Simultaneous Correspondences Estimation and Non-rigid Structure Reconstruction

Yuchao Dai[1], and Hongdong Li[1,2]
[1] Research School of Engineering, Australian National University
[2]ARC Centre of Excellence for Robotic Vision (ACRV)
Email: {yuchao.dai, hongdong.li}@anu.edu.au

Abstract—Given multi-view correspondences, it has been shown that 3D non-rigid structure can be recovered through factorization based techniques. However, establishing reliable correspondences across multi-view images of non-rigid structure is not an easy task. Existing methods solve multi-view correspondences and 3D non-rigid structure in sequel, which cannot exploit the crossover constraints in each sub-problem (*i.e.*, constraints in non-rigid structure has not been enforced in establishing multi-view correspondences and verse vise). In this paper, we present a unified framework to simultaneously solve for multi-view correspondences and non-rigid structure. We formulate the problem by using the Partial Permutation Matrices (PPMs) and aim at establishing multi-view correspondences while simultaneously enforcing the low-rank constraint in non-rigid structure deformation. Additionally, our method can handle outliers and missing data elegantly under the same framework. We solve the simultaneous non-rigid structure and correspondences recovery problem via the Alternating Direction Method of Multipliers (ADMM). Experimental results on both synthetic and real images show that the proposed method achieves state-of-the-art performance on both sparse and dense non-rigid reconstruction problems.

I. INTRODUCTION

Recently, non-rigid structure-from-motion (NRSfM) has received considerable attention from the computer vision community. A great number of methods have been established, and most of the existing methods can be roughly classified as *single view template-based methods* (*e.g.* [27], [25]) and *multi-view correspondence-based methods* (*e.g.* [3],[37],[5],[8]). Template-based methods generally assume a known 3D template (*e.g.* a mesh model) of the non-rigid surface and a set of 3D-to-2D correspondences between that image and another one for which the 3D shape is known. Multi-view correspondence-based methods make use of intra-frame (and often sparse) 2D feature point correspondences to simultaneously recover camera motion and non-rigid shape.

While achieving considerable success on benchmark data sets, these existing approaches generally assume 2D-3D or 2D-2D correspondences can be readily established, which is not true in practical applications. Actually, establishing correspondences across images containing non-rigid deformation is an active research area (*e.g.*, [35], [38], [33], [40]). Additionally, even though these methods tolerate mismatches to some extent, their performance will degrade dramatically with the increase of mismatches. When there are repetitive patterns in the images, the problem becomes even harder.

Recently, in the area of single view template-based surface reconstruction, there have been attempts to simultaneously solving for 3D-2D correspondences and non-rigid shape. Shaji *et al.*[29] solved the joint optimization problem by a branch-and-bound algorithm. Furthermore, Sánchez-Riera *et al.*[28] presented an approach to simultaneously solve for 2D-3D correspondences, camera pose and non-rigid shape of the surface, where priors on the camera pose and non-rigid shape were modeled with the Gaussian Mixture Model (GMM). However these strategies cannot simply be extended to NRSfM where multi-view 2D-2D correspondences are desired.

In this paper, we move a further step and remove the need for multi-view correspondences in recovering 3D non-rigid structure. Under our new formulation, the input is a collection of feature points from multiple images and the outputs are 3D non-rigid structure and multi-view correspondences. Specifically, we search for non-rigid structure and a reordering of feature point positions such that the deformable shapes satisfy the low rank constraint in the linear combination model [3] (other models such as union of subspaces [41] can also be used) while the reordered feature point position matrix fulfils the imaging constraint. Therefore, under our formulation, constraint in non-rigid structure helps in establishing multi-view correspondences while multi-view correspondences further constrain deformation of non-rigid structure. In this way, we can achieve both better 3D non-rigid reconstruction and multi-view correspondences. Additionally, our formulation can handle outliers and missing data elegantly under the same framework.

