

# A Simple Prior-Free Method for Non-rigid Structure-from-Motion Factorization

Yuchao Dai · Hongdong Li · Mingyi He

Received: 1 January 2013 / Accepted: 27 November 2013 / Published online: 31 December 2013  
© Springer Science+Business Media New York 2014

**Abstract** This paper proposes a simple “prior-free” method for solving the non-rigid structure-from-motion (NRSfM) factorization problem. Other than using the fundamental low-order linear combination model assumption, our method does not assume any extra prior knowledge either about the non-rigid structure or about the camera motions. Yet, it works effectively and reliably, producing optimal results, and not suffering from the inherent basis ambiguity issue which plagued most conventional NRSfM factorization methods. Our method is very simple to implement, which involves solving a very small SDP (semi-definite programming) of fixed size, and a nuclear-norm minimization problem. We also present theoretical analysis on the uniqueness and the relaxation gap of our solutions. Extensive experiments on both synthetic and real motion capture data (assuming following the low-order linear combination model) are conducted, which demonstrate that our method indeed outperforms most of the existing non-rigid factorization methods. This work offers not only new theoretical insight, but also a practical, everyday solution to NRSfM.

**Keywords** Non-rigid structure-from-motion · Nuclear norm minimization · Prior-free · Uniqueness · Rank minimization

## 1 Introduction

Non-rigid objects are ubiquitous in our surroundings, e.g., from a walking pedestrian to a talking face, from waving tree branches to rattling water surfaces and so on. Non-rigid structure-from-motion (NRSfM) aims to recover 3D non-rigid/deformable structure and camera motions from 2D correspondences across multiple images, which has been receiving increasing attentions from the community.

This paper revisits NRSfM and focuses on the *factorization* framework for NRSfM, originally proposed by Bregler et al. (2000), as an important extension to the well-known Tomasi–Kanade factorization (Tomasi and Kanade 1992) from rigid scene to non-rigid scene, assuming that the non-rigid shape deformation follows a low-order linear combination model. To date, a large body of research has been devoted to this topic, and numerous different methods/algorithms have been proposed. However, despite all the efforts, this problem remains a difficult and still active research topic—no fully satisfactory solution is available today. For example, in a 2004 textbook (Hartley and Zisserman 2004, p. 444): “currently there is not a satisfactory solution”. In CVPR 2009 (Paladini et al. 2009): “although this low-rank shape model has proved a successful representation, the NRSfM problem is inherently under-constrained”. In ICCV 2011 Gotardo and Martinez (2011b): “in the absence of any (extra) prior knowledge on 3D shape deformation, computing NRSfM is still a difficult, underconstrained problem” (see e.g., Bartoli et al. 2008; Del Bue 2008; Torresani et al. 2008; Paladini et al. 2009; Akhter et al. 2008; Gotardo and Martinez 2011a).

---

Y. Dai (✉) · M. He  
School of Electronics and Information, Northwestern  
Polytechnical University, Xi'an, Shaanxi, China  
e-mail: daiyuchao@gmail.com

M. He  
e-mail: myhe@nwpu.edu.cn

Y. Dai · H. Li  
Australian National University, Canberra, ACT, Australia  
e-mail: hongdong.li@anu.edu.au

H. Li  
NICTA, Canberra, ACT, Australia

One of the primary causes to such difficulty is due to the inherent *basis ambiguity* (ambiguity between deformable shape bases and time-varying coefficients or between trajectory bases and point-varying coefficients) of the non-rigid problem (Xiao et al. 2004). To overcome this, most existing work rely on introducing various *prior* knowledge to the problem at hand. They do so by assuming various constraints about the non-rigid scene (Bartoli et al. 2008), about the non-rigid shape bases (Xiao et al. 2004), about the time-varying coefficients (Torresani et al. 2008), about the deformation (Olsen and Bartoli 2008), or about the camera motion (Gotardo and Martinez 2011a), etc. For instance, many methods require that the camera moves smoothly, or the deformation trajectory is slow and smooth (Akhter et al. 2008; Gotardo and Martinez 2011a). However, these additional constraints not only limit the practical applicability of the factorization framework, but also obscure a clear theoretical understanding to the problem. We would like to answer: in order to solve the non-rigid factorization problem effectively, are these extraneous priors really *essential*?

In this paper, we propose a novel and simple solution to non-rigid factorization. Our method does not assume any extra prior knowledge about the problem other than the low-rank constraint, hence it is “prior-free”. Nevertheless, it does not suffer from the basis ambiguity difficulty, but is able to recover both camera motion and non-rigid shape accurately and reliably. Experiments on both synthetic and real benchmark data-sets show that: the proposed method, despite of being prior-free, outperforms most other (often prior-based) linear factorization methods. Furthermore, we give theoretical analysis on the uniqueness and degenerate cases of the solution.

To better present this paper, and also to put our contributions in context, we briefly review recent progress in factorization based NRSfM.

### 1.1 Related Works

As one of the significant problems in geometric computer vision, NRSfM has received considerable attention from the computer vision community. A great number of methods have been published, and most of the existing methods can be roughly classified into two categories: The first category is *template-based NRSfM methods* (e.g., Salzmann et al. 2008; Brunet et al. 2011; Perriollat et al. 2011) which assume a known 3D template (e.g. a mesh model) of the non-rigid surface. The second category is *correspondence-based methods* where inter-frame (and often sparse) feature point correspondences are available. Most structure-from-motion techniques used for rigid scene reconstruction (e.g., Pollefeys et al. 2004; Snavely et al. 2008; Dai et al. 2010; Li 2010) rely on feature point correspondences.

In this paper, we focus on the second category of methods, i.e. the correspondence-based methods for non-rigid structure from motion. Moreover, in the following discussions regarding related works, we confine ourselves exclusively to the matrix-factorization framework (Tomasí and Kanade 1992; Bregler et al. 2000).

Ever since Bregler et al.’s (2000) seminal work, researchers have been actively applying the factorization framework to various non-rigid problems. However, they soon noticed that, different from its rigid counterpart, the non-rigid factorization problem appeared to be much more difficult. In a 2004 paper, Xiao et al. (2004) proved that the problem itself is indeed ill-posed or under-constrained, in the sense that, based on the orthonormality constraint alone, one cannot recover the non-rigid shape bases and the corresponding shape coefficients *uniquely*. There is always a fundamental ambiguity between the shape bases and the shape coefficients.

To resolve this ambiguity, Xiao et al. (2004) suggested to add extraneous “basis constraints” so as to make the system well-constrained. In the same spirit of adding extra priors to regularize an otherwise under-constrained problem, Torresani et al. (2008) introduced Gaussian prior on the shape coefficients. Del Bue (2008) introduced special shape priors. Akhter et al. (2008) proposed to use a fixed set of Discrete Cosine Transform (DCT) bases in the trajectory dual space rather than the shape space. Recently, Akhter et al. (2012) further extended this method to temporal and spatial bilinear bases. DCT based method has also been extended to trajectory reconstruction problem (Park et al. 2010) and multi static cameras case (Zaheer et al. 2011; Angst and Pollefeys 2012). Temporally smooth deformation prior has also been used such as in Bartoli et al. (2008) and Aanaes and Kahl (2002). Other priors imposed speciality on the model such as assuming a quadratic model (Fayad et al. 2010), local-rigidity (Taylor et al. 2010), or extend to non-linear models (Rabaud and Belongie 2009; Gotardo and Martinez 2011b). Valmadre and Lucey (2012) introduced general trajectory priors for non-rigid reconstruction. Angst and Pollefeys (2012) provided a unified point of view for NRSfM factorization through modelling with tensor algebra. They also extended the tensor formulation to the case of a camera network where multiple static affine cameras observe the same deforming and moving non-rigid object.

Akhter et al. (2009) made an important theoretical progress, which reveals that: although the ambiguity in the shape bases is inherent, the 3D shape itself can be recovered uniquely without ambiguity. In a slightly earlier paper, Hartley and Vidal proved a similar result under perspective camera model (Hartley and Vidal 2008). Despite the significance of these theoretical results, neither paper has provided a practical algorithm to the problem, and our work aims to fill this gap.

In our implementations, we utilize the nuclear norm (or trace norm) heuristics—a heuristics widely used for rank minimization in the field of Compressive Sensing (CS) (Candès and Plan 2010) and has been widely used in computer vision and machine learning. Potentially, it allows our method to be further improved, e.g. to run much faster or to tackle challenging cases such as cases with missing data or outliers. Note that rank minimization formulation have also been applied to the rigid structure-from-motion problem (Dai et al. 2010; Angst et al. 2011).

Part of the work was published as a conference version in CVPR 2012 (Dai et al. 2012). However, this journal version substantively extends the conference version not only in deep theoretical analysis, but also in experiments.

## 1.2 Organization

In Sect. 2, we formulate the non-rigid structure from motion problem under the low-order linear combination model with discussion on the orthonormality constraint and inherent basis ambiguity. In Sect. 3, we state the main theory for recovering the Euclidean corrective matrix, i.e., null-space representation and the intersection theorem. Section 4 presents our algorithm solution for motion and shape recovery, which consists of recovering the Euclidean corrective matrix, recovering the camera motion and recovering the non-rigid shape in sequel. Based on recent progress in affine constrained rank minimization, Sect. 5 gives theoretical analysis to these formulations and solutions, i.e., uniqueness, relaxation gap and degenerate cases. Section 6 presents the experimental results with corresponding analysis while Sect. 7 closes the paper with future work.

## 2 Problem Statement

### 2.1 Formulation

The task of non-rigid factorization is to factorize an *image measurement matrix*  $W$  as the product of *camera motion* (projection) matrix  $M$  and a non-rigid shape matrix  $S$ , such that  $W = MS$ . We assume the measurement matrix is already centralized, therefore the camera motions reduce to pure rotations (Bregler et al. 2000).

Let us consider a monocular camera (moving or static) observing a non-rigid shape containing  $P$  3D-points by  $F$  frames, and denote  $S_{ij}$  as the  $j$ th 3D shape point in the  $i$ th frame,  $\mathbf{x}_{ij} = [u_{ij}, v_{ij}]^T$  as the image point of  $S_{ij}$ . Then

$$\mathbf{x}_{ij} = R_i S_{ij}, \quad (1)$$

where  $R_i \in \mathbb{R}^{2 \times 3}$  stands for the first two rows of the  $i$ th camera rotation, hence  $R_i R_i^T = I_2$ , i.e.,  $R_i^T$  is a Stiefel matrix (Paladini et al. 2009).

Based on the low-order linear combination model, the non-rigid shape  $S_i \in \mathbb{R}^{3 \times P}$  can be represented as a linear combination of  $K$  *shape bases*  $B_k \in \mathbb{R}^{3 \times P}$  with time-varying *shape coefficients*  $c_{ik}$  as:  $S_i = \sum_{k=1}^K c_{ik} B_k$ . Under the orthographic camera model, the coordinates of the 2D image points observed at frame  $i$  are given by:  $W_i = R_i S_i$ , where  $W_i = [\mathbf{x}_{i1} \dots \mathbf{x}_{iP}]$ . Use this representation, and stack all the  $F$  frames of measurements and all the  $P$  points in a matrix form, we obtain:

$$\begin{aligned} W &= \begin{bmatrix} \mathbf{x}_{11} & \dots & \mathbf{x}_{1P} \\ \vdots & & \vdots \\ \mathbf{x}_{F1} & \dots & \mathbf{x}_{FP} \end{bmatrix} = \begin{bmatrix} R_1 S_1 \\ \vdots \\ R_F S_F \end{bmatrix} \\ &= \begin{bmatrix} c_{11} R_1 & \dots & c_{1K} R_1 \\ \vdots & \ddots & \vdots \\ c_{F1} R_F & \dots & c_{FK} R_F \end{bmatrix} \begin{bmatrix} B_1 \\ \vdots \\ B_K \end{bmatrix} \\ &= R(C \otimes I_3)B \doteq \Pi B. \end{aligned} \quad (2)$$

In this formula, we call  $R = \text{blkdiag}(R_1, \dots, R_F) \in \mathbb{R}^{2F \times 3F}$  the camera motion (rotation) matrix. Since  $\Pi \in \mathbb{R}^{2F \times 3K}$  and  $B \in \mathbb{R}^{3K \times P}$ , it is easy to see:  $\text{rank}(W) \leq \min(\text{rank}(\Pi), \text{rank}(B)) \leq 3K$ . In general case (i.e. non degenerate basis, noise free), the rank of  $W$  is  $\text{rank}(W) = 3K$ . Additionally, the non-rigid shape matrix  $S = (C \otimes I_3)B$  is low rank too, as  $\text{rank}(S) \leq \min(\text{rank}(C \otimes I_3), \text{rank}(B)) \leq 3K$ . Notice that  $S$  is not a general rank  $3K$  matrix but owns its special structure due to its  $C \otimes I_3$  factor.

### 2.2 Orthonormality Constraint

From a measurement matrix  $W$  one can compute its rank  $3K$  approximation  $W = \hat{\Pi} \hat{B}$  via SVD (singular value decomposition). However, this decomposition is not unique as any nonsingular matrix  $G \in \mathbb{R}^{3K \times 3K}$  can be inserted between  $\hat{\Pi}$  and  $\hat{B}$  to obtain a new valid factorization as  $W = \hat{\Pi} \hat{B} = \hat{\Pi} G G^{-1} \hat{B} = \Pi B$ .

A particular matrix  $G$  that rectifies  $\hat{\Pi}$  to be a canonical Euclidean form is called the *Euclidean corrective matrix*, because once such a  $G$  is determined, one obtains  $R(C \otimes I_3) = \hat{\Pi} G$  and the shape bases  $B = G^{-1} \hat{B}$ .

Denote the  $i$ th double rows of  $\hat{\Pi}$  as  $\hat{\Pi}_{2i-1:2i} \in \mathbb{R}^{2 \times 3K}$ , and the  $k$ th column-triplet of  $G$  as  $G_k \in \mathbb{R}^{3K \times 3}$ , we have:

$$\hat{\Pi}_{2i-1:2i} G_k = c_{ik} R_i, \quad i = 1, \dots, F, k = 1, \dots, K. \quad (3)$$

Orthonormality constraints (i.e. rotation constraints) in the  $\Pi$  matrix can be imposed to recover a Gram matrix  $Q_k \in \mathbb{R}^{3K \times 3K}$  formed by  $Q_k = G_k G_k^T$  as  $\hat{\Pi}_{2i-1:2i} Q_k \hat{\Pi}_{2i-1:2i}^T = c_{ik}^2 I_2$ . Since  $c_{ik}$  is not known, one can only establish two linear equations over  $Q_k$  as:

$$\begin{aligned} \hat{\Pi}_{2i-1} Q_k \hat{\Pi}_{2i-1}^T &= \hat{\Pi}_{2i} Q_k \hat{\Pi}_{2i}^T, \\ \hat{\Pi}_{2i-1} Q_k \hat{\Pi}_{2i}^T &= 0. \end{aligned} \quad (4)$$

### 2.3 Inherent Ambiguity

In doing the above non-rigid factorization, Xiao et al. (2004) discovered that, the solutions are however fundamentally ambiguous, in the sense that one cannot expect to find the shape bases and shape coefficients uniquely. Such an inherent ambiguity largely explains why non-rigid factorization is fundamentally more difficult than its rigid counterpart.

