Organic Priors in Non-Rigid Structure from Motion

Suryansh Kumar[1] sukumar@vision.ee.ethz.ch
Luc Van Gool[1,2] vangool@vision.ee.ethz.ch

[1]ETH Zürich Switzerland, [2]KU Leuven Belgium

Abstract. This paper advocates the use of organic priors in classical non-rigid structure from motion (NRSfM). By organic priors, we mean invaluable intermediate prior information intrinsic to the NRSfM matrix factorization theory. It is shown that such priors reside in the factorized matrices, and quite surprisingly, existing methods generally disregard them. The paper's main contribution is to put forward a simple, methodical, and practical method that can effectively exploit such organic priors to solve NRSfM. The proposed method does not make assumptions other than the popular one on the low-rank shape and offers a reliable solution to NRSfM under orthographic projection. Our work reveals that the accessibility of organic priors is independent of the camera motion and shape deformation type. Besides that, the paper provides insights into the NRSfM factorization—both in terms of shape and motion—and is the first approach to show the benefit of single rotation averaging for NRSfM. Furthermore, we outline how to effectively recover motion and non-rigid 3D shape using the proposed organic prior based approach and demonstrate results that outperform prior-free NRSfM performance by a significant margin. Finally, we present the benefits of our method via extensive experiments and evaluations on several benchmark datasets.

Keywords: Organic Priors, Non-Rigid Structure from Motion, Rank Minimization, Rotation Averaging, Matrix Factorization.

1 Introduction

Non-rigid structure from motion (NRSfM) factorization is a classical problem in geometric computer vision [9,27]. The problem's primary objective is to recover 3D shape of a deforming object from a given set of image key-points tracked across multiple images. As a result, it is sometimes referred as solving an inverse graphics problem [46]. An effective solution to NRSfM is of significant importance to many computer vision and geometry processing applications [10,39].

It is now widely accepted that the NRSfM problem is challenging to work out if the shape deforms arbitrarily across images, as it becomes equivalent to a non-rigid shape recovery problem using a single image at a time, which is ill-posed. Accordingly, several assumptions and priors are often used to make the problem solvable and computationally tractable. For instance, the deforming shape spans a low-rank space [13], smooth temporal shape deformation [1,6],

shape or trajectory lies in the union of linear subspace [26,29–31,35,54] and the local surface deformation is rigid or near rigid [14,37,45,48]. Other favored prior assumptions include smooth camera motion [28,43], a piece-wise planar deformation model [14,32–34], a Gaussian shape prior distribution [51], the availability of a 3D shape template [46], and shapes across frames must align [36]. Despite that, NRSfM remains a challenging and active research problem.

Meanwhile, there exist several popular methods to solve NRSfM [3, 9, 13, 36, 51]. Here, we will concern ourselves with the theory of matrix factorization for NRSfM elaborated in 1999-2000 by Bregler *et al.* [9][1]. It is a simple yet an effective approach to solve NRSfM. In the context of matrix factorization, one of the commonly used prior assumptions is that the non-rigid shape spans a low-rank space *i.e.*, the shape at each instance can be represented as a linear combination of a small set of basis shapes [9]. This paper adheres to such an assumption and shows that other important prior information resides within the *intermediate* factorized matrices —termed as organic priors. Surprisingly, most existing methods, if not all, ignore them. We used the word «organic» because they come naturally by properly conceiving the algebraic and geometric construction of NRSfM factorization [9,13,49]. Furthermore, this paper contends that the use of external priors and assumptions not only restricts the practical use of NRSfM methods, but also constrains the understanding and broader use of the well-known theory [13]. Yet, unlike [13], we advocate the use of organic priors, which is predicated on the proposition put forward by Kumar [28]. In this paper, we will show how to procure organic priors and exploit them effectively.

One of the critical innovations in NRSfM factorization that disputed the use of extraneous priors was introduced in [12,13]. The algorithm proposed in that paper does not use any prior other than the low-rank shape assumption. Nevertheless, despite its theoretical elegance and challenging argument, it fails to perform well on benchmark datasets [3,15,25,50]. Recently, Kumar [28] highlighted the possible reasons and exploited its missing pieces to gain performance. It was shown that a better rotation and shape could be estimated using the prior-free method's theory [12,13]. Still, [28] based his work on a smooth camera motion assumption that requires a brute force, heuristic search in the rotation space. In contrast, this paper puts forward a systematic method for NRSfM factorization that encourages the use of organic priors extracted from the factorized matrices. Experiments on synthetic and real benchmarks show that our approach consistently provides excellent 3D reconstruction results. This indicates the strength of matrix factorization theory for NRSfM. In summary, our contributions are

- A methodical approach for solving NRSfM that provides outstanding results using simple matrix factorization idea under the low-rank shape assumption.
- An approach that endorses the use of organic priors rather than extraneous priors or assumptions. Our method introduce single rotation averaging to estimate better rotation while being free from smooth camera motion heuristics.

[1] See, however, *C. Tomasi and T. Kanade*, **pp. 137-154**, IJCV (1992) for the original matrix factorization theory for shape and motion estimation, although devoted to the rigid SfM problem [49].

- 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 organic shape priors and yields shape reconstructions superior to other popular NRSfM factorization methods [13,24,28].

Further, we proffer the benefits of L_1 single rotation averaging for NRSfM factorization, which is excellent at providing robust rotation solutions. Although most of NRSfM factorization focuses on sparse key-point sets, our method is equally effective for dense feature points and compares favorably with well crafted state-of-the-art dense NRSfM methods such as [15,26,29].

2 Overview and Key Strategy

General Definition and Classical Setup. In NRSfM, a measurement matrix $\mathbf{W} \in \mathbb{R}^{2F\times P}$ is defined as a matrix containing the image coordinates ($\mathbf{w}_{f,p} \in \mathbb{R}^{2\times 1}$) of P feature points tracked across F image frames. $\mathbf{W}$ is generally mean-centered and given as an input to the factorization method [9]. Under an orthographic camera model assumption, the NRSfM factorization theory proposes to decompose the $\mathbf{W}$ into a product of a rotation matrix $\mathbf{R} \in \mathbb{R}^{2F\times 3F}$ and a non-rigid shape matrix $\mathbf{X} \in \mathbb{R}^{3F\times P}$ such that $\mathbf{W} \approx \mathbf{RX}$.

A practical method for NRSfM factorization was initially proposed by Bregler *et al.* [9]. Using the linear model proposition, a non-rigid shape $\mathbf{X}_i \in \mathbb{R}^{3\times P}$ for i^{th} frame was represented as a linear combination of K basis shapes $\mathbf{B}_k \in \mathbb{R}^{3\times P}$ *i.e.*, $\mathbf{X}_i = \sum_{k=1}^{K} c_{ik}\mathbf{B}_k$, where c_{ik} denotes the shape coefficients. Using such a shape representation, the $\mathbf{W}$ matrix is decomposed as follows:

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_{11}\dots\mathbf{w}_{1P} \\ \dots \\ \mathbf{w}_{F1}\dots\mathbf{w}_{FP} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1\mathbf{X}_1 \\ .. \\ \mathbf{R}_F\mathbf{X}_F \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{11}\mathbf{R}_1\dots\mathbf{c}_{1K}\mathbf{R}_1 \\ \dots \\ \mathbf{c}_{F1}\mathbf{R}_F\dots\mathbf{c}_{FK}\mathbf{R}_F \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ .. \\ \mathbf{B}_K \end{bmatrix} \tag{1}$$
$$\Rightarrow \mathbf{W} = \mathbf{R}(\mathbf{C}\otimes\mathbf{I}_3)\mathbf{B} = \mathbf{MB}$$

where, $\mathbf{R}_i \in \mathbb{R}^{2\times 3}$ denotes the i^{th} frame rotation matrix, and $\otimes$ the Kronecker product. $\mathbf{M} \in \mathbb{R}^{2F\times 3K}$, $\mathbf{B} \in \mathbb{R}^{3K\times P}$ and $\mathbf{I}_3$ is the 3×3 identity matrix. It is easy to infer from the above construction that $\texttt{rank}(\mathbf{W}) \leq 3K$.