To the best of our knowledge, this is the first approach to address the problem of simultaneous multi-view correspondences and non-rigid structure reconstruction from multi-view images. Our main contributions are stated as: 1)

1) Given the input of feature points in each image, our new framework simultaneously recovers multi-view correspondences and 3D non-rigid structure;
2) Constraints in non-rigid structure and multi-view correspondences are simultaneously enforced, thus better conditioning both problems;
3) Outliers and missing data in feature point position measurements are handled elegantly under the same framework.

II. RELATED WORK

Here, we briefly review existing work on both non-rigid reconstruction from multi-view correspondences and multi-view feature point matching.

Non-rigid structure from motion factorization: Bregler *et al.*[3] introduced the linear combination model to represent non-rigid/deformable shapes, resulting in a matrix factorization framework to camera motion and non-rigid structure recovery. Since then, researchers have been actively applying the factorization framework to various non-rigid reconstruction problems. However, a seminal work by Xiao *et al.*[37] showed that the problem is indeed ill-posed or under-constrained. To deal with the underlying ambiguity, Xiao *et al.*[37] suggested to add "basis constraints", resulting in a closed-form solution to NRSfM. With the same spirit of adding extra priors to regularize an otherwise under-constrained problem, different priors have been introduced, *e.g.*Gaussian prior on the shape coefficients [34], special shape priors [7], and temporally smooth deformation prior [2]. Recently, Dai *et al.*[5] showed that by enforcing the low rank constraint properly (on the reshuffled shape matrix rather than original shape matrix), non-rigid factorization problem can be solved efficiently without any ambiguity. To handle increased complexity in long sequence motion capture, Zhu *et al.*[41] represented non-rigid deformation by using union of subspaces rather than sum of subspaces, which simultaneously doing clustering and non-rigid reconstruction. Other attempts toward complex non-rigid motion include separable spatiotemporal priors [30], Procrustean normal distribution [19], and Procrustean Markov process [20].

Recently, there have been attempts to formulating non-rigid reconstruction in a machine learning framework, *e.g.*, by using diffusion map [31] or spatially-smooth mapping [16]. However, the requirements for training data sets are not always fulfilled in practical applications. Meanwhile, another more recent trend in NRSfM is to extend the sparse non-rigid reconstruction techniques to dense scenario where spatial constraint has been utilized to regularize the problem [11].

Non-rigid matching: Multi-view correspondences based non-rigid structure reconstruction require establishing 2D-2D correspondences across multiple images. For dense non-rigid reconstruction, optical flow and its extensions have been utilized to achieve dense motion estimation [12][26]. For sparse non-rigid reconstruction, there are recent work such as [33][35][38] and [40] that recast non-rigid feature correspondence as a graph matching problem. To remove outliers in measurements, Ransac [10] has also been used [36]. However Ransac-like approaches quickly become impractical as the number of transformation parameters or graph nodes increases. A most recent work by Collins and Bartoli [4] proposed to resolve that difficulty in 2D-3D matching by detecting incorrect correspondences using the isometry constraint directly in 3D.

Our work in building correspondences is most related to [39] and [17], where sparse and low-rank decomposition

model is used to find feature correspondences. In this paper, we show how problem-specific knowledge (for example low-rank condition in reshuffled non-rigid 3D shape) can help in finding correspondences.

III. FORMULATION

In this section, we formulate the problem of simultaneous multi-view correspondences and 3D non-rigid reconstruction as the search for 3D non-rigid structure and a reordering of the feature points such that the non-rigid structure satisfies the low-rank constraint under the linear combination model and the reordered feature point coordinate matrix relates to the non-rigid structure via image model (affine transformation in our case).

