analysis of C_s and C_t

In the experiment section we mentioned about the observation of C_s and C_t matrix. Since, no ground-truth data's are available to quantify these matrices, we provide a visual observation for the same. We used the spectral clustering [10] to group the trajectories and shapes after convergence to infer the output of C_s and C_t matrix. Figure 1 shows the output of this experiment. Visually it can be observed that local low-rank linear subspace are properly procured —both spatially and temporally.

3. Rotation Estimate

We used the method proposed by Dai *et al*. [3] to estimate rotation which only depends on the K value (model complexity) and therefore, it can efficiently handle dense feature correspondence over multiple frame to estimate rotation. Assuming that a single non-rigid deforming object constitutes a global relative camera pose over frames is a reasonable choice and works efficiently. Most of the past approaches also used this assumption to solve rotation [3, 2, 1, 6, 7]. Quantitative results on several datasets also shows that high-quality reconstruction can be obtained under such assumption. *Additionally, it has also been observed that different camera path can lead to different reconstruction results. Consequently, we plan to investigate it in our future work.*

Note: For technical details on the compactness of grassmannians, kindly refer to [9] for comprehensive theory. Nevertheless, there are many other books and notes on differential manifolds which provides information on the compactness of grassmannians.

4. Convergence Curve

Figure 2 shows the empirical convergence of our algorithm implementation. In practice, we find that our algorithm converges in 100-150 iterations on almost all the datasets, we tested so far. The theoretical convergence proof is left as a

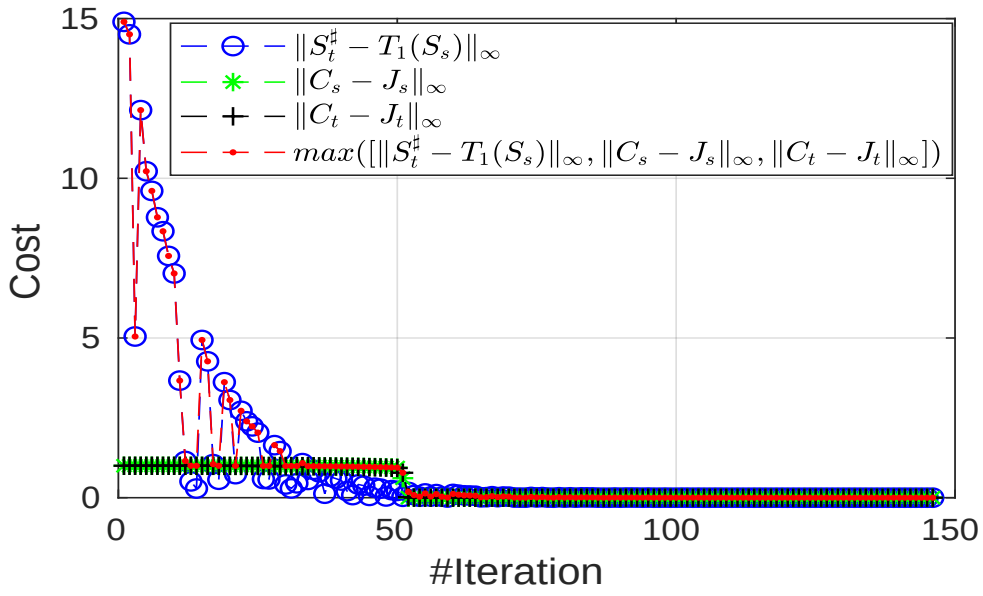


Figure 2: Empirical convergence curve of our algorithm using ADMM.

future work.

5. Practical Challenges for dense NRSfM systems

From a practical view point, estimating robust dense feature correspondences of a deforming surface/object across frames in itself is a very challenging problem to solve. Such challenges comes from the fact that illumination of the deforming object keeps changing over time. As a result, color based weighting term to regularize neighboring terms (like depth continuity, flow continuity, motion continuity etc.) in optical flow and 3D reconstruction algorithm's [8] may lead to wrong solution. Nonetheless, for our method we assume that a fairly good dense feature correspondences over images are provided as input. At last we would like to state that while developing dense NRSfM system using our or any other similar approaches aforementioned challenges needs to be addressed first.

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. 4

[2] 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

[3] 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. 4

[4] 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. 4

[5] S. Kumar, A. Cherian, Y. Dai, and H. Li. Scalable dense non-rigid structure-from-motion: A grassmannian perspective. *arXiv preprint arXiv:1803.00233*, 2018. 1

[6] 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. 4

[7] 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. 4

[8] S. Kumar, Y. Dai, and H. Li. Monocular dense 3d reconstruction of a complex dynamic scene from two perspective frames. In *IEEE International Conference on Computer Vision*, pages 4649–4657, Oct 2017. 5

[9] J. Milnor and J. D. Stasheff. *Characteristic Classes.(AM-76)*, volume 76. Princeton university press, 2016. 4

[10] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002. 4