Since there is no general way to solve for $\mathbf{R}, \mathbf{C}$, and $\mathbf{B}$ directly, rank $3K$ factorization of $\mathbf{W}$ via its Singular Value Decomposition (`svd`) gives a natural way to solve the problem under the orthonormality constraint of the rotation space [3, 9, 13]. As it is well-known that factorization of $\mathbf{W}$ via `svd` is not unique [9, 49], there must exist a corrective matrix $\mathbf{G} \in \mathbb{R}^{3K\times 3K}$ such that $\mathbf{W} = (\hat{\mathbf{M}}\mathbf{G})(\mathbf{G}^{-1}\hat{\mathbf{B}}) = \mathbf{MB}$. And therefore, once the `svd` of $\mathbf{W}$ is computed, a general rule of thumb in NRSfM factorization is to first solve for the $\mathbf{G}$ matrix, followed by the estimation of $\mathbf{R}$ and $\mathbf{X}$, respectively [3,13].

(a) Background on Corrective and Rotation Matrix Estimation. To solve for $\mathbf{G}$, orthonormality constraints are enforced [2,9]. Few works proposed in the past solve for the full $\mathbf{G} \in \mathbb{R}^{3K\times 3K}$ matrix (*i.e.*, for all its matrix entries)

to estimate the rotation matrix [2, 8, 52, 53]. In contrast, Dai *et al.* [13] argued that rather than solving for the full $\mathbf{G}$, simply solve for $\mathbf{G}^1 \in \mathbb{R}^{3K\times 3}$ (first 3 columns or first corrective triplet) leveraging Akhter *et al.*'s [2] theory. Yet, there exist K such triplets (see Fig.1). Even if we don't deviate from [2] theory, the question that still remains with the use of [13] rotation estimation theory is: *Q1. Do we utilize all possible rotations that can be recovered from [13] rotation estimation theory?* The answer is no, as recently argued by Kumar [28]. He proposed to inspect all K column triplets in the corrective matrix ($\mathbf{G}$), and recover K possible rotation solutions $\mathbf{R}^k \in \mathbb{R}^{2F\times 3}$, where $k \in \{1, 2, \ldots, K\}$. He then selected the one $\mathbf{R}^k$ that provides a smooth camera motion trajectory. Yet, this solution is heuristic in nature and requires a qualitative inspection of all the K rotations. The point to note is that, similar to Dai *et al.* [13], Kumar's [28] solution at the end does not fully utilize all the K rotation priors and eventually ends up aborting the rest of the near smooth or non-smooth $(K-1)$ solutions. We call those $(K-1)$ rotation solutions «`organic priors in the rotation space`». Our proposed method utilizes all those organic rotation priors to estimate a better and more informed rotation matrix.

(b) Background on the Shape Matrix Estimation. After solving for the rotation, the goal is to estimate the shape with the rank K constraint. Generally, an initial solution to the shape can be estimated in a closed form using $\mathbf{X}_{init} = \texttt{pinv}(\mathbf{R})\mathbf{W}$[2]. Yet, this may produce a planar solution as outlined in [52]. In spite of that, $\mathbf{X}_{init}$ provides useful information about the true shape and can be used as a shape variable ($\mathbf{X}$) initialization in the following rank-optimization problem:

$$\underset{\mathbf{X}^\sharp}{\text{minimize}} \frac{1}{2}\sum_{i=1}^{F}\sum_{j=1}^{P} \|\mathbf{w}_{ij} - \mathbf{R}_i\mathbf{x}_{ij}\|^2, \text{ s.t. } \text{rank}(\mathbf{X}^\sharp) \leq K \tag{2}$$

where, $\mathbf{x}_{ij}$ denote the 3D point j in the i^{th} view, and $\mathbf{w}_{ij}$ is its corresponding projection. $\mathbf{X}^\sharp \in \mathbb{R}^{F\times 3P}$ is the reshape of the shape matrix ($\mathbf{X}$) for K shape basis constraint [4, 13].

There exist several approaches to solve the Eq.(2) optimization [4, 13, 20, 42]. Among them, relaxed rank-minimization via appropriate matrix-norm minimization is widely used to recover a low-rank shape matrix providing favorable accuracy and robust results [13, 15, 24, 28, 52]. In this paper, we exploit the singular values of the $\mathbf{X}_{init}$, which we call «`organic priors in the shape space`» to recover better solution than the recent state-of-the-art [24, 41]. Although Kumar [28] work is the first to propose and utilize such priors for better shape reconstructions, in this paper we show that we can do better[3]. This brings us to the next question:

Q2. Can we make better use of the organic shape prior to solve for the shape matrix? We will show that we can. When solving the relaxed rank minimization optimization problem of the shape matrix [13], it is not beneficial to equally penalize all the singular values of the shape matrix. Hence, for effective shape

[2] `pinv()` symbolizes Moore–Penrose inverse of a matrix, also known as pseudoinverse.
[3] Familiarity with [13, 28] gives a good insight on our paper's novelty.

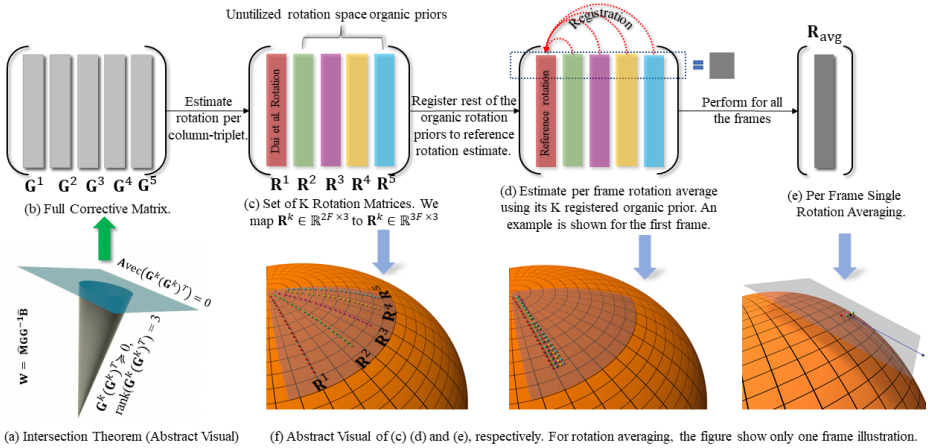


Fig. 1: A visual illustration of our rotation estimation approach for $K = 5$. (a)-(b) Use the [13] intersection theorem to recover all $\mathbf{G}^k$s'. (b)-(c) Recover $\mathbf{R}^k \in \mathbb{R}^{2F\times 3}$ corresponding to each column triplet. (c) Map per frame 2×3 rotation to $SO(3)$ via a cross product, taking care of the determinant sign. (d) Register $(K-1)$ rotation to the reference rotation (e) Perform per frame single rotation averaging to recover $\mathbf{R}_{\text{avg}}$.

recovery, one can use the $\mathbf{X}_{init}$ singular values prior to regularize the shape matrix rank-optimization [28]. In particular, perform a weighted nuclear norm (WNN) minimization of the shape matrix, and assign the weights to the shape variable that is inversely proportional to the $\mathbf{X}_{init}$ singular values magnitude [28]. In this paper, we go a step further. We propose to preserve the first component of the shape during its WNN minimization, *i.e.*, to not penalize the first singular value shape prior from $\mathbf{X}_{init}$. We empirically observed that the first singular value of $\mathbf{X}_{init}$, more often than not, does contain rich information about the true shape. Penalizing the first singular value during WNN minimization of the shape matrix may needlessly hurts the overall results. Consequently, we introduce a mix of partial sum minimization of singular values [40] and WNN minimization [28] to recover a better shape matrix estimate.