Factorization based non-rigid structure from motion aims at factorizing the *multi-view image measurement matrix* $\mathbf{W}$ as the product of *camera motion* (projection) matrix $\mathbf{M}$ and a non-rigid shape matrix $\mathbf{S}$, such that $\mathbf{W} = \mathbf{MS}$. In real world application, rigid structure usually makes the dominant component in the scene, therefore, camera motion estimation can be solved by utilizing rigid structure in the scene (*i.e.* by using rigid factorization [32]). For the rest of the paper, we generally assume the camera motions $\mathbf{M}$ have been recovered. We will leave the problem of simultaneous camera motion, non-rigid structure and multi-view correspondences for future work. Note that with recovered camera motion, the remaining problem seems similar to the trajectory reconstruction problem [24]. However, we claim that under our problem formulation, we use the inherent structure constraint in non-rigid deformation rather than treating each trajectory independently.

Prerequisite: More specifically, we consider a monocular camera observing a non-rigid structure. Under an affine camera model, the image measurement $\mathbf{m}_{ij} = [u_{ij}, v_{ij}]^T$ and 3D point $\mathbf{S}_{ij}$ on the deformable shape are related by the camera motion $\mathbf{R}_i$ as:

$$\mathbf{m}_{ij} = \mathbf{R}_i \mathbf{S}_{ij}, \tag{1}$$

where $\mathbf{R}_i \in \mathbb{R}^{2 \times 3}$ denotes the first two rows of the i-th camera rotation. The coordinates of the 2D image points observed at frame i are given by: $\mathbf{M}_i = \mathbf{R}_i \mathbf{S}_i$, where $\mathbf{M}_i = [\mathbf{m}_{i1} \cdots \mathbf{m}_{iP}]$ and $\mathbf{S}_i = [\mathbf{S}_{i1} \cdots \mathbf{S}_{iP}]$. Use this representation, and stack all the F frames of measurements and all the P points in a matrix form, we reach:

$$\mathbf{M} = \begin{bmatrix} \mathbf{R}_1 \mathbf{S}_1 \\ \vdots \\ \mathbf{R}_F \mathbf{S}_F \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 & & \\ & \ddots & \\ & & \mathbf{R}_F \end{bmatrix} \begin{bmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_F \end{bmatrix} = \mathbf{RS}, \tag{2}$$

where $\mathbf{R} = \mathrm{blkdiag}(\mathbf{R}_1, \cdots, \mathbf{R}_F) \in \mathbb{R}^{2F \times 3F}$ expresses the camera motion matrix.

Under the low-order linear combination model, the non-rigid shape $\mathbf{S}_i \in \mathbb{R}^{3 \times P}$ can be expressed as a linear combination of K shape bases $\mathbf{B}_k \in \mathbb{R}^{3 \times P}$ with time-varying shape coefficients c_{ik} as: $\mathbf{S}_i = \sum_{k=1}^{K} c_{ik} \mathbf{B}_k$. Therefore, the non-rigid shape matrix $\mathbf{S}$ is low rank as $\mathrm{rank}(\mathbf{S}) \leq 3K$.

In [5], Dai *et al.* proposed to enforce the low rank constraint in the reshuffled deformable shape rather than $\mathbf{S}$ directly, which

re-arranges the rows of S that correspond to X, Y, and Z coordinate separately, and in an $F \times 3P$ block matrix form,

$$S^\sharp = g(S) = [P_X \ \ P_Y \ \ P_Z](I_3 \otimes S). \tag{3}$$

where $P_X, P_Y, P_Z \in \mathbb{R}^{F \times 3F}$ are row-selection matrices (of 0-1 values, similar to a row-permutation matrix) [1] 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.

Modeling correspondences: In the absence of point correspondences across frames, simply stacking up of all the feature point position matrices $\{M_i\}_{i=1}^F$ does not yield valid point trajectories. However, there exists a reordering of the columns of the position matrix in each frame that yields coherent point trajectories.