Later, Akhter et al. (2009) showed that, quite surprisingly, the fundamental ambiguity does not necessarily lead to an ambiguous shape. They further proved that using the orthogonality constraints alone is in fact sufficient to recover a unique (unambiguous) non-rigid shape, (provided that a previously-overlooked rank 3 constraint on  $Q_k$  (Eq. (4)) is accounted for). However, apart from its evident theoretical value, their paper did not propose any optimization algorithm (other than a local search algorithm due to Brand (2005)) to efficiently find the correct  $G_k$ . Instead, the authors argued that “the real difficulty in achieving good 3D reconstructions for non-rigid structures ... is not the ambiguity of the [basis] constraints, but the complexity of the underlying non-linear optimization”. In this paper, we will challenge this argument, by providing a simple yet efficient (optimization) solution to the NRSfM factorization.

### 3 Main Theory

In this section and the following section, we will present our main theory and solution for solving the NRSfM factorization problem in a prior-free way. First, we prove that the Euclidean corrective matrix must lie in the intersection of a linear subspace and positive semi-definite low-rank cone. Second, based on these theoretical results, we present our solution in sequel, which includes estimating the Euclidean corrective matrix, recovering the camera rotations and recovering the non-rigid shapes. In contrast to existing methods which mainly rely on the explicit basis representations, our method solves the camera motion and non-rigid shape directly without explicit parameterization with basis matrix and coefficients, thus without dealing with the basis ambiguity.

From now on, we assume the measurement matrix  $W$  is already truncated to rank  $3K$  (by e.g. SVD); the number of shape bases  $K$  has been estimated; and all the shape bases are sufficiently generic and non-degenerate. We start with a known result of NRSfM factorization.

**Theorem 1** *All the solutions  $Q_k$  to the linear system Eq. (4) form a linear subspace of dimensionality  $(2K^2 - K)$ .*

This result is a direct consequence of Xiao et al.’s Theorem in Xiao et al. (2004). It shows that the above linear system is inherently under-determined (as by  $(2K^2 - K)$  rank deficient), no matter how many frames are given. Note that when

$K = 1$  (rigid case), the dimension of the solution space is 1, suggesting that solving rigid problem is not ambiguous at all.

On the other hand, this result also provides us with the true dimensionality of the solution space of  $Q_k$ , and note that  $Q_k$  is precisely what we are after. However, in their paper, this practical implication has not been explicitly exploited.

In the following, we will show how one can take advantage of this result, and derive a practical algorithm that directly leads to a parameterization of this solution space. More precisely, we will prove that the solution space of  $Q_k$  is actually the *null-space* of a certain matrix  $A$  which can be directly obtained from the input image measurements. Practical usefulness of this representation is obvious.

#### 3.1 Null-Space Representation

First, denote  $\text{vec}()$  as the vectorization operator, and  $\mathbf{q}_k = \text{vec}(Q_k)$ . Using  $\text{vec}(AXB^T) = (B \otimes A)\text{vec}(X)$ , we rewrite the linear system Eq. (4) as:

$$\begin{bmatrix} \hat{\Pi}_{2i-1} \otimes \hat{\Pi}_{2i-1} - \hat{\Pi}_{2i} \otimes \hat{\Pi}_{2i} \\ \hat{\Pi}_{2i-1} \otimes \hat{\Pi}_{2i} \end{bmatrix} \mathbf{q}_k \doteq A_i \mathbf{q}_k = \mathbf{0}, \quad (5)$$

where  $\hat{\Pi}_{2i-1}$ ,  $\hat{\Pi}_{2i}$  denote the  $2i - 1$ th and  $2i$ th row of  $\hat{\Pi}$ .

Stacking all such equations from all frames ( $i = 1, \dots, F$ ), we then have

$$A \mathbf{q}_k = \mathbf{0}, \quad (6)$$

where  $A = [A_1^T, A_2^T, \dots, A_F^T]^T$ . This is a linear system of equations over the unknown  $9K^2$ -vector  $\mathbf{q}_k$ .

Note that  $\mathbf{q}_k$  has  $(3K)(3K + 1)/2$  independent entries. It may appear that, given enough frames, i.e. when  $2F \geq (3K)(3K + 1)/2$ ,  $\mathbf{q}_k$  should be able to compute via linear least squares. However, this is not the case, because all valid solutions reside in a  $2K^2 - K$  dimensional space as shown in Theorem 1. Moreover, Eq. (6) shows, the solution space is nothing but the null-space of  $A$ . In addition, it is easy to verify that the minimum required number of frames for computing the null-space linearly is  $F \geq (5K^2 + 5K)/4$ .

Finally, we stress that, since matrix  $A$  can be readily solved from the input measurements, finding a set of orthonormal bases that span the null-space is simple (e.g. by SVD of  $A$ ). We denote the basis of the null space spanned by  $[\phi_1, \phi_2, \dots, \phi_{2K^2-K}]$ . This provides an explicit parameterization of the solution space. Specifically, we have reached:

**Theorem 2** *All the valid solutions  $\mathbf{q}_k = \text{vec}(Q_k)$  to the linear system Eq. (6) can be parameterized as  $\mathbf{q}_k = \sum_{l=1}^{2K^2-K} \alpha_l \phi_l$ .*



### 3.2 The Intersection Theorem

Combining all the proceeding results, we now arrive at the central theorem of this paper:

**Theorem 3** (Intersection theorem) *Under non-degenerate and noise-free conditions, any correct solution of  $Q_k$  (i.e. the Gram matrix of a column-triplet of the Euclidean corrective matrix  $G_k$ ) must lie in the intersection of the  $(2K^2 - K)$ -dimensional null-space of  $A$  and a rank 3 positive semi-definite matrix cone, i.e.,  $Q_k$  belongs to*

$$\{A \text{ vec}(Q_k) = \mathbf{0}\} \cap \{Q_k \succeq 0\} \cap \{\text{rank}(Q_k) = 3\}. \quad (7)$$

*Proof* Denote  $G_k$  as column triplet of a corrective rectifying transform, and  $Q_k = G_k G_k^T$ , then  $\text{rank}(Q_k) = \text{rank}(G_k) = 3$ ,  $Q_k \succeq 0$ . Additionally,  $\text{vec}(Q_k)$  lies in the null space of  $A$  as  $\text{vec}(Q_k)$  gives rectification with zero error. Thus  $G_k$  is a solution to the equation system which means that the equation system is well-defined.

Denote  $\tilde{G}_k$  as solution to the equation system Eq. (7), then  $\text{rank}(\tilde{Q}_k) = 3$ ,  $\tilde{Q}_k \succeq 0$  and  $\text{vec}(\tilde{Q}_k)$  lies in the null space of the system of linear constraints on the elements  $\tilde{Q}_k$ . Thus all the solutions to the equation system satisfy the condition for rectifying transforms and achieve perfect structure recovery. There is no difference between these solutions.

## 4 Solution

Armed with the above results (in particular Theorem 3), we are now ready to present our simple prior-free method to the NRSfM factorization problem.

Recall that our goal is to recover both the camera motion matrix  $R$  and the non-rigid shape matrix  $S$  from image measurement matrix  $W$ , such that  $W = RS = R(C \otimes I_3)B$ . Note that due to the inherent basis ambiguity, it is hopeless to recover a unique  $B$  or  $C$ . While in previous work many researchers chose to use a pre-defined special shape bases  $B$  or trajectory bases in the dual trajectory space (or enforce arbitrary priors on the shape bases or shape coefficients,) to pin down the undetermined degrees-of-freedom, in this work we will show how one can directly estimate  $S$  without fixing  $B$  or  $C$ .

Our method consists of three steps to be applied in sequel: (1) Estimate the (Gram of) Euclidean corrective matrix  $G_k$ , (2) Estimate the camera rotations  $R$  and (3) Estimate the non-rigid shape  $S$ .

### 4.1 Step 1: Estimate $G_k$ by Trace-Norm Minimization

Our main intersection theorem (Theorem 3) naturally leads to a simple algorithm to solve for  $G_k$ , that is: to find the intersection of the aforementioned null-space and the rank 3 positive semi-definite matrix cone.

Because the rank-function itself is not very numerically-stably, measurements noise will increase the *numerical rank* of  $Q_k$  dramatically, we slightly relax the  $\text{rank}(Q_k) = 3$  condition to a *rank-minimization* problem, i.e.  $\min \text{rank}(Q_k)$  (Under noise-less condition, these two conditions are equivalent to each other with zero relaxation gap). Note that however, rank-minimization is an NP-hard problem in general, and is very difficult to solve exactly. We therefore further relax it to a nuclear-norm minimization form, i.e.,  $\min \|Q_k\|_*$ . The nuclear norm of a matrix is the tightest convex bound of its rank, and for a positive semi-definite matrix, its nuclear norm is reduced to its trace norm (Recht et al. 2010; Angst et al. 2011). This is the case for our problem in question, because  $Q_k$  is a positive semi-definite matrix. Thus we have  $\|Q_k\|_* = \text{trace}(Q_k)$ .

Then we arrive at the following trace norm minimization formulation to solve the Gram of Euclidean corrective matrix  $Q_k$ .

$$\begin{aligned} &\min \text{trace}(Q_k), \text{ such that,} \\ &Q_k \succeq 0, \\ &A \text{ vec}(Q_k) = \mathbf{0}. \end{aligned} \quad (8)$$

To avoid the trivial solution  $Q_k = 0$ , we express  $\text{vec}(Q_k)$  in explicit form of the null space representation thus excluding the case of having all-zero weights.<sup>1</sup>

It is easy to recognize that the above trace norm minimization problem is a standard *semi-definite programming (SDP)*. Also note that this SDP is actually very small, and of a fixed size of  $2K^2 - K$  which is independent of the size of the measurement matrix. Thus this SDP problem can be solved very efficiently by any off-the-shelf SDP solver. Once  $Q_k$  is found, we use SVD to extract an exact rank 3  $G_k$ . This solved  $G_k$  can be directly used to find  $R$  and then  $S$ . Alternatively, if higher accuracy is desired, one can further improve the numerical accuracy of  $G_k$  by feeding it as an initial point to a non-linear refinement procedure, such as via the following unconstrained minimization:

$$\min \sum_{i=1}^F \left[ \left( 1 - \frac{\hat{\Pi}_{2i} G_k G_k^T \hat{\Pi}_{2i}^T}{\hat{\Pi}_{2i-1} G_k G_k^T \hat{\Pi}_{2i-1}^T} \right)^2 + \left( 2 \frac{\hat{\Pi}_{2i-1} G_k G_k^T \hat{\Pi}_{2i}^T}{\hat{\Pi}_{2i-1} G_k G_k^T \hat{\Pi}_{2i-1}^T} \right)^2 \right],$$

where the objective function is nothing but the orthonormality condition and this refinement is similar to a bundle-adjustment (Triggs et al. 2000) process.

<sup>1</sup> We explicitly express the solution as a linear combination of  $2K^2 - K$  basis vectors of the null space, then use sum-to-one constraint to fix the scale freedom, i.e.,  $\sum_{l=1}^{2K^2-K} \alpha_l = 1$ , where  $\alpha_l$  are coefficients and  $\phi_l$  are the bases of the null space of  $A$ . See Theorem 2 for detailed explanation.

## 4.2 Step 2: Compute Rotation Matrix $R$

Conventionally, once  $G_k \in \mathbb{R}^{3K \times 3}$  is solved (w.l.o.g., let's denote it as  $G_1$ ), which is merely a single column-triplet in the full Euclidean corrective matrix  $G \in \mathbb{R}^{3K \times 3K}$ , the commonly used next step is to solve for the other  $K - 1$  column-triplets  $[G_2, \dots, G_K]$  and use them to populate the entire matrix  $G$ . Brand (2005) proposed a linear method to solve for the big  $G$  (for affine case). Because these  $G_k$ s always have rotation ambiguity, in order to align them, Procrustes method must be employed subsequently (c.f. Xiao et al. 2004; Akhter et al. 2009). Once the big  $G$  is obtained, one then is allowed to compute the camera motion  $R$ , the shape coefficients matrix  $C$  and the shape bases matrix  $B$ , and then reconstruct the non-rigid shape  $S$ . However, the above Procrustes-based approach is not only rather involved, but also not numerically stable. More importantly, it is *not necessary*, as shown in Akhter et al. (2008).

In this work, we adopt a simpler approach that directly computing the camera motion  $R$  (and *directly* computes the non-rigid shape  $S$ ) from a single column-triplet  $G_k$ , without the need to fill in a big and full  $G$  matrix. The method goes as follows. Once  $G_k$  is solved, the rotation at every frame  $i = 1, \dots, F$  can be solved by using:

$$\hat{\Pi}_{2i-1:2i} G_k = c_{ik} R_i, \quad i = 1, \dots, F. \quad (9)$$

Note that we need not care about the unknown value of  $c_{ik}$ . However, there is a sign ambiguity in the obtained rotation. Assume the estimated rotation at the  $f$ th frame as  $R_f$ , denote the unit vector orthogonal to  $R_f$  as  $\mathbf{r}_f$ , then both  $[R_f; \mathbf{r}_f^T]$  and  $[-R_f; \mathbf{r}_f^T]$  are valid rotation. Different sign selection for rotations results in recovered shape with cheirality (Hartley and Zisserman 2004), i.e.,  $S_f = [S_{fx}; S_{fy}; S_{fz}]$  and  $S_f = [-S_{fx}; -S_{fy}; S_{fz}]$ . This ambiguity can be fixed via constraining the rotation between any two consecutive frames not more than  $\pm 90^\circ$  (Akhter et al. 2009). Finally, the full motion matrix  $R$  is formed as  $R = \text{blkdiag}([R_1, R_2, \dots, R_F])$ .

## 4.3 Step 3: Estimate $S$ by Rank-Minimization

Now we show how to solve the non-rigid shape matrix  $S$ . Most conventional methods do this *indirectly*, in the sense that they often start from solving the big  $G$  matrix firstly, and then use pre-selected special shape bases  $B$  (such as the first  $K$  frame (Xiao et al. 2004), or DCT bases in the dual trajectory space (Akhter et al. 2008), or assume the the shape coefficients are also DCT-expandable (Gotardo and Martinez 2011a)), then the corresponding coefficient matrix  $C$  can be determined, and also the shape matrix  $S = (C \otimes I_3)B$ .

In the following subsections, we will provide simpler, more direct methods for solving  $S$ .

### 4.3.1 Pseudo Inverse Method

Recall that our goal is to solve  $S$  through the equation of  $W = RS = R(C \otimes I_3)B$  given  $W$  and  $R$ . This equation is under-determined because  $R$  is a short matrix of size  $2F \times 3F$ . There should be no unique but an infinite family of solutions of  $S$ . However, we also notice that: the low-order linear combination model, i.e.  $S = (C \otimes I_3)B$  immediately suggests that  $\text{rank}(S) \leq 3K$ .

However, special care has to be taken here that  $S$  is not a general rank  $3K$  matrix but owns its special structure as  $S = (C \otimes I_3)B$ . Here we relax the constraint and solve for general rank  $3K$  matrix first.

Taking into account of both the above arguments, we reach: a valid *relaxed* solution to the shape matrix  $S$  must lie in the intersection of a low rank matrix set  $\{\text{rank}(S) \leq 3K\}$  and the solution space of the linear equation  $W = RS$ . As usual we equally transform the low-rank condition to rank-minimization formulation (under noise-less and non-degenerate case). Now the shape matrix  $S$  must be a solution to the following rank minimization problem:

$$\begin{aligned} &\min \text{rank}(S), \text{ such that,} \\ &W = RS. \end{aligned} \quad (10)$$

This rank minimization formulation can be relaxed to nuclear norm minimization as:

$$\begin{aligned} &\min \|S\|_*, \text{ such that,} \\ &W = RS. \end{aligned} \quad (11)$$

**Remarks.** We now make two important remarks: (1) The above rank minimization problem (in the context of NRSfM) accepts the ground truth shape matrix  $S$  as a solution; (2) Computationally, the above solution may be solved by the (unique) Moore–Penrose pseudo inverse, i.e.,  $S^* = R^\dagger W = R^T W$ .