3 Proposed Approach

First, we provide details of our approach to solve for the rotation matrix, followed by the shape matrix estimation.

3.1 Rotation Estimation

To put our work in context, we highlight some previous efforts that took a similar direction towards enhancing the rotation estimate for NRSfM.

Relation to previous methods. As mentioned before, there exist K corrective column triplets in the $\mathbf{G}$ matrix (Fig.1(b)). Brand [8] and Akhter *et al.* [2] solves for all corrective triplets jointly. Xiao *et al.* [53] proposed to independently solve

for each corrective triplet ($\mathbf{G}^k$) and align $\mathbf{G}^k$'s using the Procrustes method up to sign [43]. Lee *et al.* [36] proposed an additional constraint on the rotation by posing NRSfM as a shape alignment problem. By comparison, the Dai *et al.* [13] method is a simple and effective way to compute rotation. It estimates only the 1^{st} column-triplet of $\mathbf{G}$ *i.e.*, $\mathbf{G}^1$ to recover $\mathbf{R}$ (see Fig.1).

On the contrary, we propose to first compute all the column-triplets *i.e.*, $\mathbf{G}^k \; \forall \; k \in \{1, 2, \ldots, K\}$ and their corresponding rotation matrices $\mathbf{R}^k \in \mathbb{R}^{2F\times 3}$, using Dai *et al.* [13]. Later, we use all K rotation estimates per frame to estimate a better rotation matrix $\mathbf{R} \in \mathbb{R}^{2F\times 3F}$ via the theory of single rotation averaging [23]. Rather than aligning $\mathbf{G}^k$s as in [2, 53], we register the rotations. Our approach consists of the following steps applied in sequel:

- Recover $\mathbf{G}^k$ and its corresponding $\mathbf{R}^k \in \mathbb{R}^{2F\times 3}$ using [13], $\forall \; k \in \{1, \ldots, K\}$.
- Map $\mathbf{R}^k \in \mathbb{R}^{2F\times 3} \mapsto \mathbf{R}^k \in \mathbb{R}^{3F\times 3}$ via the cross product of per frame 2×3 orthographic rotation estimates, while correcting for the sign of the determinant, if applicable.
- Take the rotation due to the first column-triplet *i.e.*, $\mathbf{R}^1 \in \mathbb{R}^{3F\times 3}$ as the reference rotation matrix and register the other $(K-1)$ *i.e.*, $\mathbf{R}^2$ to $\mathbf{R}^K$ rotation estimates to it. After registration, filter the rotation sample if the distance w.r.t its reference rotation sample is greater than δ [23] (see below for details).
- Perform per frame single rotation averaging of all the aligned rotation priors to recover $\mathbf{R}_{avg} \in \mathbb{R}^{3F\times 3}$. Later, convert $\mathbf{R}_{avg}$ per frame to orthographic form and place it in the block diagonal structure to construct $\mathbf{R} \in \mathbb{R}^{2F\times 3F}$.

Before performing single rotation averaging, we align all rotation priors due to the global ambiguity (see Fig.1(b)-Fig.1(c) visual). We align the other $(K-1)$ rotations to $\mathbf{R}^1$ using the following optimization.

$$\underset{R^k_{\text{reg}}}{\text{minimize}} \sum_{f=1}^{F} \|\mathbf{R}^1_f - \mathbf{R}^k_f (\mathbf{R}^k_{\text{reg}})^T\|^2_{\mathcal{F}}; \text{ s.t. } \mathbf{R}^k_{\text{reg}} \in SO(3), \; \forall \; k \in \{2, \ldots, K\} \tag{3}$$

where, $k \in \mathbb{Z}$. In the paper, $\|.\|_{\mathcal{F}}$ denotes the Frobenius norm. Using Eq.(3) optimization, we recover $(K-1)$ $\mathbf{R}^k_{\text{reg}} \in \mathbb{R}^{3\times 3}$ to register the organic rotation priors for averaging. Next, we perform single rotation averaging per frame.

Single rotation averaging. Given a set of $n \geq 1$ rotations $\{R_1, R_2, \ldots, R_n\} \subset SO(3)$, the goal of single rotation averaging is to find the average of a set of rotations [23]. It can also be conceived as finding a rotation sample on the $SO(3)$ manifold that minimizes the following cost function

$$\underset{R\in SO(3)}{\text{argmin}} \sum_{i=1}^{n} d^p(R_i, R) \tag{4}$$

$d()$ denotes a suitable metric function. We use $p = 1$ for its robustness and accuracy as compared to $p = 2$ [22]. For our problem, we have K rotation samples for each frame (see Fig.1(c)). Accordingly, we modify Eq.(4) as:

$$\underset{\mathbf{R}_f\in SO(3)}{\text{argmin}} \sum_{k=1}^{K} d^1(\tilde{\mathbf{R}}^k_f, \mathbf{R}_f) \tag{5}$$

Here, $\tilde{\mathbf{R}}_f^k$ is the k^{th} registered rotation for the f^{th} frame, *i.e.*, averaging across rows after registration (Fig.1(c)). We solve Eq.(5) for all the frames using the Weiszfeld algorithm for L_1 rotation averaging [22]. The average is computed in the local tangent space of $SO(3)$ centered at the current estimate and then back-projected onto $SO(3)$ using the exponential map (Fig.1(d)-Fig.1(e)). Yet, instead of initializing using the chordal L_2 mean, we use the starting point using the following equation proposed recently by Lee and Civera [38].

$$\mathbf{S}_o = \underset{\mathbf{S}\in\mathbb{R}^{3\times 3}}{\operatorname{argmin}} \sum_{i=1}^{K}\sum_{j=1}^{3}\sum_{k=1}^{3} \left|(\mathbf{R}_i - \mathbf{S})_{jk}\right|_1 \tag{6}$$

Eq.(6) is an element-wise L_1 norm matrix entries, minimizing the sum of absolute differences from the $\mathbf{R}_i$ at (j,k) value *i.e.*, $(\mathbf{S_o})_{jk} = \text{median}(\{\mathbf{R}_i\}_{i=1}^K)\ \forall\ j,k \in \{1,2,3\}$. After median computation, its equivalent rotation representation is obtained by projecting $\mathbf{S}_o$ onto $SO(3)$ using Υ operator. For $\Psi \in \mathbb{R}^{3\times 3}$ matrix, we define $\Upsilon_{SO(3)}(\Psi) = \mathbf{UDV}^T$, where $\mathbf{UDV}^T$ is `svd` of Ψ and $\mathbf{D} = \texttt{diag}(1,1,-1)$ if $\det(\mathbf{UV}^T) < 0$ or $\mathbf{I}_{3\times 3}$ otherwise.

Algorithm 1: L_1 Single Rotation Avg.