We utilize the Partial Permutation Matrices (PPMs) to model the reordering of point positions, which has been widely used in modeling correspondences [22], [21], [39]. Here, we denote $P_f \in \{0,1\}^{N_f \times N}$ as the PPM that selects and reorders the N inlier point coordinates from the N_f measurements in frame f,

$$\mathcal{P}^k = \{P^\mathbf{k} \in \mathbb{R}^{\mathbf{N_f} \times \mathbf{N}} | P^\mathbf{k}_{\mathbf{ij}} \in \{\mathbf{0}, \mathbf{1}\}, \sum_\mathbf{i} P^\mathbf{k}_{\mathbf{ij}} = \mathbf{1}, \tag{4}$$

$$\forall j = 1, \dots, N, \sum_j P^\mathbf{k}_{\mathbf{ij}} \le \mathbf{1}, \forall \mathbf{i} = \mathbf{1}, \dots, \mathbf{N_k}\},$$

the resultant point position matrix defined upon PPMs $\{P_f\}_{f=1}^F$ is obtained as

$$W(\{P_f\}_{f=1}^F) = \begin{bmatrix} M_1 P^1 \\ \vdots \\ M_F P^F \end{bmatrix}. \tag{5}$$

Note a global permutation ambiguity exists, *i.e.*, permutating the order of all the trajectories does not change any underlying constraints.

To account for the outliers in point position measurements, we express the reordered measurement matrix W as a combination of clean data from non-rigid shape RS and sparse outliers E, that is

$$W = RS + E. \tag{6}$$

Note that, by utilizing the PPMs, we are able to handle missing data in measurements. With the introduction of outliers E, we are even able to deal with sparse outliers. Therefore, our unified framework can handle missing data and outliers in the same framework.

---

[1] $P_X(i, 3i - 2) = 1, P_Y(i, 3i - 1) = 1, P_Z(i, 3i) = 1$, while all the other positions being zero.

Piecing all the constraint together, we reach an optimization model for simultaneous multi-view correspondences and non-rigid structure reconstruction.

$$\begin{aligned} \min_{S, S^\sharp, E, \{P_f\}_{f=1}^F} & \ \|S^\sharp\|_* + \lambda\|E\|_1 \\ \text{s.t.} \ & W = RS + E, \\ & S^\sharp = g(S), \\ & W(\{P_f\}_{f=1}^F) = \begin{bmatrix} M_1 P^1 \\ \vdots \\ M_F P^F \end{bmatrix}, \end{aligned} \tag{7}$$

where the nuclear norm $\|\|_*$ is used as a convex surrogate to the rank function, λ is a trade-off parameter to balance the non-rigid structure term and sparse outlier term. Due to the discrete nature of the PPMs, our formulation is a non-convex non-linear optimization problem.

a) Constraining W directly: : In the above formulation, we enforce low rank constraint on the reshuffled deformable shape $S^\sharp$, which expresses the low order linear combination model equivalently. Note that if only point correspondences are desired, a simplified model can be achieved by enforcing constraints on the re-ordered measurement matrix W only. The problem formulation can be easily derived by removing constraints on S and $S^\sharp$ in Eq. (7),

$$\begin{aligned} \min_{L, E, \{P_f\}_{f=1}^F} & \ \|L\|_* + \lambda\|E\|_1, \\ \text{s.t.} \ & W = L + E, \\ & W(\{P_f\}_{f=1}^F) = \begin{bmatrix} M_1 P^1 \\ \vdots \\ M_F P^F \end{bmatrix}. \end{aligned} \tag{8}$$

Given feature points in multiple images, Eq. (8) estimates a reordering such that the reordered measurements matrix is low rank. The solution of Eq. (8) is similar to the solution of Eq. (7). Essentially, low rank constraint in image measurement matrix originates from the low rank constraint in 3D non-rigid deformation. A low rank structure in 3D deformation naturally projects to a low rank structure in 2D image measurements. However, a low rank structure pursuing in image measurement matrix does not necessarily correspond to the underlying low rank structure in 3D non-rigid deformation. A related analysis is available in [6]. In the experiments section, we will compare point correspondence recovery performance of both Eq. (7) and Eq. (8).