**Remark 1** is simply the main conclusion of Akhter et al. (2009), which states that: once  $R$  is fixed, there is no ambiguity in finding a low rank shape matrix  $S$  (assuming the shape is sufficiently non-degenerate).

**Remark 2** was a bit surprising to the authors, as it seems to suggest a simple linear and closed-form solution to our original rank-minimization problem of (11)-which is NP-hard to solve in general.

Fortunately, recent progress in Compressive Sensing has confirmed the correctness of our Remark 2. In particular, we use the following result due to Liu et al. (2013): *the Moore–Penrose pseudo-inverse solution  $S = R^\dagger W$  is the **unique** minimizer of the nuclear norm minimization problem (11), and is also **one** minimizer of the rank minimization problem (10).* The reader is referred to Liu et al. (2013) for a detailed proof. From this result, we see that the pseudo-inverse solution is

indeed a possible solution that satisfies two necessary conditions (i.e., satisfying both the imaging equation and *relaxed* low rank conditions), though care must be given when discussing the equality of the approximation (i.e., relaxation gap) as well as the existence and uniqueness (multiplicity) issue associated with both the original rank minimization problem and the relaxed nuclear norm minimization problem (to be presented in Sect. 5.3).

We have numerically tested the pseudo-inverse method on both synthetic data and real data. Judging only from the numerical result in terms of 3D shape-recovery accuracy, the pseudo inverse solution outperforms Xiao et al.'s shape basis method by a large margin, and achieves comparable performance with metric-projection (Paladini et al. 2009) and EM-PPCA (Torresani et al. 2008). The reader is referred to the second last column of Table 2. Note that however, these methods (including our pseudo inverse method) are still inferior to the more recent DCT trajectory basis method (Akhter et al. 2008) and the CSF method (Gotardo and Martinez 2011a) (both methods rely on strong smoothness prior).

This inferiority prompted the authors to think further: perhaps the rank  $3K$  condition enforced upon  $S$  is not strong enough or sufficient? Pondering this question has led us to the following “block matrix method” which gives more favorable performance and theoretical guaranty.

#### 4.3.2 Block Matrix Method

In the pseudo-inverse method, we mainly make use of the rank  $3K$  condition on deformable shape  $S$  that is  $\text{rank}(S) \leq 3K$ . This  $3F \times P$  matrix  $S$  is simply a stack of  $P$  3D points  $[X_i, Y_i, Z_i]^T$  over  $F$  frames.

However, we realize that, since in reality there are *in fact* only  $K$  shape bases (rather than  $3K$ ), the shape matrix  $S$  is not a fully-generic rank  $3K$  matrix, but has its special block structure indicated as  $S = (C \otimes I_3)B$ . Ignoring this special structure will add spurious degrees-of-freedom (hence ambiguities) to the problem. Next we will show how to get a stronger (yet meaningful) rank-minimization formulation, considering the nature of the NRSfM factorization.

In particular, if we re-arrange the rows of  $S$  that correspond to  $X$ ,  $Y$ , and  $Z$  coordinate separately, and in an  $F \times 3P$  block matrix form, denoted by  $S^\sharp$  in below<sup>2</sup>:

$$S^\sharp = \begin{bmatrix} X_{11} & \dots & X_{1P} & Y_{11} & \dots & Y_{1P} & Z_{11} & \dots & Z_{1P} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ X_{F1} & \dots & X_{FP} & Y_{F1} & \dots & Y_{FP} & Z_{F1} & \dots & Z_{FP} \end{bmatrix}.$$

<sup>2</sup> An identical rearrangement was used in Akhter et al. (2008) but with different motivation and purpose.

Notice that under the linear combination model, the non-rigid shape is expressed as  $S_i = \sum_{k=1}^K c_{ik} B_k$ , where

$$S_i = \begin{bmatrix} X_{i1} & X_{i2} & \dots & X_{iP} \\ Y_{i1} & Y_{i2} & \dots & Y_{iP} \\ Z_{i1} & Z_{i2} & \dots & Z_{iP} \end{bmatrix}.$$

Thus, it is easy to see that the low order constraint in the linear combination model has been *equivalently* expressed in the low rank condition on  $S^\sharp$  rather than the low rank condition on  $S$ . Then we must have:  $\text{rank}(S^\sharp) \leq K$ . Since the row rank and column rank of a matrix are equal, this condition shows that non-rigid shape has a low rank property (rank  $K$ ) in both shape space and trajectory space. Note that this rank  $K$  condition on  $S^\sharp$  is much stronger than the above rank  $3K$  condition on  $S$  (A rank  $K$  condition on  $S^\sharp$  guarantees a rank  $3K$  condition on  $S$ , however the inverse conclusion does not exist in general.). The form of  $S^\sharp$  captures the essence of the  $K$ -order linear combination model.

By defining proper row-selection matrices  $P_X, P_Y, P_Z \in \mathbb{R}^{F \times 3F}$  (of 0–1 values, similar to a row-permutation matrix<sup>3</sup>), we can compactly represent the relationship between  $S^\sharp$  and  $S$  as:

$$S^\sharp = [P_X \ P_Y \ P_Z](I_3 \otimes S). \quad (12)$$

Now, to solve for this re-arranged block shape matrix  $S^\sharp$ , we use, again, a rank-minimization formulation:

$$\begin{aligned} & \min \text{rank}(S^\sharp), \text{ such that,} \\ & W = RS, \\ & S^\sharp = [P_X \ P_Y \ P_Z](I_3 \otimes S). \end{aligned} \quad (13)$$

We relax the above rank minimization model with the nuclear norm heuristics, i.e.,  $\min \|S^\sharp\|_*$ , thus reach the following nuclear norm minimization model:

$$\begin{aligned} & \min \|S^\sharp\|_*, \text{ such that,} \\ & W = RS, \\ & S^\sharp = [P_X \ P_Y \ P_Z](I_3 \otimes S). \end{aligned} \quad (14)$$

In principle, this nuclear norm minimization problem may be solved by a standard SDP solver such as Sedumi (Sturm 1999) and SDPT3 (Toh et al. 1999). However, unlike the case of Sect. 4.1 where the resulted-SDP has a small and fixed size, here this SDP is of size  $F \times 3P$ , which renders an off-the-shelf SDP solver very inefficient when either  $P$  or  $F$  is large.

<sup>3</sup>  $P_X(i, 3i - 2) = 1, P_Y(i, 3i - 1) = 1, P_Z(i, 3i) = 1$ , while all the other positions being zero.

Below, we give an efficient approximate numerical implementation, based on the *fixed point continuation* algorithm (Ma et al. 2011).

**4.3.2.1 Fast Numerical Implementation** We first rewrite the above minimization Eq. (14) in its Lagrange multiplier form as:

$$\min_{\mu} \mu \|S^{\sharp}\|_* + \frac{1}{2} \|W - RS\|_F^2, \text{ such that,} \\ S^{\sharp} = [P_X \ P_Y \ P_Z](I_3 \otimes S), \quad (15)$$

where  $\mu$  is the continuation (homotopy) parameter which diminishes as the algorithm iterates.

Next, the gradient of  $\frac{1}{2} \|W - RS\|_F^2$  with respect to  $S^{\sharp}$  is obtained as:

$$g(S^{\sharp}) = \frac{\partial \frac{1}{2} \|W - RS\|_F^2}{\partial S^{\sharp}} \\ = [P_X \ P_Y \ P_Z](I_3 \otimes (R^T(RS - W))). \quad (16)$$

Then, we solve the minimization of Eq. (15) via the following two-line iteration update (c.f. Ma et al. 2011):

$$\begin{cases} Y^{(k)} = S^{\sharp(k)} - \tau g(S^{\sharp(k)}), \\ S^{\sharp(k+1)} = S_{\tau\mu}(Y^{(k)}), \end{cases} \quad (17)$$

where  $\tau$  is the step size of gradient descent, and  $S_v(\cdot)$  is the *matrix shrinkage* operator<sup>4</sup>(c.f. Ma et al. 2011). Once we have recovered the re-arranged shape  $S^{\sharp}$ , we first project it to the nearest rank  $K$  matrix (note: not  $3K$ ) via SVD, then rearrange it to  $S$ .

Thus we propose the fixed point continuation based algorithm for non-rigid shape recovery in Algorithm 1, where the initial  $S^{\sharp(0)}$  is initialized by the pseudo-inverse method.

Convergence analysis of fixed point continuation algorithm for general matrix rank minimization has been given in Goldfarb and Ma (2011). Currently, the main computation takes place in the singular value shrinkage (SVT), where SVD is utilized during each iteration. Recent progress in compressive sensing and matrix completion such as SVT without SVD (Cai and Osher 2010) and alternating direction method (ADM) (Lin et al. 2010; Yang and Yuan 2013) can be used to further speed up the implementation.

<sup>4</sup> **Matrix Shrinkage Operator:** Assume  $X \in \mathbb{R}^{m \times n}$  and the SVD of  $X$  is given by  $X = U \text{Diag}(\sigma) V^T$ ,  $U \in \mathbb{R}^{m \times r}$ ,  $\sigma \in \mathbb{R}_+^r$ ,  $V \in \mathbb{R}^{n \times r}$ . For any  $v > 0$ , the matrix shrinkage operator  $S_v(\cdot)$  is defined as  $S_v(X) := U \text{Diag}(s_v(\sigma)) V^T$ , where  $s_v(\sigma)$  is defined as:

$$s_v(\sigma) := \bar{\sigma}, \text{ with } \bar{\sigma}_i = \begin{cases} \sigma_i - v, & \text{if } \sigma_i - v > 0, \\ 0, & \text{otherwise.} \end{cases}$$

```
1: Initialization: Given  $S^{\sharp(0)}$ , select  $\bar{\mu} > 0$  and  $\tau > 0$ . Fix the
   sequence of  $\mu_l$  as  $\mu_{l+1} = \max\{\mu_l \eta_{\mu}, \bar{\mu}\}$ ,  $l = 1, 2, \dots, L-1$ .
2: for  $\mu = \mu_1, \mu_2, \dots, \mu_L$  do
3:   while Not converged do
4:     Compute  $Y^{(k)} = S^{\sharp(k)} - \tau g(S^{\sharp(k)})$ ;
5:     Update  $S^{\sharp(k+1)} = S_{\tau\mu}(Y^{(k)})$ ;
6:     Compute stopping criteria  $\varepsilon^{(k+1)} = \frac{\|S^{\sharp(k+1)} - S^{\sharp(k)}\|_F}{\|S^{\sharp(k)}\|_F}$ .
7:     if  $\varepsilon^{(k+1)} < \epsilon$  then
8:       Inner iteration converged;
9:     end if
10:   end while
11: end for
```

**Algorithm 1:** Fixed point continuation (FPC) for non-rigid shape recovery in the block matrix method

### 4.3.3 Closed-Form Solution with Smooth Deformation Constraint

In this subsection, we introduce a modified version of our pseudo-inverse method by incorporating the smoothness prior. Although this topic is deviated from the main thesis of prior-free of the present paper, our intention is however to illustrate how sensible priors—if available—can be applied to further improve the quality of the solution.

By introducing smooth deformation prior,<sup>5</sup> we can formulate the non-rigid shape recovery problem as minimizing a data term evaluated on the image measurements and a regularization term (smoothness), thus reaching the following formulation:

$$\min_S \frac{1}{2} \|W - RS\|_F^2 + \frac{1}{2} \lambda \|HS\|_F^2, \quad (18)$$

where the first term measures the reconstruction error of the recovered motion and non-rigid shape while the second term measures the smoothness constraint.  $H$  can be chose as different pattern to characterize various kinds of smoothness, e.g. first order smoothness as in Eq. (19), second order smoothness, etc.

$$H_{ij} = \begin{cases} 1, & j = i, i = 1, \dots, 3(F-1), \\ -1, & j = i + 3, i = 1, \dots, 3(F-1), \\ 0, & \text{Otherwise.} \end{cases} \quad (19)$$

The optimization problem in Eq. (18) admits an analytical (closed-form) solution too, as shown below. First, we take the derivative of Eq. (18) with respect to  $S$  and set it to zero:

$$(R^T R + \lambda H^T H)S = R^T W, \quad (20)$$

<sup>5</sup> Smooth deformation prior has been used in Aanaes and Kahl (2002), Olsen and Bartoli (2008), Valmadre and Lucey (2012), etc.



which has a closed-form solution as:

$$S = (R^T R + \lambda H^T H)^{\dagger} R^T W. \quad (21)$$

The rotation matrix  $R$  is of row full rank  $2F$  thus  $R^T R$  is of rank  $2F$  generally. The smoothness matrix  $H$  is rank deficient too, thus  $H^T H$  is of rank  $3F - 3$  (for first order smoothness). In general case,  $R^T R + \lambda H^T H$  is full rank, thus invertible. We can replace the pseudo-inverse operation with matrix inverse, thus speeding up the implementation (matrix inverse is much easier and efficient than matrix pseudo-inverse).

There is a trade-off parameter  $\lambda$  in the closed-form solution. It trades off between reconstruction ability and smoothness. When  $\lambda$  approaches 0, the solution approaches  $R^{\dagger} W$ , i.e. the pseudo-inverse solution. When  $\lambda$  is large enough, the solution approaches a rigid shape, which minimizes the combination of  $\|HS\|_F^2$  and  $\|W - RS\|_F^2$ . When  $\lambda$  approaches  $+\infty$ , the solution approaches trivial solution  $S = 0$ , which will make  $\|HS\|_F^2 \rightarrow 0$ .

The reader may wonder how to select a good trade-off parameter as it trades-off between reconstruction error and smoothness. Here we give some empirical results. On the motion capture dataset “Stretch”, given the ground truth motion  $R$  and noisy measurements  $W$ , we recover the deformable shape with different trade-off parameter  $\lambda$  ranging from  $10^{-15}$  to  $10^5$ . Relationships between the normalized mean 3D error, the smooth cost function  $\|HS\|_F^2/(F-1)$ , the reprojection error on the measurements  $\sqrt{\|W - RS\|_F^2/(FP)}$  and the trade off parameter  $\lambda$  are illustrated in Fig. 1. It is obviously to observe that there is a large interval of  $\lambda$  where the normalized mean 3D error is almost constant which suggests that the normalized mean 3D error is not sensitive the selection of  $\lambda$ . Meanwhile, we confirm that when  $\lambda$  approaching 0, the result approaches the pseudo-inverse solution  $S = R^{\dagger} W$  which consists with the theoretical analysis above.