Data: Set of rotation $\{\mathbf{R}_i\}_{i=1}^K$, $\epsilon_t = 1e^{-3}$
Result: L_1 mean *i.e.*, median rotation
Set $\mathbf{S}_o := \text{median}(\{\mathbf{R}_i\}_{i=1}^K)$;
/*Project median on $SO(3)$*/
Set $\mathbf{R}_o := \Upsilon_{SO(3)}(\mathbf{S}_o)$; Set $\mathbf{R}_{\text{avg}} := \mathbf{R}_o$;
while do
  $\mathbf{v}_i := \log(\mathbf{R}_i\mathbf{R}_{\text{avg}}^T)\ \forall\ i = 1, 2.., K$;
  $\boldsymbol{\Delta}\mathbf{v} := \frac{\sum_{i=1}^K \mathbf{v}_i/\|\mathbf{v}_i\|}{\sum_{i=1}^K 1/\|\mathbf{v}_i\|}$;/*Weiszfeld step*/
  $\mathbf{R}_{\text{avg}} := \exp(\boldsymbol{\Delta}\mathbf{v})\mathbf{R}_{\text{avg}}$
  if $\|\boldsymbol{\Delta}\mathbf{v}\| < \epsilon_t$ **then**
    break;
  end
end
return $\mathbf{R}_{\text{avg}}$;

Algorithm 1 provide our implementation for single rotation averaging. Empirically, after registration and minor filtering of the rotation samples[4], we observed per frame rotation samples are reasonably close, which is good for convergence [23]. Averaging per frame rotation priors, we recover $\mathbf{R}_{\text{avg}} \in \mathbb{R}^{3F\times 3}$. For more details on single rotation averaging and its convergence analysis refer [23,38].

To compute $\mathbf{R} \in \mathbb{R}^{2F\times 3F}$ from $\mathbf{R}_{\text{avg}} \in \mathbb{R}^{3F\times 3}$, we take $\mathbf{R}_{\text{avg}}$'s per frame 3×3 matrix, drop its 3^{rd} row and place it to the diagonal of $\mathbf{R}$, and perform this step for all frame[5].

3.2 Shape Estimation

Once we estimated the rotation matrix, our goal is to recover the shape matrix. An easy way to compute shape is $\mathbf{X}_{init} = \texttt{pinv}(\mathbf{R})\mathbf{W}$, which is consistent with the assumption of low rank shape matrix and it minimizes the re-projection error. To show the merit of our rotation estimation, we tabulate the pseudo inverse shape reconstruction result using our rotation compared to BMM [13] in Tab.(1). Clearly, our rotation improves the $\mathbf{X}_{init}$, *i.e.*, the initial shape solution,

[4] filter if sample is too far to the reference rotation after registration.
[5] After registration, if samples are filtered out due to its distance from the reference rotation (more than δ), then per frame rotations is less than K.

Dataset	Drink	Pickup	Yoga	Stretch	Dance	Face	Walking	Shark
BMM-PI [13]	0.4449	0.5989	0.6523	0.4784	0.5764	0.4848	0.5100	0.8784
Ours-PI	**0.2195**	**0.2985**	**0.2740**	**0.2238**	**0.3014**	**0.2995**	**0.2702**	**0.3053**

Table 1: Pseudo inverse (PI) shape results comparison with BMM [13] via e_{3d} metric. Compared to [13], our approach dramatically improves PI shape accuracy, showing the benefit of using rotation organic prior. e_{3d} definition is provided in Sec.§4

by a large margin. However, $\mathbf{X}_{init}$ may be a 3D reconstruction co-planar in every frame and "the catch is that there are in fact many solutions which minimize the rank" [52]. Therefore, further optimization of the shape matrix is recommended. Let's review Dai *et al.*'s [13] relaxed rank-minimization and recent improvement over it to better place our approach.

Relation to previous methods. Given rotation matrix $\mathbf{R} \in \mathbb{R}^{2F\times 3F}$, Dai *et al.* [13] perform the following optimization for low-rank shape matrix estimation.

$$\underset{\mathbf{X}^{\sharp},\mathbf{X}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{W}-\mathbf{R}\mathbf{X}\|_{\mathcal{F}}^{2}+\mu\|\mathbf{X}^{\sharp}\|_{*}; \text{subject to: } \mathbf{X}^{\sharp}=\mathbf{\Phi}(\mathbf{X}) \tag{7}$$

Here, $\mathbf{\Phi}(.)$ is a function that maps $\mathbf{X} \in \mathbb{R}^{3F\times P}$ to $\mathbf{X}^{\sharp} \in \mathbb{R}^{F\times 3P}$. μ is a scalar constant and $\|.\|_{*}$ denotes the nuclear norm of the matrix, which is a convex bound of the matrix rank and can give a good solution to rank minimization problems under restricted isometry property constraints [44]. Eq.(7) can be solved efficiently using the ADMM optimization strategy [7]. [13] optimizes Eq.(7) by penalizing each singular value of $\mathbf{X}^{\sharp}$ equally. Yet, we have an initial shape prior $\mathbf{X}_{init}$ that we can exploit to recover a better shape. In the same vein, Kumar [28] introduced WNN minimization to Eq.(7), which shows highly effective results with the use of $\mathbf{X}_{init}$ singular values as prior. [28] suggested the following changes

$$\underset{\mathbf{X}^{\sharp},\mathbf{X}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{W}-\mathbf{R}\mathbf{X}\|_{\mathcal{F}}^{2}+\mu\|\mathbf{X}^{\sharp}\|_{*,\theta}; \text{subject to: } \mathbf{X}^{\sharp}=\mathbf{\Phi}(\mathbf{X}) \tag{8}$$

Here, θ is the weight assigned to $\mathbf{X}^{\sharp}$ based on the $\mathbf{X}^{\sharp}_{init}$ singular values. It is known that for a low-rank shape matrix, a few top singular values contain most of the shape information. Thus, when optimizing Eq.(8) the first singular value should be penalized the least and vice-versa, using the following relation

$$\theta_i = \xi\big(\sigma_i(\mathbf{X}^{\sharp}_{init})+\gamma\big)^{-1} \tag{9}$$

where, ξ and γ are small positive scalars, $\sigma_i(\mathbf{X}^{\sharp}_{init})$ denotes the i^{th} singular value of $\mathbf{X}^{\sharp}_{init}$ and θ_i denotes its weight. It is observed and empirically tested that such a strategy provides significantly better minima after optimization [28].

Nevertheless, for shape estimation, contrary to Kumar [28], we propose the mixed use of partial sum minimization of singular values and weighted nuclear norm optimization of $\mathbf{X}^{\sharp}$. Based on our extensive empirical study over several non-rigid shapes, we found that the first singular value of $\mathbf{X}^{\sharp}_{init}$ contains useful

information. Thus, penalizing it during WNN minimization may hurt performance unnecessarily. Therefore, we propose to preserve the first singular value of the shape during optimization, leading to the following optimization problem

$$\underset{\mathbf{X}^\sharp,\mathbf{X}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{W}-\mathbf{R}\mathbf{X}\|_{\mathcal{F}}^2+\mu\|\mathbf{X}^\sharp\|_{r=N,\theta};\ \text{subject to: } \mathbf{X}^\sharp=\mathbf{\Phi}(\mathbf{X}) \tag{10}$$

We use $N=1$ for all our experiments and assign the weights θ using Eq.(9) for the rest of the singular values in the shape matrix optimization via ADMM [7].