IV. SOLUTION

In this section, we develop a numerical solver for simultaneous multi-view correspondences and non-rigid reconstruction by adapting the classic Alternative Direction Method of Multipliers (ADMM) to non-convex optimization problem. To solve the problem, we introduce Lagrange multipliers Γ and

Υ to remove the equality constraints. The resulting augmented Lagrangian cost function is expressed as:

$$\mathcal{L}(\mathtt{S}^\sharp, \mathtt{S}, \mathtt{E}, \{\mathtt{P}^f\}_{f=1}^F, \Gamma, \Upsilon) = \|\mathtt{S}^\sharp\|_* + \lambda\|\mathtt{E}\|_1$$
$$+ < \Gamma, \mathtt{S}^\sharp - g(\mathtt{S}) > + < \Upsilon, \mathtt{W} - \mathtt{RS} - \mathtt{E} > \qquad (9)$$
$$+ \beta/2(\|\mathtt{S}^\sharp - g(\mathtt{S})\|_F^2 + \|\mathtt{W} - \mathtt{RS} - \mathtt{E}\|_F^2),$$

where $< \mathtt{X}, \mathtt{Y} >= \mathrm{tr}(\mathtt{X}^T\mathtt{Y})$, and β is a penalty parameter. The ADMM works by minimizing Eq. (9) with respect to $\mathtt{S}^\sharp, \mathtt{S}, \mathtt{E}$ and $\{\mathtt{P}^f\}_{f=1}^F$ one at a time while fixing the others. During each iteration, we update $\mathtt{S}^\sharp, \mathtt{S}, \mathtt{E}, \{\mathtt{P}^f\}_{f=1}^F$ and Lagrange multipliers Γ, Υ in sequel

$$\mathtt{S}_{k+1}^\sharp = \arg\min_{\mathtt{S}^\sharp} \mathcal{L}(\mathtt{S}^\sharp, \mathtt{S}_k, \mathtt{E}_k, \{\mathtt{P}_k^f\}_{f=1}^F, \Gamma_k, \Upsilon_k), \qquad (10)$$

$$\mathtt{S}_{k+1} = \arg\min_{\mathtt{S}} \mathcal{L}(\mathtt{S}_{k+1}^\sharp, \mathtt{S}, \mathtt{E}_k, \{\mathtt{P}_k^f\}_{f=1}^F, \Gamma_k, \Upsilon_k), \qquad (11)$$

$$\mathtt{E}_{k+1} = \arg\min_{\mathtt{E}} \mathcal{L}(\mathtt{S}_{k+1}^\sharp, \mathtt{S}_{k+1}, \mathtt{E}, \{\mathtt{P}_k^f\}_{f=1}^F, \Gamma_k, \Upsilon_k), \qquad (12)$$

$$\{\mathtt{P}_{k+1}^f\}_{f=1}^F = \arg\min_{\{\mathtt{P}^f\}_{f=1}^F} < \Upsilon_k, \mathtt{W}_k - \mathtt{RS}_{k+1} - \mathtt{E}_k >$$
$$+ \beta/2(\|\mathtt{W}_k - \mathtt{RS}_{k+1} - \mathtt{E}_{k+1}\|_F^2), \qquad (13)$$

$$\Gamma_{k+1} = \Gamma_k + \beta(S_{k+1}^\sharp - g(\mathtt{S}_{k+1})), \qquad (14)$$

$$\Upsilon_{k+1} = \Upsilon_k + \beta(\mathtt{W}_{k+1} - \mathtt{RS}_{k+1} - \mathtt{E}_{k+1}), \qquad (15)$$

$$\beta_{k+1} = \min(\beta_m, \eta\beta_k). \qquad (16)$$

The update of $\mathtt{S}_{k+1}^\sharp$ involves a convex optimization problem for which a closed form solution exists. By introducing the matrix shrinkage operator $\mathcal{S}_v(\cdot)$ [2], the optimal solution to Eq. (10) can be obtained as:

$$\mathtt{S}_{k+1}^\sharp = \mathcal{S}_{1/\beta}(g(\mathtt{S}_k) - \Gamma_k/\beta). \qquad (17)$$