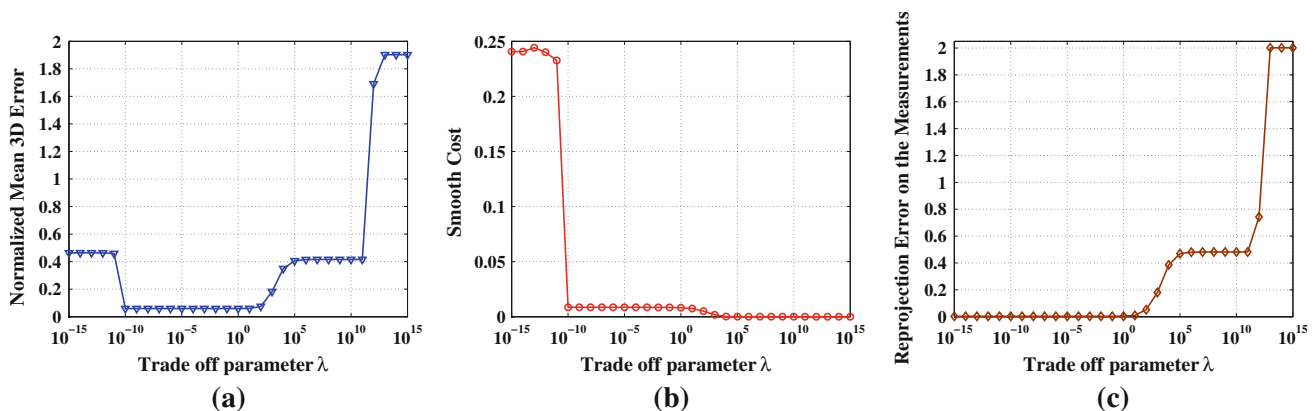
#### 4.3.4 Connecting the Block Matrix Method and the Pseudo-Inverse Method

In this part, we briefly discuss the relationship between our block matrix method and pseudo-inverse method.

First, in this work, we aim at solving the non-rigid structure from motion problem in a prior-free way by capturing the essential low-rank constraint in the linear combination model. Our first attempt was to utilize the low-rank property of the non-rigid shape matrix  $S$  directly. This effort had lead us first to the closed-form pseudo-inverse solution, which itself is an important step which provides new insight to the understanding of the problem. Specifically, by experimenting the pseudo-inverse method we realized that it cannot fully capture the structured low-rank constraint of the NRSfM model. This had motivated us to explore further, and we made the second discovery of the reshuffled shape matrix form  $S^{\sharp}$ . A low rank constraint on  $S^{\sharp}$  captures completely the low-order linear combination model in explaining the image measurements. Therefore, the pseudo-inverse method works as the starting point in our prior-free solution for NRSfM.

Second, currently, the block matrix method for non-rigid shape recovery does not own a closed-form solution as in the pseudo-inverse method. Therefore, we proposed a fixed-point continuation based iterative method to solve it efficiently. The pseudo-inverse solution is utilized as initialization for the block matrix method to achieve accelerated convergence.

Third, in the theoretical analysis section, we will provide further analysis to the pseudo-inverse method as well as the block matrix method. According to the analysis, we note that valid non-rigid shapes lie in an affine subspace, where the pseudo-inverse solution achieves both nuclear norm and Frobenius norm minimization. The pseudo-inverse method explains away the image measurements with minimum Frobenius norm, where its solution appears to be a col-



**Fig. 1** Performance evaluation under varying trade off parameter  $\lambda$  on motion capture sequence Stretch. (a) Normalized mean 3D error; (b) smooth cost function  $\|HS\|_F^2/(F-1)$ ; and (c) reprojection error on the image measurements

lection of coplanar shapes at each frame. This offers deeper insight to the NRSfM problem.

## 5 Theoretical Analysis

In the above sections, we have presented our main theory and numerical implementations of our simple “prior free” method to factorization based NRSfM, where we have applied the nuclear norm heuristics to the rank minimization formulation for motion and shape recovery. Then a theoretical problem naturally arises, i.e., can our formulation solve the original problem exactly and uniquely? In this section, we give further theoretical analysis to the formulation, regarding the uniqueness of the rank minimization, and relaxation gap in nuclear norm heuristics. Finally, we relate our theoretical results with existing work in NRSfM.

We first review the formulation of affine constrained rank minimization and nuclear norm minimization. Then we review theoretical results on null space analysis to affine constrained rank minimization. Finally, we apply these results to the pseudo-inverse method and block matrix method.

### 5.1 Affine Constrained Rank Minimization

#### 5.1.1 Affine Rank Minimization

The affine rank minimization problem (ARMP) (Recht et al. 2010) aims at finding a matrix of minimum rank that satisfies a given system of linear equality constraints, i.e.,

$$\min \text{rank}(\mathbf{X}), \text{ such that, } \mathcal{A}(\mathbf{X}) = \mathbf{b}, \quad (22)$$

where  $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$  is a linear mapping operator. The linear mapping can always be expressed in matrix form as:

$$\mathcal{A}(\mathbf{X}) = \mathbf{A} \text{vec}(\mathbf{X}), \quad (23)$$

where  $\text{vec}(\cdot)$  denotes vectorization of  $\mathbf{X}$  with its columns stacked in order on top of one another, and  $\mathbf{A} \in \mathbb{R}^{m \times n_1 n_2}$ .

#### 5.1.2 Nuclear Norm Minimization

The affine rank minimization problem Eq. (22) is generally NP-hard (Recht et al. 2010). A common trick is to replace the rank function with the nuclear norm of the matrix, defined as  $\|\mathbf{X}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{X})$  where  $\sigma_i(\mathbf{X})$  are the  $i$ th singular values and  $r = \text{rank}(\mathbf{X})$ .

By introducing the nuclear norm heuristics, the above affine rank minimization problem are approximated by the following nuclear norm minimization formulation:

$$\min \|\mathbf{X}\|_*, \text{ such that, } \mathcal{A}(\mathbf{X}) = \mathbf{b}, \quad (24)$$

which is a convex optimization problem and can be recast as semi-definite programming (SDP).

### 5.2 Null Space Analysis

When we use the nuclear-norm relaxation to solve an original rank-minimization problem, a key theoretical question naturally arises: can the nuclear-norm relaxation always recover the exact rank-minimization solution, or is the relaxation gap tight? Conventionally, this kind of “relaxation gap questions” are answered through some statistical analysis of the incoherent conditions, or restricted isometry property (RIP), etc. developed in the field of Compressive Sensing. Recht et al. (2010) extended it to matrix rank minimization problem and proposed to use RIP to characterize the condition of unique recovery. However RIP is not invariant with respect to any invertible map, and relies crucially on particular stochastic assumption about the measurement matrix (e.g. with Gaussian entries).

In this section, we take a different, and more general viewpoint to look at the problem. As for any relaxation method, while we need to characterize the properties of the nuclear-norm relaxation itself, we also need to analyze when the original problem of rank-minimization admits unique solution. In fact, in the ARMP setting, the ability to recover  $\mathbf{X}$  from  $\mathcal{A}(\mathbf{X}) = \mathbf{b}$  depends exclusively on the null-space of  $\mathcal{A}$ , which is invariant under invertible map.

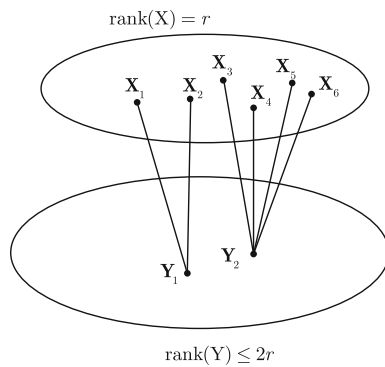
First, we review recent progress for the rank minimization problem. Based on the analysis of the null-space property of  $\mathcal{A}$ , Eldar et al. (2012) presented the following theorem regarding the uniqueness of rank minimization problem.

**Theorem 4** (Eldar et al. 2012) *When  $\mathbf{X}$  and  $\mathbf{X}'$  are each rank  $r$ , then  $\mathbf{X} - \mathbf{X}'$  is at most rank  $2r$ . To guarantee that (22) reconstruct all rank  $r$  matrices, a necessary and sufficient condition is that there are no rank  $2r$  (or less) matrices in the null space of  $\mathcal{A}$ .*

Here we provide our simplified proof.

*Proof Sufficiency:* Suppose there are no rank  $2r$  or less matrices in the null space of  $\mathcal{A}$ , if the rank minimization formulation is not not unique, i.e.,  $\mathbf{X}_1$  and  $\mathbf{X}_2$  achieve identical measurements  $\mathcal{A}(\mathbf{X}_1) = \mathcal{A}(\mathbf{X}_2)$  with rank  $r$ , then  $\mathbf{X}_1 - \mathbf{X}_2$  lies in the null space of  $\mathcal{A}$ , additionally  $\text{rank}(\mathbf{X}_1 - \mathbf{X}_2) \leq 2r$ , which conflicts the assumption. Therefore, if there are no rank  $2r$  or less matrices in the null space of  $\mathcal{A}$ , then (22) reconstructs all rank  $r$  matrices.

*Necessity:* Suppose (22) reconstructs all rank  $r$  matrices, i.e. the rank minimization solution is unique. If there is a rank  $2r$  or less matrices  $\mathbf{Y}$  in the null space of  $\mathcal{A}$ , then we can always construct two rank  $r$  matrices  $\mathbf{X}_1$  and  $\mathbf{X}_2$  such that  $\mathcal{A}(\mathbf{X}_1) = \mathcal{A}(\mathbf{X}_2)$ .



**Fig. 2** Relationship between rank  $r$  matrix manifold and rank  $2r$  matrix manifold

One direct way is: Denote the SVD of  $Y$  as  $Y = U_Y \Sigma_Y V_Y^T$ ,  $\Sigma_Y = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{2r})$ , by letting

$$\Sigma_{X_1} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r, 0, \dots, 0)$$

and

$$\Sigma_{X_2} = \text{diag}(0, \dots, 0, -\sigma_{r+1}, -\sigma_{r+2}, \dots, -\sigma_{2r}),$$

we obtain  $X_1 = U_Y \Sigma_{X_1} V_Y^T$  and  $X_2 = U_Y \Sigma_{X_2} V_Y^T$  such that  $\text{rank}(X_1) = \text{rank}(X_2) = r$  and  $\mathcal{A}(X_1) = \mathcal{A}(X_2)$ . This conflicts with the assumption. Therefore we can conclude that if (22) reconstruct all rank  $r$  matrices, then there are no rank  $2r$  or less matrices  $Y$  in the null space of  $\mathcal{A}$ .

In Fig. 2, we illustrates the relationship between rank  $r$  manifold and rank  $2r$  manifold. Let two rank  $r$  matrices be  $X_1$  and  $X_2$ , then  $Y_1 = X_1 - X_2$  lies in the matrix manifold  $\text{rank}(Y) \leq 2r$ . Note that, given a matrix  $Y_2$  in the matrix manifold  $\text{rank}(Y) \leq 2r$ , we can always construct pairs of rank  $r$  matrices  $X_3$  and  $X_4$  or  $X_5$  and  $X_6$  such that  $Y_2 = X_3 - X_4$  and  $Y_2 = X_5 - X_6$ . It is worth noting that the relationship is not necessarily full coverage, i.e., it is possible that some matrix  $X$  with  $\text{rank}(X) = r$  can not be reached from any matrix  $Y$  with  $\text{rank}(Y) \leq 2r$ ,  $Y \neq 0$  such that  $\text{rank}(X + Y) = r$ .

The above analysis gives the condition of uniform recovery under rank minimization formulation, for the nuclear norm heuristics relaxed version, we have the following result.

**Theorem 5** (Oymak and Hassibi 2010; Oymak et al. 2011) *Null space condition for low-rank recovery: Let  $\mathcal{A}$  be a linear measurement operator, then one can recover all matrices  $X_0$  with  $\text{rank}(X_0) = r$  via nuclear norm minimization (24) if and only if for any  $X \in \mathcal{N}(\mathcal{A})$  we have*

$$\sum_{i=1}^r \sigma_i(X) < \sum_{i=r+1}^{n_1} \sigma_i(X). \quad (25)$$

This theorem gives the uniform recovery condition for rank minimization via nuclear norm relaxation, i.e. recovering all the low rank matrices  $\text{rank}(X_0) = r$  such that

$\mathcal{A}(X) = \mathcal{A}(X_0)$  via nuclear norm minimization. When this condition is not satisfied, it does not necessarily mean we can not recover a specific rank  $r$  matrix  $X$  from nuclear norm minimization. When the nuclear norm minimization can not recover the low rank solution, there are two possibilities: (1) the nuclear norm minimizer in the solution space does not achieve rank minimization and (2) there are multiple rank minimizers and the nuclear norm minimizer is one of the rank minimizers.

Theorem 5 is much stronger than Theorem 4. Given the condition that  $\sum_{i=1}^r \sigma_i(X) < \sum_{i=r+1}^{n_1} \sigma_i(X)$ , we must have  $\text{rank}(X) > 2r$ , thus there are no rank  $2r$  or less matrices in the null space of  $\mathcal{A}$ , the condition for Theorem 4 is satisfied naturally. This can be explained that to recover all rank  $r$  matrices through nuclear norm minimization, the original rank minimization problem should have unique solution as well as the relaxation gap from rank minimization to nuclear norm minimization is zero.

### 5.3 Analysis of the Pseudo-Inverse Method

In this subsection and following subsection, we apply the above theoretical results to our formulation of pseudo-inverse method and block matrix method, giving analysis on uniqueness, general low rank solution and relaxation gap. Note that in the theoretical analysis, we make the assumption of noiseless image measurement and non-degenerate shapes.

By introducing matrix vectorization operator and Kronecker product, the shape recovery equation  $W = RS$  can be equivalently expressed as:

$$\text{vec}(W) = (I_P \otimes R) \text{vec}(S) = R^* \text{vec}(S), \quad (26)$$

where  $I_P \in \mathbb{R}^{P \times P}$  denotes the identity matrix. Thus the deformable shape recovery problem is transformed to:

$$\begin{aligned} &\min \text{rank}(S), \text{ such that,} \\ &\text{vec}(W) = R^* \text{vec}(S). \end{aligned} \quad (27)$$

Therefore, it is easy to check that this problem falls into the family of affine rank minimization problem Eq. (22) except that  $R^*$  has a special block matrix form.

In NRSfM, the motion matrix  $R \in \mathbb{R}^{2F \times 3F}$  is a row full rank matrix and  $R^\dagger = R^T$ . Therefore the pseudo-inverse solution  $S^* = R^\dagger W$  lies in the range space of  $R^T$ . Denote the null space of  $R$  as  $\mathcal{N}(R)$ . The solution space to the homogeneous equation  $RS = 0$  can be expressed as  $S = \mathcal{N}(R)C$ , where  $C \in \mathbb{R}^{F \times P}$  is an arbitrary coefficient matrix.

Given the pseudo inverse solution  $S^* = R^T W$  and the null space  $\mathcal{N}(R)$ , the solution system to  $W = RS$  can be expressed as:

$$S = R^T W + \mathcal{N}(R)C, \quad (28)$$

which can be equivalently formulated as:

$$S = \begin{bmatrix} R^T & \mathcal{N}(R) \end{bmatrix} \begin{bmatrix} W \\ C \end{bmatrix} = \begin{bmatrix} R^T & \mathcal{N}(R) \end{bmatrix} M. \quad (29)$$

Note that  $R^T$  and  $\mathcal{N}(R)$  are orthogonal to each other, i.e.  $R\mathcal{N}(R) = 0$  and  $R^T \oplus \mathcal{N}(R) = \mathbb{R}^{3F \times 3F}$ . Therefore  $\begin{bmatrix} R^T & \mathcal{N}(R) \end{bmatrix}$  is a full rank (non-singular) matrix and  $\text{rank}(S) = \text{rank}(M)$ . To make  $S$  lies in rank  $3K$  subspace, i.e.  $\text{rank}(M) = 3K$ ,  $C$  has to lie in the row space of  $W$ , i.e.  $C = DW$ , where  $D \in \mathbb{R}^{F \times 2F}$  is an arbitrary matrix.