Shape optimization. We optimized Eq.(10) using ADMM [7]. Introducing the Lagrange multiplier in Eq.(10) gives us

$$\begin{aligned}\mathcal{L}_\rho(\mathbf{X}^\sharp,\mathbf{X})&=\frac{1}{2}\|\mathbf{W}-\mathbf{R}\mathbf{X}\|_{\mathcal{F}}^2+\mu\|\mathbf{X}^\sharp\|_{r=N,\theta}\\&+\frac{\rho}{2}\|\mathbf{X}^\sharp-\mathbf{\Phi}(\mathbf{X})\|_{\mathcal{F}}^2+<\mathbf{Y},\mathbf{X}^\sharp-\mathbf{\Phi}(\mathbf{X})>\end{aligned} \tag{11}$$

$\mathbf{Y}\in\mathbb{R}^{F\times 3P}$ is the Lagrange multiplier and $\rho>0$ is the penalty parameter. We obtain the solution to each variable solving the following sub-problems over iterations (indexed with the variable t):

$$\mathbf{X}_{t+1}=\underset{\mathbf{X}}{\text{argmin}}\ \mathcal{L}_{\rho t}(\mathbf{X}^\sharp,\mathbf{X}_t);\ \mathbf{X}^\sharp_{t+1}=\underset{\mathbf{X}^\sharp}{\text{argmin}}\ \mathcal{L}_{\rho t}(\mathbf{X}^\sharp_t,\mathbf{X}) \tag{12}$$

Using Eq.(11)-Eq.(12), we derive the following expression for $\mathbf{X}$, assuming $\mathbf{X}^\sharp$ is constant.

$$\mathbf{X}\simeq\underset{\mathbf{X}}{\text{argmin}}\frac{1}{2}\|\mathbf{W}-\mathbf{R}\mathbf{X}\|_{\mathcal{F}}^2+\frac{\rho}{2}\left\|\mathbf{X}-\left(\mathbf{\Phi}^{-1}(\mathbf{X}^\sharp)+\frac{\mathbf{\Phi}^{-1}(\mathbf{Y})}{\rho}\right)\right\|_{\mathcal{F}}^2 \tag{13}$$

The closed form solution for $\mathbf{X}$ is obtained by taking the derivative of Eq.(13) w.r.t the corresponding variable and equating it to zero. The closed form expression is used during the ADMM iteration until convergence to recover optimal $\mathbf{X}$. Similarly, rewriting the Eq.(11) by assuming $\mathbf{X}^\sharp$ as variable and $\mathbf{X}$ as constant, we get the following expression for $\mathbf{X}^\sharp$

$$\mathbf{X}^\sharp\simeq\underset{\mathbf{X}^\sharp}{\text{argmin}}\ \mu\|\mathbf{X}^\sharp\|_{r=N,\theta}+\frac{\rho}{2}\left\|\mathbf{X}^\sharp-\left(\mathbf{\Phi}(\mathbf{X})-\frac{\mathbf{Y}}{\rho}\right)\right\|_{\mathcal{F}}^2 \tag{14}$$

To solve Eq.(14), we used the theory of Partial Singular Value Thresholding (PSVT) [40]. Let $\mathcal{P}_{N,\tau}[\mathbf{Q}]$ denote the PSVT operator operating on matrix $\mathbf{Q}$. The operator preserves the leading N singular values and penalizes the others with soft-thresholding parameter τ [6]. For completeness, let's go over the following:

Theorem 1. *Oh et al. [40] proposed the following optimization problem to solve*

$$\underset{\mathbf{P}}{\text{argmin}}\ \tau\|\mathbf{P}\|_{r=N}+\frac{1}{2}\|\mathbf{P}-\mathbf{Q}\|_{\mathcal{F}}^2 \tag{15}$$

[6] For more discussion on partial sum minimization of singular values, cf. the supplementary material. For a comprehensive theory refer to [40].

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	0.0550 ($K = 3$)

Table 2: Statistical comparison on the MoCap dataset [3]. Our method provides favorable 3D reconstruction results. Contrary to the R-BMM [28], our approach provides a methodical way to solve NRSfM factorization irrespective of the camera motion assumption. The value of K used is generally same as [12, 13]. The 2$^{\text{nd}}$ best results are underlined. `To have clear spacing, we put comparison with other methods as suggested by the reviewers in the supplementary material`.

where, $\tau > 0$ and $\mathbf{P}$, $\mathbf{Q} \in \mathbb{R}^{m \times n}$ be real valued matrices which can be decomposed by Singular Value Decomposition (SVD). Then, the optimal solution can be expressed by the PSVT operator defined as:

$$\mathcal{P}_{N,\tau}[\mathbf{Q}] = \mathbf{U}_Q(\mathbf{\Sigma}_{Q1} + \mathcal{S}_\tau[\mathbf{\Sigma}_{Q2}])\mathbf{V}_Q^T \tag{16}$$

where, $\mathbf{\Sigma}_{Q1} = \mathbf{diag}(\sigma_1, \sigma_2, .., \sigma_N, .., 0)$ and $\mathbf{\Sigma}_{Q2} = \mathbf{diag}(0, ..\sigma_{N+1}, .., \max(m, n))$. Symbol $\mathcal{S}_\tau$ is the soft-thresholding operator defined as $\mathcal{S}_\tau(\sigma) = \text{sign}(\sigma)\max(|\sigma| - \tau, 0)$. $\mathbf{Q} = \mathbf{U}_Q(\mathbf{\Sigma}_{Q1} + \mathbf{\Sigma}_{Q2})\mathbf{V}_Q^T$

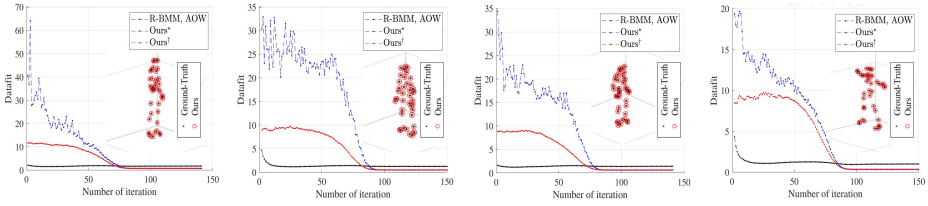
For a detailed derivation and proof, we refer to Oh *et al.* [40]. Using the theorem, we substitute $N = 1$, $\tau = (\mu\theta)/\rho$ and write the solution of $\mathbf{X}^\sharp$ in Eq.(14) as:

$$\mathbf{X}^\sharp = \mathcal{P}_{1,\frac{\mu\theta}{\rho}}\Big[\big(\mathbf{\Phi}(\mathbf{X}) - \rho^{-1}\mathbf{Y}\big)\Big] \tag{17}$$

We use the above expression of $\mathbf{X}^\sharp$ during the ADMM optimization [7] to recover the optimal shape matrix. The θ values are assigned according to Eq.(9) for $N > 1$. The Lagrange multiplier ($\mathbf{Y}$) and penalty parameter (ρ) are updated over ADMM iteration (say for $t+1$ iteration) as $\mathbf{Y}_{t+1} = \mathbf{Y}_t + \rho(\mathbf{X}^\sharp_{t+1} - \mathbf{\Phi}(\mathbf{X}_{t+1}))$; $\rho_{t+1} = \text{minimum}(\rho_{\max}, \lambda\rho_t)$. Where, $\rho_{\max}$ refers to the maximum value of 'ρ' and λ is an empirical constant. $\mathbf{Y}$ and ρ are updated during the ADMM [7] iteration until convergence criteria is satisfied. The criteria for the ADMM iteration to stop are $\|\mathbf{X}^\sharp - \mathbf{\Phi}(\mathbf{X})\|_\infty < \epsilon$, or, $\rho_{t+1} \geq \rho_{\max}$