For the update of $\mathtt{S}$, as it only includes linear and least-squares terms. Therefore, its update can be easily obtained in closed-form. More specifically, the closed-form solution is given by:

$$\mathtt{S}_{k+1} = (\mathtt{I} + \mathtt{R}^T\mathtt{R})^{-1}(g^{-1}(\Gamma_k)/\beta \qquad (18)$$
$$+ \mathtt{R}^T\Upsilon_k/\beta + g^{-1}(\mathtt{S}_{k+1}^\sharp) + \mathtt{R}^T(\mathtt{W}_k - \mathtt{E}_k)).$$

By using the soft-thresholding operator [9] $\tau_v(x) = \mathrm{sign}(x)\max(0, |x| - v)$, the optimal solution for $\mathtt{E}$ is achieved as:

$$\mathtt{E}_{k+1} = \tau_{\lambda/\beta}(\mathtt{W}_k - \mathtt{RS}_{k+1} + \Upsilon_k/\beta), \qquad (19)$$

where the soft-shresholding operator works on $\mathtt{W}_k - \mathtt{RS}_{k+1} + \Upsilon_k/\beta$ element-wisely.

---

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

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

Require: Camera motion $\mathtt{R}$, point positions measurements $\mathtt{M}_1, \cdots, \mathtt{M}_F, \lambda$;
  Initialize: $\mathtt{S}_0, \mathtt{S}_0^\sharp, \mathtt{E}_0, \{\mathtt{P}_0^f\}_{f=1}^F, \beta_0$.
  while Not converged **do**
    1. Update $\mathtt{S}^\sharp, \mathtt{S}$ and $\mathtt{E}$ by Eq. (17), Eq. (18) and Eq. (19);
    2. Update $\{\mathtt{P}^f\}_{f=1}^F$ as
    $\mathtt{P}_f = \arg\min\|\mathtt{M}_f\mathtt{P}_f - (\mathtt{RS}_{k+1} + \mathtt{E}_{k+1} - 1/\beta\Upsilon_k)_f\|_F^2$ by using the Hungarian algorithm;
    3. Update Γ, Υ and β as Eq. (14), Eq. (15) and Eq. (16);
    4. Check the convergence conditions:
    $\|\mathtt{S}_{k+1}^\sharp - g(\mathtt{S}_{k+1})\|_\infty \le \epsilon_1, \|\mathtt{W}_{k+1} - \mathtt{RS}_{k+1}\|_\infty \le \epsilon_2$.
  end while
Ensure: Non-rigid shape $\mathtt{S}$, reshuffled shape $\mathtt{S}^\sharp$ and outlier pattern $\mathtt{E}$.
Algorithm 1: Simultaneous correspondences and non-rigid structure reconstruction via ADMM.

As there is no bilinear element between partial permutation matrices across all the frames, problem (13) can be decomposed into F sub-problems, each of which only involves one PPM $\mathtt{P}^f$,

$$\mathtt{P}_{k+1}^f = \arg\min_{\mathtt{P}^f} \mathrm{tr}\left[(\mathtt{M}_f\mathtt{P}_f)^T(\mathtt{M}_f\mathtt{P}_f)\right] \qquad (20)$$
$$- 2\mathrm{tr}\left[(\mathtt{M}_f\mathtt{P}_f)^T(\mathtt{RS}_{k+1} + \mathtt{E}_{k+1} - 1/\beta\Upsilon_k)_f\right],$$