Therefore, we conclude that for rank-minimization formulation (10), there are always multiple solutions lying in an affine subspace:

$$S = R^T W + \mathcal{N}(R)DW. \quad (30)$$

From this implicit rank minimization formulation, we can regularize the solution subspace in order to obtain solutions with better or desired properties such as smooth solution and sparse solution.

### 5.3.1 Frobenius Norm Minimization and Nuclear Norm Minimization

Denote the SVD of  $S^* = R^T W$  as  $S^* = U_1 D_1 V_1^T$ , the SVD of  $\mathcal{N}(R)C$  as  $\mathcal{N}(R)C = U_2 D_2 V_2^T$ . Note that  $R^T$  and  $\mathcal{N}(R)$  are orthogonal spaces to each other, thus  $U_1$  is orthogonal to  $U_2$ . The Frobenius norm of all the solutions  $\|S\|_F^2 = \sum \sigma_{i1}^2 + \sum \sigma_{i2}^2$ , where  $\sigma_{i1}$  and  $\sigma_{i2}$  are diagonal elements of  $D_1$  and  $D_2$  respectively. Apparently, when  $C = 0$ , i.e. the pseudo-inverse solution  $S^* = R^T S$  achieves Frobenius norm minimization. Additionally, Liu et al. (2013) showed that the pseudo-inverse solution i.e.  $S^* = R^T$  achieves nuclear norm minimization.

Thus we have shown that within the general solutions  $S = R^T W + \mathcal{N}(R)DW$ , the pseudo-inverse solution  $S = R^T W$  achieves both Frobenius norm minimization and nuclear norm minimization.

### 5.3.2 Degenerate Shape of Pseudo-Inverse Solution

For the pseudo-inverse solution  $S^* = R^T W$ , due to the diagonal property of the rotation matrix  $R$ , for the  $f$ th frame, we have  $S_f = R_f^T W_f$ , where  $R_f \in \mathbb{R}^{2 \times 3}$  denotes the rotation at the  $f$ th frame and  $W_f$  is the corresponding image measurement at the  $f$ th frame. Thus, it is easy to observe that  $\text{rank}(S_f) = 2$ , i.e., the non-rigid shape degenerates to be coplanar at each frame. This is due to the Frobenius norm minimization property of the pseudo-inverse solution, which explains the image measurements with minimal energy. We illustrate the degenerate shape of pseudo inverse solution of the Pickup sequence in Fig. 3, where each line in Fig. 3a corresponds to the projection of non-rigid shape at each frame.

Note that, after low rank projection, the solution becomes non-degenerate.

### 5.3.3 Geometric Interpolation

For the rotation at each frame  $R_f \in \mathbb{R}^{2 \times 3}$ , the null space of  $R_f$  can be obtained as  $\mathbf{r}_f \in \mathbb{R}^{3 \times 1}$  such that  $R_f \mathbf{r}_f = \mathbf{0}$ .  $\mathbf{r}_f$  can be retrieved from the SVD of  $R_f$  directly. Due to the diagonal property of  $R$ ,  $\mathcal{N}(R)$  is easily arrived at as the diagonal matrix of  $[\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_F]$ .

For the non-rigid shape at each frame  $f$ , we have

$$S_f = R_f^T W_f + C_f \otimes \mathbf{r}_f, \quad (31)$$

where  $C_f \in \mathbb{R}^{1 \times P}$  denotes the  $f$ th row of  $DW$ . Thus non-rigid shape at each frame  $S_f$  can be interpolated as a combination of rank 2 coplanar shape  $R_f^T W_f$  and rank 1 3D shape  $C_f \otimes \mathbf{r}_f$ . The geometric interpolation behind is that: once recovering the rotations, we can transform the image measurement to a coplanar shape in 3D space  $R_f^T W_f$ , the depth  $C_f$  of the direction orthogonal to the plane  $\mathbf{r}_f$  is to determine, thus satisfying the low rank constraint on the non-rigid shape. Below we give an illustration of the geometric interpolation in Fig. 4.

## 5.4 Analysis of the Block-Matrix-Method

We give further analysis of our block matrix method based on null space analysis of affine rank minimization problem.

### 5.4.1 Formulation

In the block matrix method, we aim to find rank minimization solution  $S^\sharp$  such that  $W = RS$ , where  $S^\sharp = [P_X \ P_Y \ P_Z](I_3 \otimes S)$ .

To facilitate analysis, we introduce a slightly different arrangement of the non-rigid shape  $S$  as:

$$S^* = \begin{bmatrix} X_{11} & Y_{11} & Z_{11} & \cdots & X_{1P} & Y_{1P} & Z_{1P} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{F1} & Y_{F1} & Z_{F1} & \cdots & X_{FP} & Y_{FP} & Z_{FP} \end{bmatrix}.$$

Note that this re-arranged shape matrix  $S^*$  is a column-wise permutation version of  $S^\sharp$ .

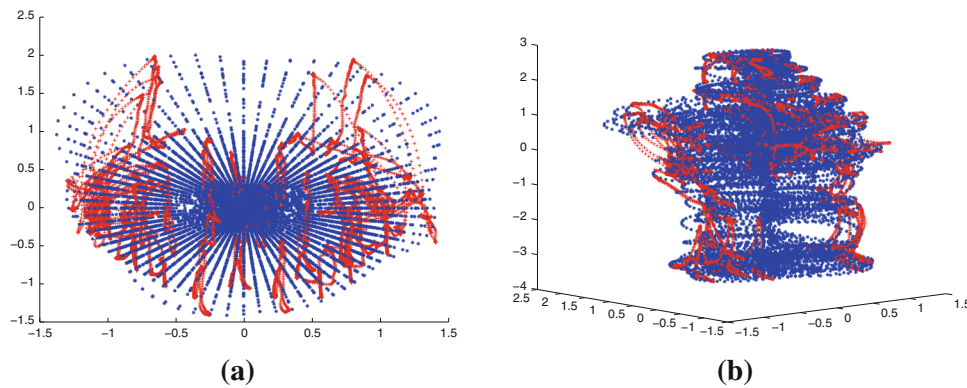
Denote the image measurement at frame  $f$  as

$$W_f = [u_{f1} \ v_{f1} \ \dots \ u_{fP} \ v_{fP}] \in \mathbb{R}^{1 \times 2P}, \quad (32)$$

let  $R_f \in \mathbb{R}^{2 \times 3}$  be the rotation at frame  $f$ , then we have

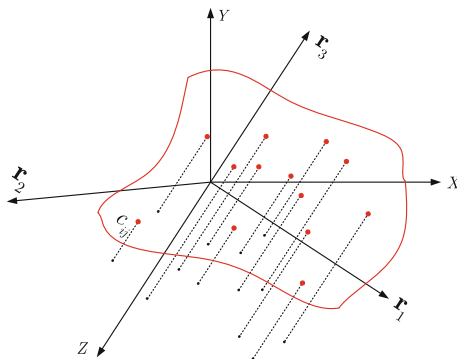
$$W_f = \mathbf{e}_f^T \left( S^* \left( I_P \otimes R_f^T \right) \right), \quad (33)$$





**Fig. 3** Non-rigid shape recovery results on the Pickup sequence from the pseudo-inverse method, note that no rank  $K$  projection has been enforced. We illustrate the recovered non-rigid shapes at each frame. Obviously, non-rigid shape degenerates to coplanar configuration at

each frame. **(a)** Non-rigid shapes at each frame, where the *blue points* show the shapes from the pseudo-inverse method and the *red points* show the ground truth 3D shapes. **(b)** Another viewpoint of the non-rigid shapes (Color figure online)



**Fig. 4** The pseudo-inverse solution gives the coplanar degenerate shape, while the component in the null space gives the coefficients along the orthogonal direction to the degenerate plane. This problem can be viewed as in a rotated coordinate (with the recovered rotations), we know planar projection of the 3D points on the plane defined by  $\mathbf{r}_1, \mathbf{r}_2$  and aim to infer the coordinate  $c_{ij}$  along  $\mathbf{r}_3$  such that the whole non-rigid shape is low rank in shape space

where  $\mathbf{e}_f \in \mathbb{R}^{F \times 1}$  is all zero except the  $f$ th component being 1. Through using vectorization and Kronecker product, we obtain

$$\text{vec}(\mathbf{W}_f) = ((\mathbf{I}_P \otimes \mathbf{R}_f) \otimes \mathbf{e}_f^T) \text{vec}(\mathbf{S}^*). \quad (34)$$

Stack image measurements at all the image frames, we obtain the following relationship:

$$\begin{bmatrix} \text{vec}(\mathbf{W}_1) \\ \vdots \\ \text{vec}(\mathbf{W}_f) \\ \vdots \\ \text{vec}(\mathbf{W}_F) \end{bmatrix} = \begin{bmatrix} ((\mathbf{I}_P \otimes \mathbf{R}_1) \otimes \mathbf{e}_1^T) \\ \vdots \\ ((\mathbf{I}_P \otimes \mathbf{R}_f) \otimes \mathbf{e}_f^T) \\ \vdots \\ ((\mathbf{I}_P \otimes \mathbf{R}_F) \otimes \mathbf{e}_F^T) \end{bmatrix} \text{vec}(\mathbf{S}^*). \quad (35)$$

The above equation can be compactly expressed as:

$$\mathbf{w} = \mathbf{R}^* \mathbf{s}, \quad (36)$$

where  $\mathbf{w} \in \mathbb{R}^{2FP \times 1}$ ,  $\mathbf{R}^* \in \mathbb{R}^{2FP \times 3FP}$  and  $\mathbf{s} \in \mathbb{R}^{3FP \times 1}$ .

Our target is to find the rank minimizer with the affine constraint defined as  $\mathbf{w} = \mathbf{R}^* \mathbf{s}$ , i.e.

$$\begin{aligned} \min \text{rank}(\mathbf{S}^*), \text{ such that,} \\ \mathbf{w} = \mathbf{R}^* \text{vec}(\mathbf{S}^*). \end{aligned} \quad (37)$$

This is standard affine constrained rank minimization problem (22). Thus we have arrived at an equivalent form of the block matrix method formulation. Nuclear norm heuristics can be utilized thus reaching:

$$\begin{aligned} \min \|\mathbf{S}^*\|_*, \text{ such that,} \\ \mathbf{w} = \mathbf{R}^* \text{vec}(\mathbf{S}^*). \end{aligned} \quad (38)$$

Uniqueness of the solution depends on the null space of  $\mathbf{R}^*$ .

Without the rank minimization constraint, the solution space of  $\text{vec}(\mathbf{S}^*)$  can be expressed as:

$$\text{vec}(\mathbf{S}^*) = (\mathbf{R}^*)^\dagger \mathbf{w} + \mathcal{N}(\mathbf{R}^*) \mathbf{c}, \quad (39)$$

In our NRSfM problem,  $(\mathbf{R}^*)^\dagger = (\mathbf{R}^*)^T$ . Therefore, the solution system is:

$$\text{vec}(\mathbf{S}^*) = [(\mathbf{R}^*)^T \quad \mathcal{N}(\mathbf{R}^*)] \begin{bmatrix} \mathbf{w} \\ \mathbf{c} \end{bmatrix}. \quad (40)$$

Our block matrix method aims to find the rank minimizer/nuclear norm minimizer in the solution space. Unlike in the pseudo inverse method case, there does not exist a closed form solution to the rank minimization/nuclear norm minimization formulation.

### 5.4.2 Null Space Analysis

Uniqueness of affine constrained rank minimization problem depends on the null space of  $R^*$ , which can be computed from the rotation  $R_f$ , therefore the null space of  $R^*$  can be easily computed from the vectors  $\mathbf{r}_f$  (orthogonal to  $R_f$ ):

$$\mathcal{N}(R^*) = [(\mathbf{I}_P \otimes \mathbf{r}_1) \otimes \mathbf{e}_1 \quad \dots \quad (\mathbf{I}_P \otimes \mathbf{r}_F) \otimes \mathbf{e}_F]. \quad (41)$$

The dimension of  $\mathcal{N}(R^*)$  is  $3FP \times FP$ . When transforming the basis vector in the null space  $\mathcal{N}(R^*)$  to  $F \times 3P$  as  $S_i^*$ , we will have  $RS_i = 0$  correspondingly. After some basic algebraic derivation, general solution to the homogeneous system  $R^* \text{vec}(S^*) = 0$  can be expressed in the form  $\sum_{i=1}^{FP} c_i S_i^*$  as:

$$\begin{aligned} \sum_{i=1}^{FP} c_i S_i^* &= \begin{bmatrix} c_{11}\mathbf{r}_1^T & c_{12}\mathbf{r}_1^T & \dots & c_{1P}\mathbf{r}_1^T \\ c_{21}\mathbf{r}_2^T & c_{22}\mathbf{r}_2^T & \dots & c_{2P}\mathbf{r}_2^T \\ \vdots & \vdots & \ddots & \vdots \\ c_{F1}\mathbf{r}_F^T & c_{F2}\mathbf{r}_F^T & \dots & c_{FP}\mathbf{r}_F^T \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{r}_1^T & & & \\ & \ddots & & \\ & & \mathbf{r}_F^T & \end{bmatrix} \left( \begin{bmatrix} c_{11} & \dots & c_{1P} \\ \vdots & \ddots & \vdots \\ c_{F1} & \dots & c_{FP} \end{bmatrix} \otimes \mathbf{I}_3 \right) \end{aligned} \quad (42)$$

It is easy to observe that the null space is fully characterized by the camera rotations  $\mathbf{r}_f$  and coefficient matrix

$$\mathbf{c} = \begin{bmatrix} c_{11} & \dots & c_{1P} \\ \vdots & \ddots & \vdots \\ c_{F1} & \dots & c_{FP} \end{bmatrix}. \text{ Additionally, we have:}$$

$$\text{rank}\left(\sum_{i=1}^{FP} c_i S_i^*\right) = 3\text{rank}(\mathbf{c}).$$

### 5.4.3 Degenerate Cases

In the above part, we have given the explicit form of null space of  $R^*$  as in Eq. (42). It is easy to observe that there even exists rank 1 solution in the null space (by letting only one  $c_{ij}$  be non-zero). According to Theorem 4, uniform recovery of deformable shape  $S^*$  is impossible, i.e., we cannot expect to recover all the low rank shape  $S^*$  such that  $R^* \text{vec}(S^*) = R^* \text{vec}(S_0^*)$  and  $\text{rank}(S^*) = \text{rank}(S_0^*)$ , where  $S_0^*$  is the ground truth non-rigid shape to recover (in rearranged form).

This shows that the block matrix method cannot achieve uniform recovery for non-rigid shape. Note that the block matrix method *equivalently* encapsulates all the constraints under the linear combination model. Thus this further informs us that for low order linear combination model, generally there are multiple realizations for the image measurements. However as we will analyze later, these results do not affect our success in real world non-rigid shape recovery.

Here we give analysis to two degenerate cases: Degenerate case (1): Given recovered camera rotations  $R_f$ , we can recover the corresponding  $\mathbf{r}_f$  through SVD. When  $\mathbf{c}$  owns low rank property, then  $\sum_{i=1}^{FP} c_i S_i^*$  will have low rank property too. We can generate arbitrary low rank  $\mathbf{c}$  (through projecting random  $\mathbf{c}$  to low rank subspace). As have illustrated in the proof of Theorem 4, from these low rank shape in the null space, we can easily construct rank  $K$  shapes  $S_1^*$  and  $S_2^*$  achieving identical measurements  $R^* \text{vec}(S_1^*) = R^* \text{vec}(S_2^*) = \mathbf{w}$ . Note that in this construction procedure, although  $\mathbf{c}$  can be arbitrary, the constructed degenerate non-rigid shape  $S^*$  depends on recovered camera motions  $R_f$ .