4 Experiments

(a) Implementation Details and Initialization. We implemented our method on a desktop machine with 32GB RAM using C++/MATLAB software. Initial rotation filtering parameter δ is set to 0.05. All the K rotations $\mathbf{R}^k \in \mathbb{R}^{2F \times 3}$ can be solved in parallel, so the minor increase in processing time compared to [13], is due to registration, filtering and rotation averaging. For *e.g.*, a 357 frame pickup sequence takes 7.46s. for registration, 0.012s. for filtering and 0.93s. for computing $\mathbf{R}_{\text{avg}}$. We ran 50 iterations of single rotation averaging (**Algorithm** 1) for



(a) Drink (b) Pickup (c) Stretch (d) Yoga

Fig. 2: Datafit curve shows the value of Eq.(10) cost function over iteration compared to R-BMM [28] and AOW [24] WNN formulation on Mocap dataset [3]. Ours* show the Eq.(10) datafit value using [13] rotation whereas, Ours† show the datafit curve using our rotation initialization. Our shape optimization gives better minima, and using our rotation as initialization, we have faster and stable convergence response (Ours†).

each frame. The weights (θ) in Eq.(10) is initialized using Eq.(9) with $\xi = 5e^{-3}$ generally and $\gamma = 1e^{-6}$. For Eq.(10) optimization via ADMM, we initialized $\rho = 1e^{-4}$, $\lambda = 1.1$, $\rho_{\max} = 1e^{10}$, $\mathbf{Y} = \text{zeros}(F, 3P)$, $\epsilon_t = 1e^{-10}$, and $\epsilon = 1e^{-10}$.

(b) Evaluation Metric. We used the popular mean normalized 3D reconstruction error metric to report our statistical results on motion capture (MoCap) benchmark [3, 50] and Garg *et al.* [15] dense NRSfM benchmark dataset. It is defined as $e_{3d} = \frac{1}{F}\sum_{i=1}^{F} \|\mathbf{X}_i^{est} - \mathbf{X}_i^{gt}\|_{\mathcal{F}} / \|\mathbf{X}_i^{gt}\|_{\mathcal{F}}$ with $\mathbf{X}_i^{est}$, $\mathbf{X}_i^{gt}$ symbolizing per frame estimated shape and its ground-truth (GT) value, respectively. For evaluation on recent NRSfM benchmark dataset [25], we used their supplied error evaluation metric script, which is inspired from Taylor *et al.* work [48]. The 3D reconstruction accuracy is computed after registering the recovered shape to the ground-truth shape due to global ambiguity [3,25]. To evaluate rotation estimate accuracy, we use the mean rotation error metric $e_{\mathbf{R}} = \frac{1}{F}\sum_{i=1}^{F} \|\mathbf{R}_i^{\text{GT}} - \mathbf{R}_i^{\text{est}}\|_{\mathcal{F}}$. Here, $\mathbf{R}_i^{\text{GT}}$, $\mathbf{R}_i^{\text{est}}$ denotes the ground-truth and estimated per frame rotation.

4.1 Dataset and Evaluation.

(a) MoCap Benchmark Dataset. Introduced by Akther *et al.* [3] and Torresani *et al.* [50], this dataset has become a standard benchmark for any NRSfM algorithm evaluation. It is composed of 8 real sequences, namely Drink (1102, 41), Pickup (357, 41), Yoga (307, 41), Stretch (370, 41), Dance (264, 75), Walking (260, 55), Face (316, 40) and Shark (240, 91). The last 3 sequences were introduced by Torresani *et al.* [50]. The numbers presented in bracket correspond to number of frames and points (F, P). Tab.(2) shows the comparison of our method with other competing methods. For evaluation, we keep the value of K generally same as BMM [13]. From Tab.(2), it is easy to observe that more often than not, our approach performs best or second-best than other methods, thus showing a consistent superior performance over a diverse set of object deformation type.

Compared to BMM [13], which also makes no assumption other than low-rank, our «organic prior» based method dramatically improves 3D reconstruction accuracy, thereby validating our claims made in the paper. Fig.(2) shows few qualitative results along with the convergence curve comparison with the current

Method Type →	Sparse NRSfM Methods						Dense NRSfM 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

Table 3: 3D reconstruction accuracy on dense NRSfM dataset [15]. We observed superior results compared to the well-known sparse NRSfM methods. It is interesting to observe that our results compares favorably to carefully crafted dense NRSfM methods such as DV, GM and others. The 2$^{\text{nd}}$ best performance of our method is underlined.

methods such as R-BMM [28], AOW [24]. The results show recovery of better minima and stable convergence curve using our rotation estimate initialization.

(b) Dense NRSfM Benchmark Dataset. Introduced by Garg *et al.* [15,16], it is a standard dataset to evaluate dense NRSfM methods. It comprises of 4 synthetic face sequences and 3 real video sequences of heart, back, and face deformation. The synthetic face dataset is composed of 28,880 tracked feature points. Face sequence 1 and Face sequence 2 are 10 frames long video, whereas Face sequence 3 and Face sequence 4 are 99 frames video. The video sequence for heart, back, and face dataset is 80, 150, and 120 frames long with 68295, 20561, and 28332 feature track points. Tab.(3) provides the statistical results of our approach compared to well-known dense NRSfM algorithms. For better comprehension, we classified the comparison into two sets *i.e.*, sparse NRSfM methods and dense NRSfM methods. From Tab.(3), it is easy to observe the advantage of our approach compared to well-known sparse NRSfM methods. For evaluation of our method, we use $K = 1$ for all the four sequence. For other methods [3,13], we iterate over different K and put its best possible results.

The interesting point to note is that without using any extra assumptions about the dense deforming surface such as union of linear subspaces [26, 29], variation in the deformation over frame should be smooth [16], dynamic shape prior [17], smooth trajectory constraint [5], and recent deep neural network based latent space constraint [47], our method provide impressive results and it is close to the best method [29]. Note that, contrary to our simple approach, GM [29] is a complex geometric method to implement. To conclude, our results reveal the strength of classical NRSfM factorization if organic priors are exploited sensibly.

(c) NRSfM Challenge Dataset. Jensen *et al.* [25] recently proposed this dataset. It comprises 5 different subjects, namely Articulated, Paper, Balloon, Stretch, and Tearing. Each subject's deformations is captured under 6 varying camera trajectories *i.e.*, circle, flyby, line, semi-circle, tricky and zigzag, making the dataset interesting yet challenging. For evaluation, the dataset provide a single frame ground-truth 3D

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** ($K = 6$)

Table 4: Comparison of our method with state-of-the-art on recent benchmark [25]. Results are reported in millimeters.

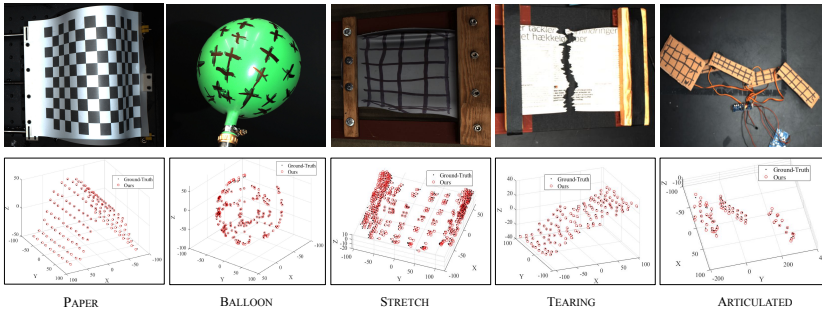


Fig. 3: Qualitative Results on NRSfM challenge dataset [25]. **Top row.** Subject image. **Bottom row.** 3D reconstruction of the respective object shape.

shape for each subject. Tab.(4) show the average 3D reconstruction accuracy comparison in millimeters with the recent and earlier state-of-the-art on this dataset *i.e.*, BMM [13], R-BMM [28], AOW [24], BP [41]. For comparison, we used the orthogonal sequence of the dataset. The value of K used by our method for comparison is provided in the bracket. Statistical results indicate that our approach provides better non-rigid shape reconstruction for most of the subject categories on this dataset. Fig.(3) show visual results obtained on this dataset.