where $\mathtt{M}_f$ denotes the position measurements at the f-th frame, and $(\mathtt{M}_f\mathtt{P}_f)^T(\mathtt{RS}_{k+1} + \mathtt{E}_{k+1} - 1/\beta\Upsilon_k)_f$ denotes the $2f-1, 2f$ rows of matrix $(\mathtt{M}_f\mathtt{P}_f)^T(\mathtt{RS}_{k+1} + \mathtt{E}_{k+1} - 1/\beta\Upsilon_k)$ which corresponds to the f-th frame. After some algebraic derivation, we reach the following equivalent formula,

$$\mathtt{P}_{k+1}^f = \arg\min_{\mathtt{P}^f} \|\mathtt{M}_f\mathtt{P}^f - (\mathtt{RS}_{k+1} + \mathtt{E}_{k+1} - 1/\beta\Upsilon_k)_f\|_F^2. \qquad (21)$$

This equation exactly expresses the problem as a linear assignment problem, which can be solved by using the Hungarian algorithm [18].

Our algorithm to perform simultaneous correspondences and non-rigid structure reconstruction is summarized in Algorithm 1. The initialization is algorithm implementation is achieved as : the PPMs are set as identity, we work out corresponding deformable shape $\mathtt{S}_0, \mathtt{S}_0^\sharp$ by state-of-the-art NRSfM methods such as BMM. The outlier pattern is initialized as $\mathtt{E}_0 = 0$.

V. Experiments

In this section, we report experimental results on various configurations. We have tested our method on real sequences including Drink (1102/41), Pick-up (357/41), Yoga (307/41), Stretch (370/41), and Dance (264/75) used in [1], and Face (316/40), Shark (240/91) and Walking (260/55) in [34], where (F/P) denotes the number of frames (F) and number of points (P). To facilitate the comparison, we use the same error metrics as reported in [1] and [13]. e_{3D} measures the

normalized mean 3D error in the reconstructed 3D points, which is defined as

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

where σ_{fx}, σ_{fy} and σ_{fz} are the standard deviations in X, Y and Z coordinates of the original shape at frame f. To evaluate the performance in building correspondences, we use the metric of correspondence accuracy, which is defined as the ratio between the number of corrected estimated correspondence and the number of point positions.

We compare our methods against the state-of-the-art non-rigid structure from motion methods, which include (1) Point trajectory approach (PTA) [1]; (2) Column space fitting (CSF2) [15] and Block matrix method (BMM) [5].

A. Tolerance to mismatches

The state-of-the-art NRSfM methods achieve considerable performance improvement on benchmark data sets (*i.e.*, CMU mocap). However, their performance will degrade when mismatches occur. Here, we present two examples to illustrate this effect. On the "Yoga" and "Pickup" sequences, we randomly permutated the feature correspondences to create mismatches in each frame. The ratio of mismatches gradually increase from 0% to 10% and experiments were ran 10 times for each ratio. Statistical results are reported in Fig. 1.
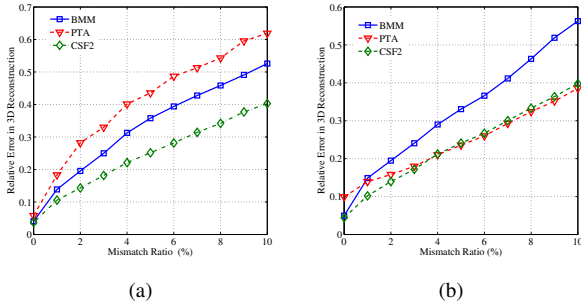

(a) (b)

Fig. 1. Performance of state-of-the-art non-rigid structure reconstruction method (PTA, CSF2, BMM) while mismatch ratio in measurement matrix gradually increases from 0% to 10%. (a) Experimental results on "Yoga"; (b) Experimental results on "Pickup".

In non-rigid reconstruction, 0.2 and above in relative error express a big change in 3D reconstruction. It is easy to observe that with 1% and 2% mismatches, performance of all the state-of-the-art methods (BMM, CSF2 and PTA) degrade dramatically from less than 0.05 to around 0.2 and 0.3. The figure clearly illustrates the performance decrease of the state-of-the-art NRSfM methods in the presence of mismatches. These motivate our simultaneous correspondence and non-rigid reconstruction framework. As illustrated in the following parts, our new framework does not depend on pre-computed correspondences, which enables it to handle even 100% mismatches.