Degenerate case (2): Given a non-rigid shape  $S$ , we can even construct camera motions  $R_f$  such that there exists multiple low rank non-rigid shapes  $S_1^*, S_2^*$  achieving identical image measurements too.

**Theorem 6** Given a general low rank non-rigid shape  $S$  and its rearranged form  $S^*$  that  $\text{rank}(S^*) = K$ , we can always construct a measuring system  $R$  such that there are multiple low-rank solutions  $S_1$  and  $S_2$  thus  $RS_1 = RS_2$  and  $\text{rank}(S_1^*) = \text{rank}(S_2^*) = K$ .

*Proof* Without loss of generality, we assume that the first  $K$  rows of  $S_1^*$  are independent, thus all the left  $F - K$  rows of  $S_1^*$  can be expressed as a linear combination of  $(S_1^*)_i, i = 1, \dots, K$  as  $(S_1^*)_f = \sum_{k=1}^K \alpha_{fk} (S_1^*)_k$ . Then if all the rotation vectors  $\mathbf{r}_i$  have similar relationship:

$$c_{fp}\mathbf{r}_f = \sum_{k=1}^K c_{kp}\alpha_{fk}\mathbf{r}_k,$$

we can construct another shape as:

$$S_2^* = S_1^* + \begin{bmatrix} c_{11}\mathbf{r}_1^T & c_{12}\mathbf{r}_1^T & \dots & c_{1P}\mathbf{r}_1^T \\ c_{21}\mathbf{r}_2^T & c_{22}\mathbf{r}_2^T & \dots & c_{2P}\mathbf{r}_2^T \\ \vdots & \vdots & \ddots & \vdots \\ c_{F1}\mathbf{r}_F^T & c_{F2}\mathbf{r}_F^T & \dots & c_{FP}\mathbf{r}_F^T \end{bmatrix}.$$

Thus for  $f = K + 1, \dots, F$ , we have

$$\begin{aligned} (S_2^*)_{fp} &= \sum_{k=1}^K \alpha_{fk} (S_1^*)_{kp} + \sum_{k=1}^K c_{kp}\alpha_{fk}\mathbf{r}_k^T \\ &= \sum_{k=1}^K \alpha_{fk} \left( (S_1^*)_{kp} + c_{kp}\mathbf{r}_k^T \right) = \sum_{k=1}^K \alpha_{fk} (S_2^*)_{kp}. \end{aligned}$$

Therefore we can conclude that  $S_2^*$  is also a rank  $K$  non-rigid shape. Additionally, we have  $RS_1 = RS_2$ .

Thus given a non-rigid shape  $S$ , we can always find rotation  $R$  such that there are multiple solutions  $RS_1 = RS_2 = \mathbf{w}$  and  $\text{rank}(S_1^*) = \text{rank}(S_2^*) = K$ . Notice that this kind of degeneration depends on the strong correlation between the camera motion and the shape deformation.

#### 5.4.4 Relationship with Existing Theoretical Results

In the above section, we have shown two degenerate cases, where multiple solutions in non-rigid shape recovery exist. Note that due to the equivalency in the block matrix method and linear combination model, our analysis of multiple solutions applies to linear combination model naturally. Generally, these two degenerate cases depend on **shape-motion coherence**, implying given non-rigid shape  $S$ , we can construct camera motions  $R$  and given camera motions  $R$ , we can construct non-rigid shape  $S$  such that multiple solutions exist.

Hartley and Vidal (2008) analyzed the uniqueness of non-rigid shape recovery under the perspective camera model. Their conclusion states as “even though there are ambiguities in the reconstruction of shape bases and shape coefficients, the reconstructed shape is actually unique”. At a first glance, this conflicts with our analysis on degenerate cases. Nevertheless, in Hartley and Vidal (2008) assumptions on the shape and motion recovery have been made such as “The meaning of the assumption of genericity will be made clear in the proof. Broadly speaking, it means that the motion of the camera is sufficiently general and independent of the shape deformation, and that the shape space is indeed  $K$ -dimensional, spanned by the  $K$  shape bases.” Obviously, their assumptions have excluded degenerate cases of **shape-motion coherence**.

In future work, to fully characterize the uniqueness in motion and shape recovery, we will extend the concept of *reconstructibility* (Park et al. 2010; Valmadre and Lucey 2012) to NRSfM problem. Reconstructibility was proposed in analyzing trajectory reconstruction problem, where the camera motions are available. For NRSfM, this is an even harder problem due to the unknown camera motions. The hardness can be verified from another viewpoint. As we know, even for rigid structure from motion, there are degenerate cases where multiple realizations are possible (Hartley and Kahl 2007). However, we would like to emphasize again that in practical applications, camera motions and shape deformation have little possibility to be coherent, thus we can recover camera motions and non-rigid shape correctly and uniquely as verified in our experiments.

## 6 Experiments

### 6.1 Setup

We compare our methods against the state-of-the-art methods, which include (1) Xiao et al.’s shape basis method (XCK) (Xiao et al. 2004); (2) Torresani et al.’s EM-PPCA (Torresani et al. 2008); (3) Metric projection (Paladini et al. 2009); (4) Point trajectory approach (PTA or trajectory basis)

(Akhter et al. 2008); (5) Column space fitting (CSF) (Gotardo and Martinez 2011a); (6) Column space fitting with complementary rank 3 spaces (CSF2) (Gotardo and Martinez 2011c) and (7) Kernel shape trajectory approach (KSTA) (Gotardo and Martinez 2011b).<sup>6</sup>

To facilitate the comparison, we use the same error metrics as reported in Akhter et al. (2008) and Gotardo and Martinez (2011a), that is:  $e_R$  measures the mean error in rotation estimation and

$$e_R = \frac{1}{F} \sum_{f=1}^F \|R_f - \tilde{R}_f\|_F,$$

where  $R_f$  is the ground truth rotation at frame  $f$  and  $\tilde{R}_f$  is the recovered rotation;  $e_{3D}$  measures the normalized mean 3D error in the reconstructed 3D points and

$$e_{3D} = \frac{1}{\sigma_{FP}} \sum_{f=1}^F \sum_{p=1}^P e_{fp}, \quad \sigma = \frac{1}{3F} \sum_{f=1}^F (\sigma_{fx} + \sigma_{fy} + \sigma_{fz}),$$

where  $\sigma_{fx}$ ,  $\sigma_{fy}$  and  $\sigma_{fz}$  are the standard deviations in  $X$ ,  $Y$  and  $Z$  coordinates of the original shape at frame  $f$ .

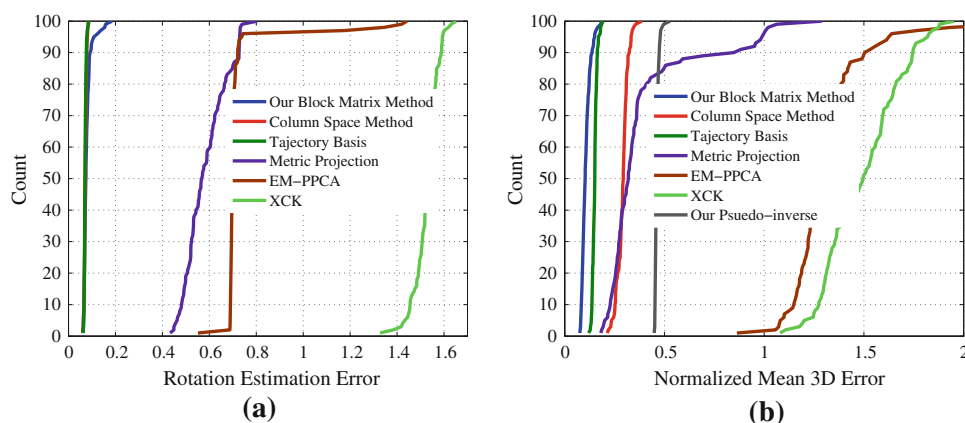
Extensive experiments were conducted to test the performance of the proposed methods, on both randomly synthetic data and real motion capture data. The random synthetic data, which satisfy the low-rank non-rigid model perfectly, are used only for the purpose of algorithm validation, for which our methods have obtained nearly perfect result (with zero error) as we expect; the results data are therefore omitted. Instead, only test results on real sequences are reported below. The real sequences we have tested include the standard sequences of Drink (1102/41), Pick-up (357/41), Yoga (307/41), Stretch (370/41), and Dance (264/75) used in Akhter et al. (2008), and Face (316/40), Shark (240/91) and Walking (260/55) in Torresani et al. (2008), where  $(F/P)$  denotes the number of frames ( $F$ ) and number of points ( $P$ ).

### 6.2 Cumulative Histograms of Error

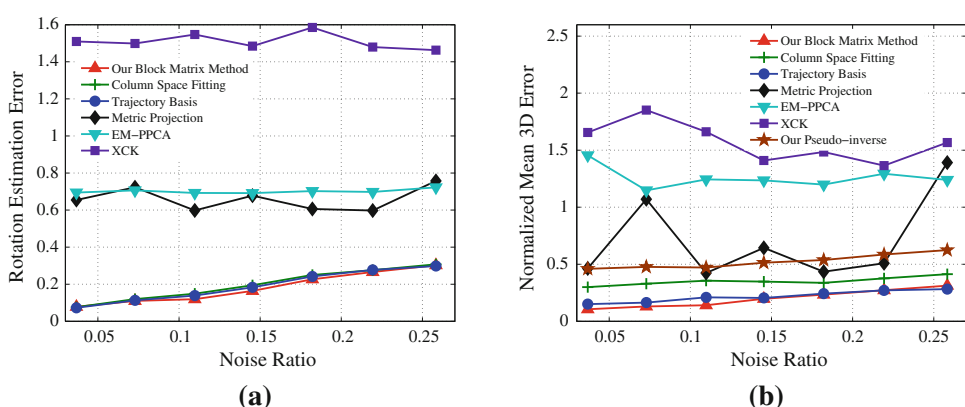
Our first experiment is aimed to give a statistical comparison between the performance of our method and several existing methods. For this purpose we used a real motion capture sequence, here e.g. the Stretch sequence. From the ground-truth 3D point clouds of the sequence as well the camera motion matrices, we re-synthesized  $F$  frames of image measurements with Gaussian random noise added in, where noise ratio is defined as  $\|\text{Noise}\|_F / \|W\|_F$ . Use the obtained data, we tested our methods, as well as several other existing methods. We repeated the random test 100 times. Then, we plotted the cumulative histograms of the rotation estimation errors, and

<sup>6</sup> Note that KSTA utilizes a non-linear mapping, which means it is out of the scope of low-order linear combination model. We provide the result mainly for the purpose of benchmarking.

**Fig. 5** Cumulative histograms of errors tested on the “Stretch” sequence. The most *left* curve gives the best performance. *Left* the rotation error; *right* the 3D reconstruction error. Our block matrix method outperforms most of the other existing methods (Color figure online)



**Fig. 6** Noise performance. (a) Rotation estimation error and (b) normalized mean 3D error



the 3D reconstruction errors, as shown in Fig. 5. This figure clearly reveals that: our block-matrix method outperforms most of the other methods.

### 6.3 Noise Performance

To analyze the behavior of our new methods under noise, we repeated the (above) first experiment at different noise levels. Example results on the “Stretch” sequence are given in Fig. 6 which plots the rotation estimation error and normalized mean 3D error as a function of the noise ratio.

It is seen, our block matrix method achieves the best performance in terms of the accuracy for rotation estimation and for non-rigid shape recovery, compared favorably with almost all the other state-of-the-art competitors.

### 6.4 Compare All Methods on All Real Sequences

In this subsection, we provide experimental results of all the nine methods we are benchmarking, on all the real sequences at hand. Tables 1 and 2 summarize our main results, where both the camera rotation error (whenever the ground truth are available) and the shape reconstruction error (normalized mean 3D error) are provided. Figure 7 shows the comparison.

**Table 1** Quantitative comparison in rotation estimation of our method with the state-of-the-art methods on motion capture data

Dataset	XCK	EM-PPCA	MP	PTA	CSF	TNM
Drink	0.336	0.291	0.286	0.006	0.006	0.007
Pick-up	0.469	0.428	0.251	0.155	0.156	0.121
Yoga	1.201	0.809	0.871	0.106	0.102	0.088
Stretch	0.949	0.760	0.817	0.055	0.049	0.068

TNM denotes our trace norm minimization method for rotation estimation

Clearly, our block matrix method achieves the best performance in shape recovery on the sequences Pickup, Yoga, Dance and Face, comparable performance other benchmark sequences Drink, Stretch and Walking. The Shark sequence is an exception, but that probably due to the fact that this Shark sequence is in fact degenerate (Torresani et al. 2008).

### 6.5 Which Component Plays More Important Role?

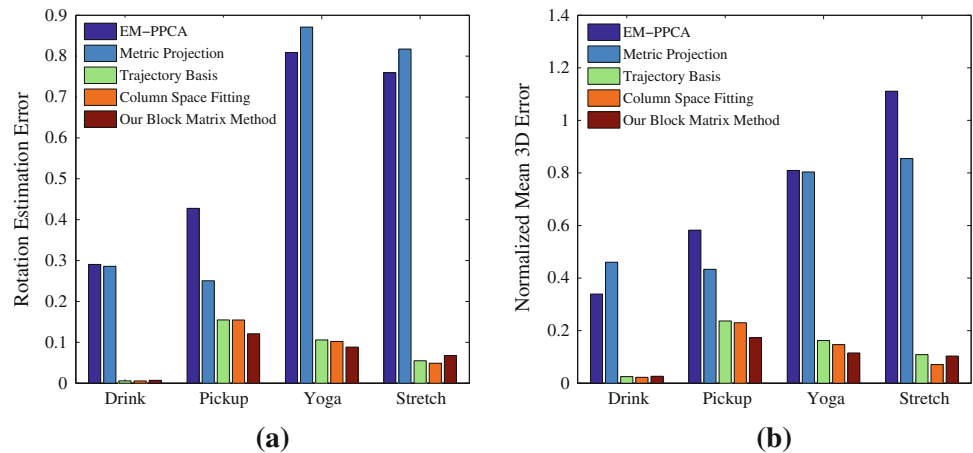
In this paper, we propose a ‘prior-free’ method to recover non-rigid shape, which consists of camera motion recovery by trace norm minimization and non-rigid shape recovery by block matrix method. It is interesting to analyze how much of the performance improvement is due to our trace norm minimization based method for estimating camera motions,



**Table 2** Quantitative comparison in non-rigid shape recovery of our proposed methods with the state-of-the-art methods on motion capture data (normalized mean 3D error)

Dataset	XCK	EM-PPCA	MP	PTA(K)	CSF(K)	CSF2(K)	KSTA(K)	PI(K)	BMM(K)
Drink	3.5186	0.3393	0.4604	0.0250(13)	0.0223(6)	0.0223(6)	<b>0.0156</b> (12)	0.4449(12)	0.0266(12)
Pick-up	3.3721	0.5822	0.4332	0.2369(12)	0.2301(6)	0.2277(3)	0.2322(6)	0.5989(12)	<b>0.1731</b> (12)
Yoga	7.4935	0.8097	0.8039	0.1625(11)	0.1467(7)	0.1464(7)	0.1476(7)	0.6523(10)	<b>0.1150</b> (10)
Stretch	4.2415	1.1111	0.8549	0.1088(12)	0.0710(8)	0.0685(8)	<b>0.0674</b> (12)	0.4784(11)	0.1034(11)
Dance	2.9962	0.9839	0.2639	0.2958(5)	0.2705(2)	0.1983(7)	0.2504(4)	0.5764(10)	<b>0.1864</b> (10)
Face	–	0.0329	0.0357	0.0444(5)	0.0363(3)	0.0314(5)	0.0339(4)	0.4848(7)	<b>0.0303</b> (7)
Walking	–	0.4917	0.5607	0.3954(2)	0.1863(2)	0.1035(5)	<b>0.1029</b> (5)	0.5100(8)	0.1298(8)
Shark	–	0.0501	0.1571	0.1796(9)	<b>0.0081</b> (3)	0.0444(5)	0.0160(3)	0.8784(3)	0.2311(4)

Bold values indicate the best performance on each dataset

**Fig. 7** Motion capture data experimental results. *Left* rotation estimation error; *right* 3D reconstruction error

and how much of the performance improvement is due to our block matrix method in non-rigid shape recovery. In this subsection, we adapted a cross-validation method to analyze it. We ran three groups of experiments, where the first one used the ground truth rotation, the second one used the rotation recovered by the point trajectory approach (PTA) (Akhter et al. 2008), and the third one used the rotation estimated by our trace norm minimization method.<sup>7</sup>

Experimental results are presented in Tables 3, 4, and 5 individually. In all the experiments, we used  $K$  as in Table 2. Note that CSF2 shares a very similar way with PTA in estimating camera rotation and the rotation recovered by CSF2 do not depend on the choice of  $K$ . In the following tables,  $PTA_0$ ,  $CSF2_0$  and  $BMM_0$  denote the non-rigid shape recovery results as in Table 2.