(d) Rotation Estimation. To validate that the single rotation averaging gives meaningful rotation, we validate our results using the ground-truth rotation available in the Akther *et al.* [3] dataset.

Tab.(5) provide the average camera rotation error $e_{\mathbf{R}}$ results on yoga, pickup, stretch, and drink sequence. The statistics show that using our approach, we can have fine rotation estimate[7]. Further, advantage of our rotation estimation on clean sequence, noisy trajectories and pseudo inverse solution can be inferred from Tab.(6) Fig.4(a), and Tab.(1), respectively.

Data	MP [42]	PTA [3]	CSF [18]	BMM [13]	R-BMM [28]	Ours
Yoga	0.8343	0.1059	0.1019	0.0883	0.0883	0.0888
Pickup	0.2525	0.1549	0.1546	0.1210	0.1217	**0.1144**
Stretch	0.8185	0.0549	0.0489	0.0676	0.0676	0.0671
Drink	0.2699	0.0058	0.0055	0.0071	0.0243	0.0072

Table 5: $e_{\mathbf{R}}$ comparison with other factorization methods.

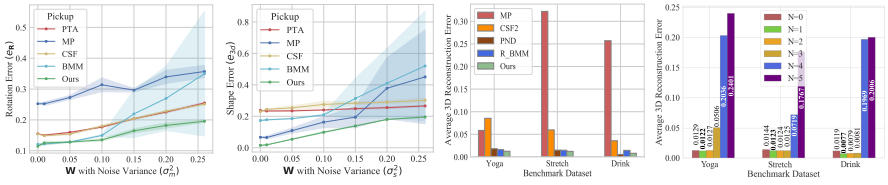
Dataset	Drink	Pickup	Yoga	Stretch	Dance	Face	Walking	Shark
BMM [13]	0.0266	0.1731	0.1150	0.1034	0.1864	0.0303	0.1298	0.2357
e_{3d} ([13] rotation)	0.0101	0.0164	0.0126	0.0126	0.1382	0.0152	0.0880	0.0563
e_{3d} (our rotation)	**0.0071**	**0.0152**	**0.0122**	**0.0124**	**0.1209**	**0.0145**	**0.0816**	**0.0550**

Table 6: 2$^{\mathbf{nd}}$ **row**: Our e_{3d} results using Dai *et al.* rotation [13]. 3$^{\mathbf{rd}}$ **row**: e_{3d} using our rotation. Indeed using organic rotation priors help improve overall performance.

(e) Other Experiments and Ablations.

(i) *Performance with noisy trajectory.* Fig.4(a), Fig.4(b) shows the rotation ($e_{\mathbf{R}}$) and shape error (e_{3d}) comparison on the noisy trajectory, respectively. We introduce noise to the 2D point trajectory with the standard deviation varying from 0.01-0.25 using `normrand()` function from MATLAB. We ran the different

[7] With $\mathbf{W} = \mathbf{RS}$ theory, even GT rotation cannot provide GT shape, cf. [13] Table(3)



(a) $e_{\mathbf{R}}$ w.r.t noise (b) e_{3d} w.r.t noise (c) missing data. (d) e_{3d} w.r.t N

Fig. 4: (a)-(b) Rotation and shape error on the noisy Pickup trajectories. Our method show stable behaviour consistently. Mean and standard deviation is shown with bold and light shaded regions, respectively. (c) Avg. 3D reconstruction for missing data experiment. (d) $N = 0$ show results when all the singular values are penalized using WNN. Our approach *i.e.*, $N = 1$ gives better results overall under same rotation.

methods 10 times for each standard deviation value and plotted the method's mean and variance. Statistical results show that our method is quite robust to noisy sequence and show much stable behaviour (both in rotation and shape estimation) than the other prior or prior-free approaches (see Fig.4(a)-4(b)).
(ii) *Performance on missing trajectory cases.* For this experiment, we used Lee *et al.* [36] and Kumar [28] setup, where we randomly set 30% of the trajectory missing from the $\mathbf{W}$. We perform matrix completion using [11] optimization and then ran our algorithm on the recovered matrix. The results are shown in Fig.4(c). Our method outperforms the state-of-the-art in most of the cases.
(iii) *Performance with change in value of* N. To show that $N = 1$ generally works best for Eq.(10), we conducted this experiment. First, we penalize all the singular values using WNN optimization ($N = 0$) and then we vary the value of N from 1 to 5 and recorded the results. Fig.4(d) shows the reconstruction results using different values of N. We observed that by penalizing all the singular values using WNN, we are unnecessarily hurting the performance. On the contrary, if we increase N value greater than 1, `more often than not`, it starts to reduce the performance. Refer supplementary material for more results and discussions.

5 Conclusion

This work reveals organic priors for NRSfM factorization irrespective of camera motion and shape deformation type. It exhibited that mindful use of such fundamental priors gives better accuracy than the prior-free methods. That said, our method uses an orthographic camera model with a low-rank shape assumption in NRSfM. Hence, by construction, it has some limitations *for e.g.*, our method may perform inadequately on high perspective distortion images having large object deformation. A recent idea by Graßhof et al. [21] can be used to overcome such a limitation. Finally, we conclude that the clever use of organic priors with matrix factorization theory is sufficient to provide excellent 3D reconstruction accuracy for both sparse and dense NRSfM.

Acknowledgement. The authors thank Google for their generous gift (ETH Zürich Foundation, 2020-HS-411).

References

1. Aanæs, H., Kahl, F.: Estimation of deformable structure and motion. In: Proceedings of the Vision and Modelling of Dynamic Scenes Workshop. vol. 2, p. 3. Citeseer (2002)
2. Akhter, I., Sheikh, Y., Khan, S.: In defense of orthonormality constraints for non-rigid structure from motion. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. pp. 1534–1541. IEEE (2009)
3. Akhter, I., Sheikh, Y., Khan, S., Kanade, T.: Nonrigid structure from motion in trajectory space. In: Advances in neural information processing systems. pp. 41–48 (2008)
4. Akhter, I., Sheikh, Y., Khan, S., Kanade, T.: Trajectory space: A dual representation for nonrigid structure from motion. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(7), 1442–1456 (2011)
5. Ansari, M.D., Golyanik, V., Stricker, D.: Scalable dense monocular surface reconstruction. In: 2017 International Conference on 3D Vision (3DV). pp. 78–87. IEEE (2017)
6. Bartoli, A., Gay-Bellile, V., Castellani, U., Peyras, J., Olsen, S., Sayd, P.: Coarse-to-fine low-rank structure-from-motion. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2008)
7. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning **3**(1), 1–122 (2011)
8. Brand, M.: A direct method for 3d factorization of nonrigid motion observed in 2d. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 2, pp. 122–128. IEEE (2005)
9. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3d shape from image streams. In: IEEE Conference on Computer Vision and Pattern Recognition. vol. 2, pp. 690–696. IEEE (2000)
10. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Numerical geometry of non-rigid shapes. Springer Science & Business Media (2008)
11. Cabral, R., De la Torre, F., Costeira, J.P., Bernardino, A.: Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2488–2495 (2013)
12. Dai, Y., Li, H., He, M.: A simple prior-free method for non-rigid structure-from-motion factorization. In: Computer Vision and Pattern Recognition (CVPR). pp. 2018–2025. IEEE (2012)
13. Dai, Y., Li, H., He, M.: A simple prior-free method for non-rigid structure-from-motion factorization. International Journal of Computer Vision **107**(2), 101–122 (2014)
14. Fayad, J., Agapito, L., Del Bue, A.: Piecewise quadratic reconstruction of non-rigid surfaces from monocular sequences. In: European conference on computer vision. pp. 297–310. Springer (2010)
15. Garg, R., Roussos, A., Agapito, L.: Dense variational reconstruction of non-rigid surfaces from monocular video. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1272–1279 (2013)
16. Garg, R., Roussos, A., Agapito, L.: A variational approach to video registration with subspace constraints. International journal of computer vision **104**(3), 286–314 (2013)