Table 3 shows non-rigid shape recovery performance with ground truth rotations. Compared with Table 2, it is clear that given ground truth rotations, PTA, CSF2 and BMM all achieve better performance in non-rigid shape recovery.

**Table 3** Performance comparison in shape recovery given ground truth rotations

Dataset	PTA	CSF2	BMM
Drink	0.0229	0.0215	0.0238
Pick-up	0.0992	0.0814	0.0497
Yoga	0.0580	0.0371	0.0334
Stretch	0.0822	0.0442	0.0456

Therefore, there still is space in camera motion estimation to achieve better reconstruction performance.

Table 4 shows non-rigid shape recovery performance with rotations estimated by PTA. We notice that (1) With identical rotations estimated by PTA, BMM achieves better performance compared with PTA except on Shark; (2) BMM with rotation estimated by PTA are still worse than original BMM with rotation estimation by trace norm minimization.

Table 5 shows non-rigid shape recovery performance with rotations estimated by our trace norm minimization method. Note that (1) With rotation by TNM, PTA achieves better shape recovery for sequences Pick-up, Yoga, Stretch, Dance and Walking while achieves comparable performance on Drink and Face. (2) With rotations by TNM, CSF2 achieves better performance on sequences Pickup, Yoga, Dance which

<sup>7</sup> We compare performance of our method with point trajectory approach (Akhter et al. 2008) and column space fitting method (Gotardo and Martinez 2011c), as other methods such as metric projection (Paladini et al. 2009) works in an alternating way.

**Table 4** Performance comparison in shape recovery given rotations estimated by PTA

Dataset	PTA	CSF2 <sub>0</sub>	CSF2	BMM <sub>0</sub>	BMM
Drink	0.0250	0.0223	0.0224	0.0266	0.0261
Pick-up	0.2369	0.2277	0.2271	0.1731	0.2203
Yoga	0.1625	0.1464	0.1566	0.1150	0.1539
Stretch	0.1088	0.0685	0.0751	0.1034	0.0797
Dance	0.2958	0.1983	0.2172	0.1864	0.2202
Walking	0.3954	0.1035	0.1346	0.1298	0.1880
Face	0.0444	0.0314	0.0333	0.0303	0.0329
Shark	0.1796	0.0444	0.1943	0.2311	0.2345

**Table 5** Performance comparison in shape recovery given rotation by our trace-norm minimization method

Dataset	PTA <sub>0</sub>	PTA	CSF2 <sub>0</sub>	CSF2	BMM
Drink	0.0250	0.0287	0.0223	0.0227	0.0266
Pick-up	0.2369	0.1939	0.2277	0.1791	0.1731
Yoga	0.1625	0.1243	0.1464	0.1179	0.1150
Stretch	0.1088	0.1035	0.0685	0.1136	0.1034
Dance	0.2958	0.2426	0.1983	0.1877	0.1864
Walking	0.3954	0.3761	0.1035	0.1938	0.1298
Face	0.0444	0.0489	0.0314	0.0319	0.0303
Shark	0.1796	0.2933	0.0444	0.1117	0.2311

is still inferior to BMM, while it achieves comparable performance on Drink and Face.

Therefore, we conclude that in our method, both rotation estimation by trace norm minimization and shape recovery by BMM play important role in achieving good 3D reconstruction.

## 6.6 Test on Temporally Reshuffled Data

A highlight of this work is that our method does not assume any prior knowledge about the problem. For instance, we do not assume the trajectories are smooth across frames, while many other method do assume this either explicitly or implicitly. So, we expect (predict) that our method is immune to random frame-order reshuffle (permutation).

To verify this point, we redo the experiments but on frame-reshuffled data, and obtain the following results as demonstrated in Fig. 8. It is seen that our methods is independent of frame ordering. On the original temporally smooth sequences, both the trajectory basis method and column space fitting methods achieve performance comparable but slightly inferior to our block matrix method, when the sequences are reshuffled, these methods performances very badly while our method remains unaffected.

To analyze the behavior of our new methods under noise on reshuffled sequence, we repeated experiment at different noise levels. Experimental results on the reshuffled “Stretch” sequence are given in Fig. 9 which plots the estimation errors as a function of the noise ratio.

Figure 10 gives a close inspection. Figure 10a, b compare the trajectories recovered by using the original sequence, and using a frame-reshuffled sequence, while Fig. 10c, d shows the normalized mean 3D error for both sequences. From this figure, the trajectory basis method fails to output acceptable results on the frame-permuted sequence, but our block matrix method achieves identical results on both cases. This is not surprising as permutating a matrix will not change its rank, and our methods does not assume any frame order or temporal smoothness.

## 6.7 Sample Reconstruction Results

For visual evaluation, we give result comparison between our block matrix method and the trajectory basis method on a more complex sequence, Dance, see Fig. 11.

We also tested the Talking Face video,<sup>8</sup> using 500 frames and 68 features tracks. Figure 12 shows three frames of the original images and the resulted 3D points, where the reprojection error is 0.4839 pixels.

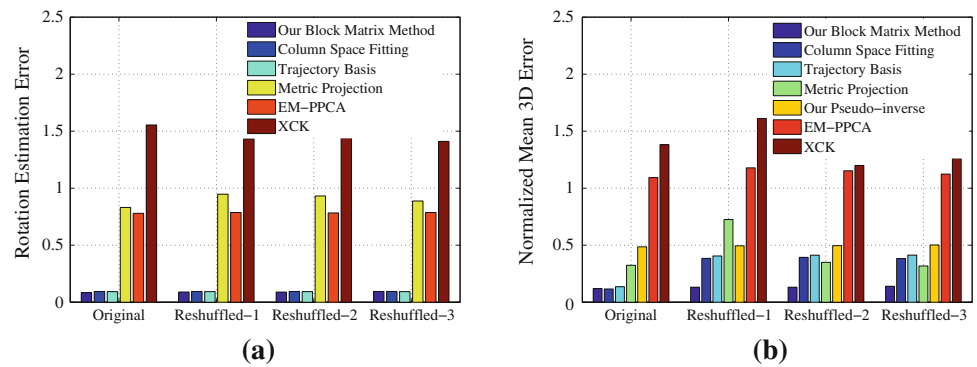
## 7 Closing Remarks

This paper advocates a prior-free approach to factorization based non-rigid structure from motion problem. Our method is purely convex, very easy to implement, and is guaranteed to converge to an optimal solution (at least approximately up to certain relaxation). It shows that, contrary to common belief, the NRSfM factorization problem can be solved unambiguously, efficiently and accurately, without using extra priors. This said, however, from a practical point of view, we do not against the use of available prior, as long as the prior is sensible and reflects the physical nature of the problem at hand. It is expected that, using good prior will further improve the quality of our solution, and make our method more applicable in complex real scenarios.

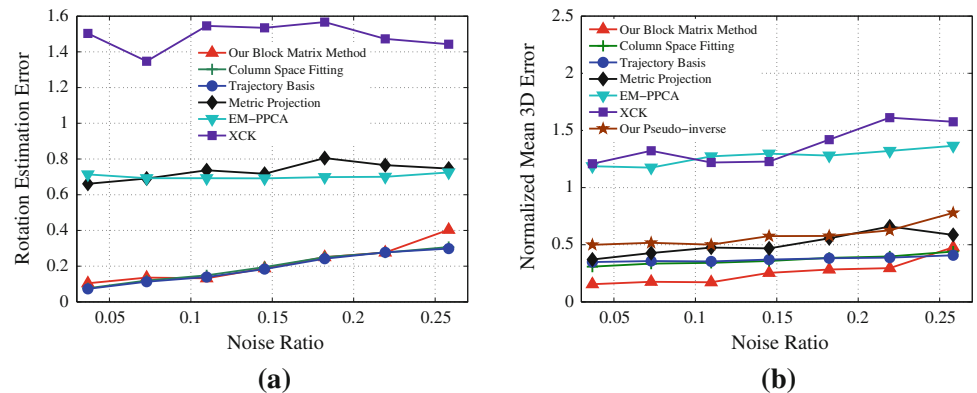
In the present paper, we have concentrated on the complete measurement case under orthographic camera model. Thanks to the recent progress in structure from motion and Compressive Sensing, the proposed method can be adapted to handling missing-data case (e.g. Buchanan and Fitzgibbon 2005; Dai et al. 2013; Tao and Yuan 2011), outlier case (e.g. Candès et al. 2011; Eriksson and van den Hengel 2010; Angst et al. 2011; Li 2009), multibody motion case

<sup>8</sup> <http://www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/>.

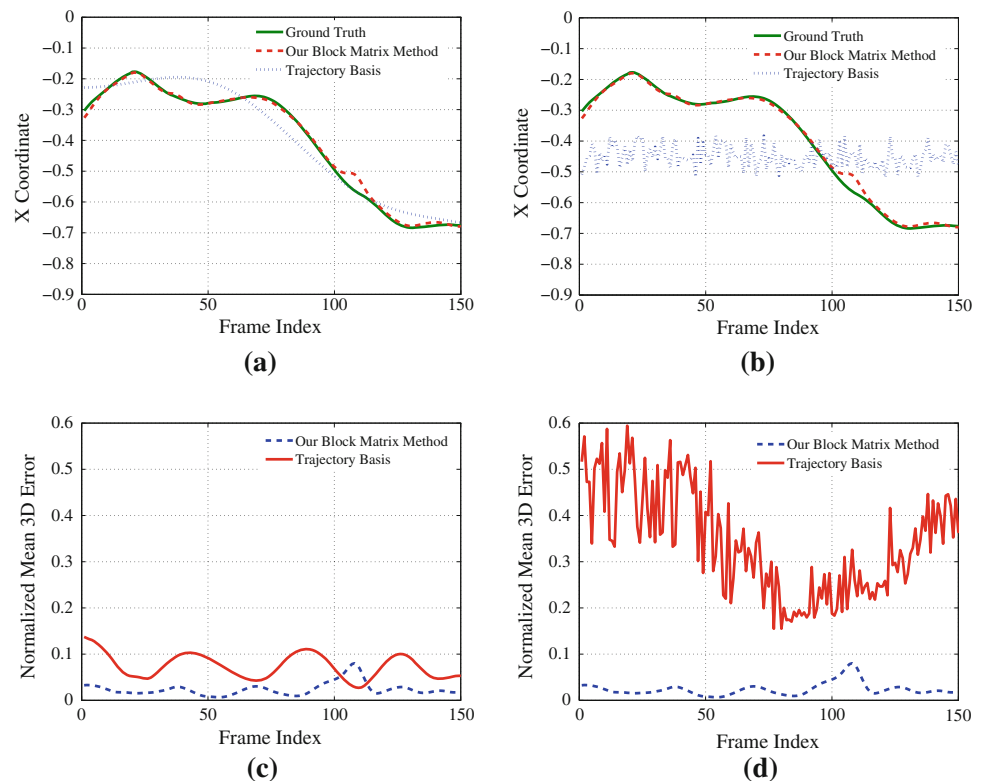
**Fig. 8** Performance on frame order reshuffled sequences. **(a)** Rotation error and **(b)** normalized mean 3D error (Color figure online)



**Fig. 9** Performance on frame order reshuffled sequences. **a** Rotation error and **b** normalized mean 3D error (Color figure online)

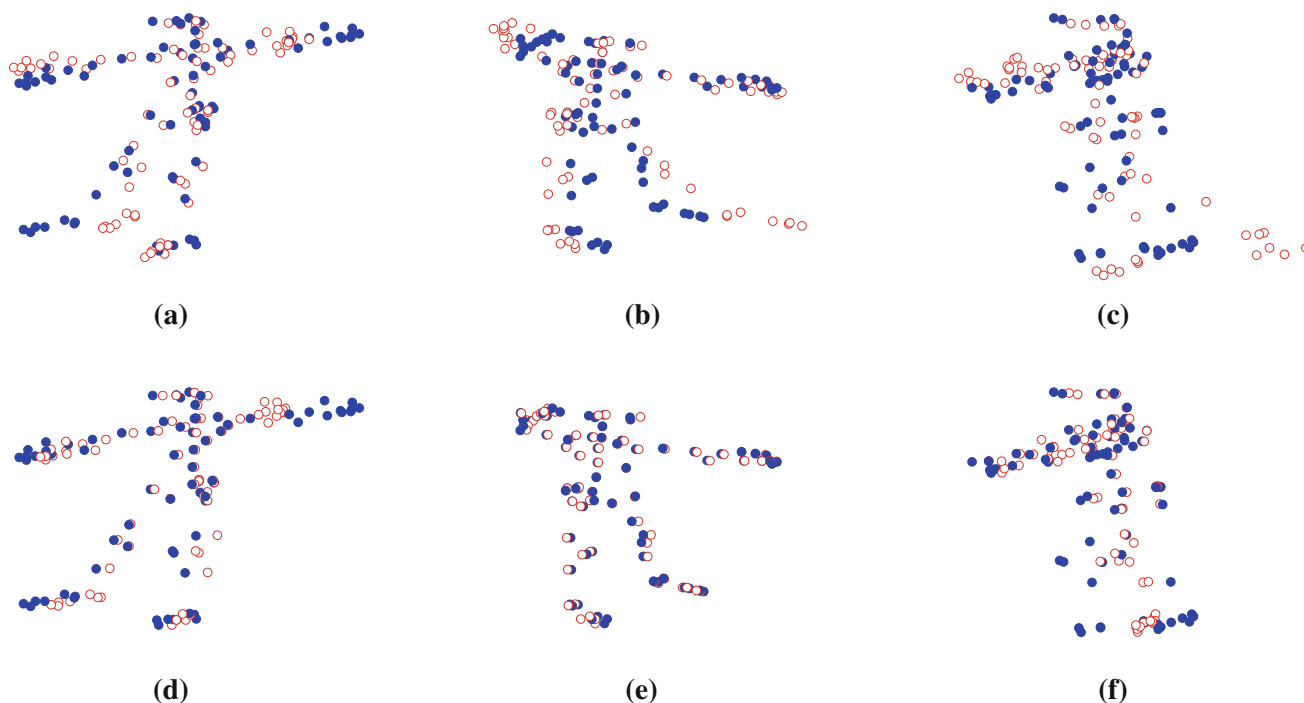


**Fig. 10** Comparison of our block matrix method versus the trajectory basis method, on an original input video sequence, as well as on a frame-reshuffled version. **(a)** Recovered trajectory of one point in the X coordinate by both methods on the original sequence. **(b)** Recovered trajectory of one point in the X coordinate by both methods on the frame-reshuffled sequence. **(c)** Normalized mean 3D error for both methods on original sequence. **(d)** Normalized mean 3D error for both methods on frame reshuffled sequence



(Li 2007), multi static cameras case (Angst and Pollefeys 2012; Zaheer et al. 2011) as well as perspective camera case (e.g., Dai et al. 2013; Xiao and Kanade 2005; Lladó et al.

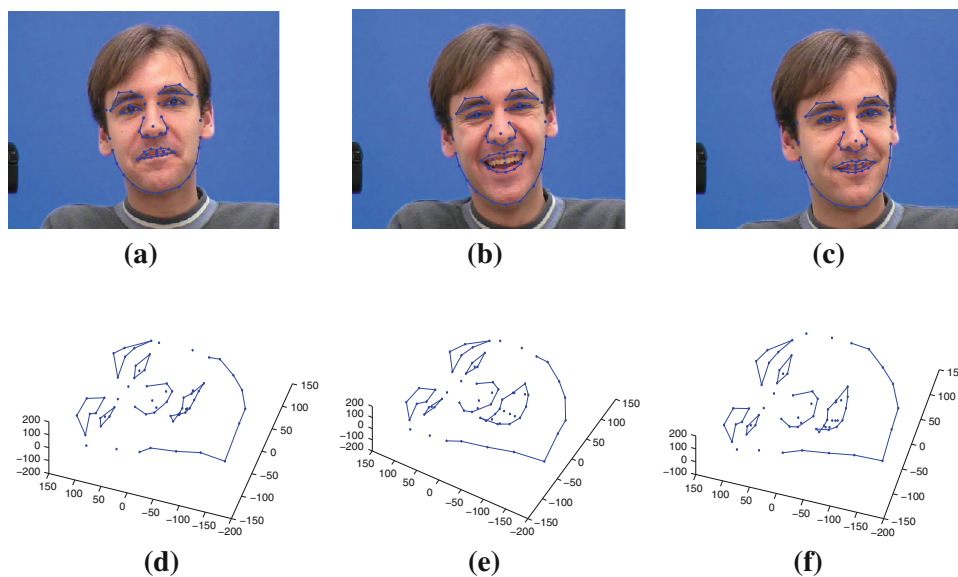
2010). Other extensions include automatic estimation of the model complexity  $K$  and the discussion of reconstructibility (Park et al. 2010) for non-rigid structure from motion.



**Fig. 11** Comparison of the 3D reconstruction results on the Dance sequence. The *blue dots* denote the ground truth 3D points, and the *red circles* show the reconstructed points. *Top row* results by the trajectory basis method (Akhter et al. 2008), where the 3D errors are 0.3011,

0.2827, 0.2814 for the three frames. *Bottom row* our result by the block matrix method, where the 3D errors are 0.2228, 0.0355, 0.1389 for the three frames (Color figure online)

**Fig. 12** The talking face sequence used in our experiment (*top row*). The *bottom row* shows the 3D reconstruction result of our method



**Acknowledgments** We wish to thank anonymous reviewers of CVPR 2012 and IJCV for their constructive comments, which greatly improve the paper. This work is funded, in part by National Natural Science Foundation of China (60736007, 61171154), and by Australia ARC-Discovery Grant and ARC-Linkage Grant. The authors wish to thank Professor Richard Hartley for guidance, supports, and invaluable discussions on this work. The first author was a China Scholarship Council (CSC)-funded PhD student to ANU, under the supervision of Professor Richard Hartley and Dr. Hongdong Li. The second author also had fruitful discussions with Professor Fredrik Kahl and Dr. Yaser Sheikh

specifically on this NRSfM topic. Dr Jing Xiao kindly provided testing images to the authors.

## References

- Aanaes, H., & Kahl, F. (2002). Estimation of deformable structure and motion. In *ECCV workshop on vision and modelling of dynamic scenes* (pp. 1–4).



- Akhter, I., Sheikh, Y., & Khan, S. (2009). In defense of orthonormality constraints for nonrigid structure from motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1534–1541).
- Akhter, I., Sheikh, Y., Khan, S., & Kanade, T. (2008). Nonrigid structure from motion in trajectory space. In *Advances in neural information processing systems* (pp. 41–48).
- Akhter, I., Simon, T., Khan, S., Matthews, I., & Sheikh, Y. (2012). Bilinear spatiotemporal basis models. *ACM Transactions on Graphics*, 31(2), 17:1–17:12. doi:10.1145/2159516.2159523.
- Angst, R., & Pollefeys, M. (2012). A unified view on deformable shape factorizations. In *Proceedings of the European conference on computer vision* (pp. 682–695). doi:10.1007/978-3-642-33783-3\_49.
- Angst, R., Zach, C., & Pollefeys, M. (2011). The generalized trace-norm and its application to structure-from-motion problems. In *Proceedings of the IEEE international conference on computer vision* (pp. 2502–2509).
- Bartoli, A., Gay-Bellile, V., Castellani, U., Peyras, J., Olsen, S., & Sayd, P. (2008). Coarse-to-fine low-rank structure-from-motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–8).
- Brand, M. (2005). A direct method for 3D factorization of nonrigid motion observed in 2D. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 122–128).
- Bregler, C., Hertzmann, A., & Biermann, H. (2000). Recovering non-rigid 3D shape from image streams. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 690–696).
- Brunet, F., Hartley, R., Bartoli, A., Navab, N., & Malgouyres, R. (2011). Monocular template-based reconstruction of smooth and inextensible surfaces. In *Proceedings of the Asian conference on computer vision* (pp. 52–66). doi:10.1007/978-3-642-19318-7\_5.
- Buchanan, A. M., & Fitzgibbon, A. W. (2005). Damped Newton algorithms for matrix factorization with missing data. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 316–322). doi:10.1109/CVPR.2005.118.
- Cai, J., & Osher, S. (2010). *Fast singular value thresholding without singular value decomposition*. Technical report, University of California, Los Angeles.
- Candès, E., Li, X., Ma, Y., & Wright, J. (2011). Robust principal component analysis? *Journal of the ACM*, 58(3), 11:1–11:37.
- Candès, E. J., & Plan, Y. (2010). Matrix completion with noise. *Proceedings of the IEEE*, 98(6), 925–936. doi:10.1109/JPROC.2009.2035722.
- Dai, Y., Li, H., & He, M. (2010). Element-wise factorization for n-view projective reconstruction. In *Proceedings of the European conference on computer vision* (pp. 396–409).
- Dai, Y., Li, H., & He, M. (2012). A simple prior-free method for non-rigid structure-from-motion factorization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2018–2025). doi:10.1109/CVPR.2012.6247905.
- Dai, Y., Li, H., & He, M. (2013). Projective multiview structure and motion from element-wise factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9), 2238–2251. doi:10.1109/TPAMI.2013.20.
- Del Bue, A. (2008). A factorization approach to structure from motion with shape priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–8).
- Eldar, Y. C., Needell, D., & Plan, Y. (2012). Uniqueness conditions for low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 33(2), 309–314. doi:10.1016/j.acha.2012.04.002.
- Eriksson, A., & van den Hengel, A. (2010). Efficient computation of robust low-rank matrix approximations in the presence of missing data using the  $L_1$  norm. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 771–778).
- Fayad, J., Agapito, L., & Del Bue, A. (2010). Piecewise quadratic reconstruction of non-rigid surfaces from monocular sequences. In *Proceedings of the European conference on computer vision* (pp. 297–310).
- Goldfarb, D., & Ma, S. (2011). Convergence of fixed-point continuation algorithms for matrix rank minimization. *Foundations of Computational Mathematics*, 11(2), 183–210. doi:10.1007/s10208-011-9084-6.
- Gotardo, P., & Martinez, A. (2011a). 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. doi:10.1109/TPAMI.2011.50.
- Gotardo, P., & Martinez, A. (2011b). Kernel non-rigid structure from motion. In *Proceedings of the IEEE international conference on computer vision* (pp. 802–809).
- Gotardo, P., & Martinez, A. (2011c). Non-rigid structure from motion with complementary rank-3 spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3065–3072).
- Hartley, R., & Kahl, F. (2007). Critical configurations for projective reconstruction from multiple views. *International Journal of Computer Vision*, 71(1), 5–47. doi:10.1007/s11263-005-4796-1.
- Hartley, R., & Vidal, R. (2008). Perspective nonrigid shape and motion recovery. In *Proceedings of the European conference on computer vision* (pp. 276–289).
- Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge: Cambridge University Press. ISBN: 0521540518.
- Li, H. (2007). Two-view motion segmentation from linear programming relaxation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–8). doi:10.1109/CVPR.2007.382975.
- Li, H. (2009). Consensus set maximization with guaranteed global optimality for robust geometry estimation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1074–1080). doi:10.1109/ICCV.2009.5459398.
- Li, H. (2010). Multi-view structure computation without explicitly estimating motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2777–2784). doi:10.1109/CVPR.2010.5540005.
- Lin, Z., Chen, M., & Ma, Y. (2010). The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. ArXiv e-prints. <http://arxiv.org/abs/1009.5055>.
- Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., & Ma, Y. (2013). Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 171–184. doi:10.1109/TPAMI.2012.88.
- Lladó, X., Bue, A. D., & Agapito, L. (2010). Non-rigid metric reconstruction from perspective cameras. *IVC*, 28(9), 1339–1353.
- Ma, S., Goldfarb, D., & Chen, L. (2011). Fixed point and Bregman iterative methods for matrix rank minimization. *Mathematical Programming, Series A*, 128(1–2), 321–353.
- Olsen, S. I., & Bartoli, A. (2008). Implicit non-rigid structure-from-motion with priors. *Journal of Mathematical Imaging and Vision*, 31(2–3), 233–244. doi:10.1007/s10851-007-0060-3.
- Oymak, S., & Hassibi, B. (2010). New null space results and recovery thresholds for matrix rank minimization. ArXiv e-prints. <http://arxiv.org/abs/1011.6326>.
- Oymak, S., Mohan, K., Fazel, M., Hassibi, B. (2011). A simplified approach to recovery conditions for low rank matrices. ArXiv e-prints. <http://arxiv.org/abs/1103.1178>.
- Paladini, M., Bue, A. D., Stosic, M., Dodig, M., Xavier, J., & Agapito, L. (2009). Factorization for non-rigid and articulated structure using metric projections. In *Proceedings of the IEEE conference on*

- computer vision and pattern recognition (pp. 2898–2905). doi:[10.1109/CVPRW.2009.5206602](https://doi.org/10.1109/CVPRW.2009.5206602).
- Park, H. S., Shiratori, T., Matthews, I., & Sheikh, Y. (2010). 3D reconstruction of a moving point from a series of 2D projections. In *Proceedings of the European conference on computer vision* (pp. 158–171).
- Perriollat, M., Hartley, R., & Bartoli, A. (2011). Monocular template-based reconstruction of inextensible surfaces. *International Journal of Computer Vision*, 95(2), 124–137. doi:[10.1007/s11263-010-0352-8](https://doi.org/10.1007/s11263-010-0352-8).
- Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., et al. (2004). Visual modeling with a hand-held camera. *International Journal of Computer Vision*, 59(3), 207–232. doi:[10.1023/B:VISI.0000025798.50602.3a](https://doi.org/10.1023/B:VISI.0000025798.50602.3a).
- Rabaud, V., & Belongie, S. (2009). Linear embeddings in non-rigid structure from motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2427–2434). Los Alamitos, CA: IEEE Computer Society. doi:[10.1109/CVPRW.2009.5206628](https://doi.org/10.1109/CVPRW.2009.5206628).
- Recht, B., Fazel, M., & Parrilo, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3), 471–501. doi:[10.1137/070697835](https://doi.org/10.1137/070697835).
- Salzmann, M., Moreno-Noguer, F., Lepetit, V., & Fua, P. (2008). Closed-form solution to non-rigid 3D surface registration. In *Proceedings of the European conference on computer vision* (Vol. 5305, pp. 581–594). doi:[10.1007/978-3-540-88693-8\\_43](https://doi.org/10.1007/978-3-540-88693-8_43).
- Snavely, N., Seitz, S., & Szeliski, R. (2008). Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2), 189–210. doi:[10.1007/s11263-007-0107-3](https://doi.org/10.1007/s11263-007-0107-3).
- Sturm, J. (1999). Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11, 625–653.
- Tao, M., & Yuan, X. (2011). Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM Journal on Optimization*, 21(1), 57–81.
- Taylor, J., Jepson, A., & Kutulakos, K. (2010). Non-rigid structure from locally-rigid motion. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2761–2768). doi:[10.1109/CVPR.2010.5540002](https://doi.org/10.1109/CVPR.2010.5540002).
- Toh, K., Todd, M., & Tutuncu, R. (1999). SDPT3—A Matlab software package for semidefinite programming. *Optimization Methods and Software*, 11, 545–581.
- Tomasi, C., & Kanade, T. (1992). Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2), 137–154. doi:[10.1007/BF00129684](https://doi.org/10.1007/BF00129684).
- Torresani, L., Hertzmann, A., & Bregler, C. (2008). Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5), 878–892.
- Triggs, B., McLauchlan, P. F., Hartley, R. I., & Fitzgibbon, A. W. (2000). Bundle adjustment—A modern synthesis. In *Proceedings of the international workshop on vision algorithms* (pp. 298–372).
- Valmadre, J., & Lucey, S. (2012). General trajectory prior for non-rigid reconstruction. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1394–1401). doi:[10.1109/CVPR.2012.6247826](https://doi.org/10.1109/CVPR.2012.6247826).
- Xiao, J., Chai, J., & Kanade, T. (2004). A closed-form solution to non-rigid shape and motion recovery. In *Proceedings of the European conference on computer vision* (Vol. 3024, pp. 573–587). doi:[10.1007/978-3-540-24673-2\\_46](https://doi.org/10.1007/978-3-540-24673-2_46).
- Xiao, J., & Kanade, T. (2005). Uncalibrated perspective reconstruction of deformable structures. In *Proceedings of the IEEE international conference on computer vision* (pp. 1075–1082). doi:[10.1109/ICCV.2005.241](https://doi.org/10.1109/ICCV.2005.241).
- Yang, J., & Yuan, X. (2013). Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. *Mathematics of Computation*, 82(281), 301–329.
- Zaheer, A., Akhter, I., Baig, M., Marzban, S., & Khan, S. (2011). Multi-view structure from motion in trajectory space. In *Proceedings of the IEEE international conference on computer vision* (pp. 2447–2453). doi:[10.1109/ICCV.2011.6126529](https://doi.org/10.1109/ICCV.2011.6126529).