17. Golyanik, V., Jonas, A., Stricker, D., Theobalt, C.: Intrinsic dynamic shape prior for dense non-rigid structure from motion. In: 2020 International Conference on 3D Vision (3DV). pp. 692–701. IEEE (2020)
18. Gotardo, P.F., Martinez, A.M.: Computing smooth time trajectories for camera and deformable shape in structure from motion with occlusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(10), 2051–2065 (2011)
19. Gotardo, P.F., Martinez, A.M.: Kernel non-rigid structure from motion. In: IEEE International Conference on Computer Vision. pp. 802–809. IEEE (2011)
20. Gotardo, P.F., Martinez, A.M.: Non-rigid structure from motion with complementary rank-3 spaces. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. pp. 3065–3072. IEEE (2011)
21. Graßhof, S., Brandt, S.S.: Tensor-based non-rigid structure from motion. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 3011–3020 (2022)
22. Hartley, R., Aftab, K., Trumpf, J.: L1 rotation averaging using the weiszfeld algorithm. In: CVPR 2011. pp. 3041–3048. IEEE (2011)
23. Hartley, R., Trumpf, J., Dai, Y., Li, H.: Rotation averaging. International journal of computer vision **103**(3), 267–305 (2013)
24. Iglesias, J.P., Olsson, C., Valtonen Örnhag, M.: Accurate optimization of weighted nuclear norm for non-rigid structure from motion. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16. pp. 21–37. Springer (2020)
25. 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)
26. Kumar, S.: Jumping manifolds: Geometry aware dense non-rigid structure from motion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5346–5355 (2019)
27. Kumar, S.: Non-Rigid Structure from Motion. Ph.D. thesis, College of Engineering & Computer Science, The Australian National University (2019)
28. Kumar, S.: Non-rigid structure from motion: Prior-free factorization method revisited. In: Winter Conference on Applications of Computer Vision (WACV 2020). pp. 51–60 (2020)
29. Kumar, S., Cherian, A., Dai, Y., Li, H.: Scalable dense non-rigid structure-from-motion: A grassmannian perspective. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 254–263 (2018)
30. Kumar, S., Dai, Y., H.Li: Spatio-temporal union of subspaces for multi-body non-rigid structure-from-motion. Pattern Recognition **71**, 428–443 (May 2017)
31. Kumar, S., Dai, Y., Li, H.: Multi-body non-rigid structure-from-motion. In: 3D Vision (3DV), 2016 Fourth International Conference on. pp. 148–156. IEEE (2016)
32. Kumar, S., Dai, Y., Li, H.: Monocular dense 3d reconstruction of a complex dynamic scene from two perspective frames. In: IEEE International Conference on Computer Vision. pp. 4649–4657 (Oct 2017)
33. Kumar, S., Dai, Y., Li, H.: Superpixel soup: Monocular dense 3d reconstruction of a complex dynamic scene. IEEE transactions on pattern analysis and machine intelligence **43**(5), 1705–1717 (2019)
34. Kumar, S., Ghorakavi, R.S., Dai, Y., Li, H.: Dense depth estimation of a complex dynamic scene without explicit 3d motion estimation. arXiv preprint arXiv:1902.03791 (2019)

35. Kumar, S., Van Gool, L., de Oliveira, C.E., Cherian, A., Dai, Y., Li, H.: Dense non-rigid structure from motion: A manifold viewpoint. arXiv preprint arXiv:2006.09197 (2020)
36. Lee, M., Cho, J., Choi, C.H., Oh, S.: Procrustean normal distribution for non-rigid structure from motion. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1280–1287 (2013)
37. Lee, M., Cho, J., Oh, S.: Consensus of non-rigid reconstructions. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4670–4678 (2016)
38. Lee, S.H., Civera, J.: Robust single rotation averaging. arXiv preprint arXiv:2004.00732 (2020)
39. Matthews, I., Xiao, J., Baker, S.: 2d vs. 3d deformable face models: Representational power, construction, and real-time fitting. International journal of computer vision **75**(1), 93–113 (2007)
40. Oh, T.H., Tai, Y.W., Bazin, J.C., Kim, H., Kweon, I.S.: 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)
41. Ornhag, M.V., Iglesias, J.P., Olsson, C.: Bilinear parameterization for non-separable singular value penalties. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3897–3906 (2021)
42. Paladini, M., Del Bue, A., Stosic, M., Dodig, M., Xavier, J., Agapito, L.: Factorization for non-rigid and articulated structure using metric projections. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2898–2905. IEEE (2009)
43. Rabaud, V., Belongie, S.: Re-thinking non-rigid structure from motion. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2008)
44. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM review **52**(3), 471–501 (2010)
45. Russell, C., Fayad, J., Agapito, L.: Energy based multiple model fitting for non-rigid structure from motion. In: CVPR 2011. pp. 3009–3016. IEEE (2011)
46. Salzmann, M., Fua, P.: Deformable surface 3d reconstruction from monocular images. Synthesis Lectures on Computer Vision **2**(1), 1–113 (2010)
47. Sidhu, V., Tretschk, E., Golyanik, V., Agudo, A., Theobalt, C.: Neural dense non-rigid structure from motion with latent space constraints. In: European Conference on Computer Vision. pp. 204–222. Springer (2020)
48. Taylor, J., Jepson, A.D., Kutulakos, K.N.: Non-rigid structure from locally-rigid motion. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2761–2768. IEEE (2010)
49. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. International Journal of Computer Vision **9**(2), 137–154 (1992)
50. Torresani, L., Hertzmann, A., Bregler, C.: Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. IEEE transactions on pattern analysis and machine intelligence **30**(5), 878–892 (2008)
51. Torresani, L., Hertzmann, A., Bregler, C.: Learning non-rigid 3d shape from 2d motion. In: Advances in Neural Information Processing Systems. pp. 1555–1562 (2004)
52. Valmadre, J., Sridharan, S., Denman, S., Fookes, C., Lucey, S.: Closed-form solutions for low-rank non-rigid reconstruction. In: 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA). pp. 1–6. IEEE (2015)

53. Xiao, J., Chai, J.x., Kanade, T.: A closed-form solution to non-rigid shape and motion recovery. In: European conference on computer vision. pp. 573–587. Springer (2004)
54. Zhu, Y., Huang, D., De La Torre, F., Lucey, S.: Complex non-rigid motion 3d reconstruction by union of subspaces. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1542–1549 (2014)