B. Simultaneously correspondences and reconstruction

To evaluate the performance of our algorithm in simultaneously estimating correspondences and reconstruction, we utilize the sequences "Yoga", "Pickup", "Stretch" and "Drink" where the camera motions are available. For each frame in the video sequence, we randomly generated permutation matrices to obtain permutated feature point positions. Then our method was applied to recover both correspondences and non-rigid structure. Furthermore, we ran the implementation constraining W only as in Eq. (8).

In Fig. 2, we illustrate the performance comparison on sequences "Pickup" and "Yoga" between our model Eq. (7) and model Eq. (8) in the aspect of permutation estimation accuracy. Meanwhile, we provide non-rigid structure reconstruction error accordingly in Fig. 2(d) and Fig. 2(b), where the normalized mean 3D error is comparable with the result given perfect correspondences. Obviously, our formulation with constraint on $S^{\sharp}$ outperforms the formulation with constraint on W by a margin. This can be understood as the low rank condition on $S^{\sharp}$ essentially expresses the low order constraint in the linear combination model equivalently. In Fig. 3, we presented 3D reconstruction of our framework on "Yoga" sequence.
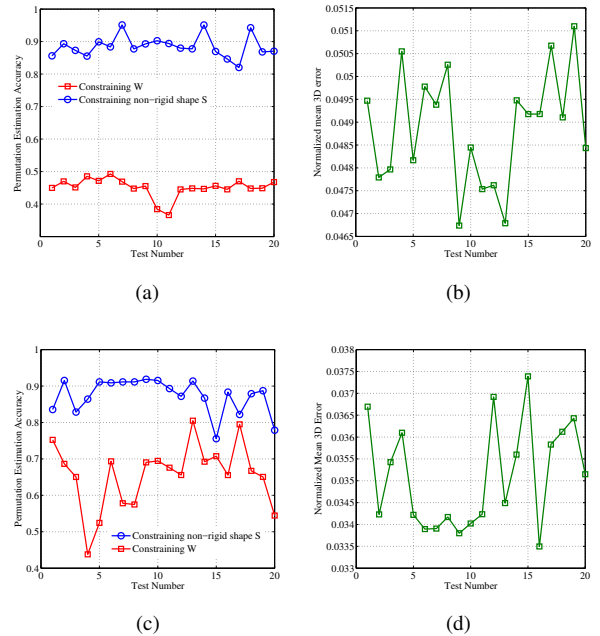

(a) (b)

(c) (d)

Fig. 2. Performance of correspondence estimation and non-rigid structure reconstruction, where the performance is reported as Permutation Estimation Accuracy and Normalized Mean 3D Error under each test. (a) Correspondence estimation accuracy for "Yoga"; (b) Normalized mean 3D error for "Yoga"; (c) Correspondence estimation accuracy for "Pickup"; (b) Normalized mean 3D error for "Pickup".

In Table I, we reported the average correspondence estimation accuracy and normalized mean 3D error in non-rigid reconstruction. To better position the performance of our method, we also provided 3D reconstruction results of the state-of-the-art non-rigid reconstruction methods (PTA, CSF2
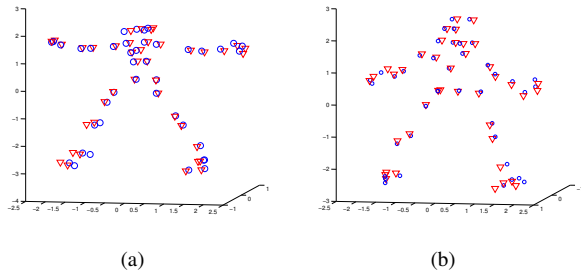
Fig. 3. 3D reconstruction results on the Yoga sequence, where red triangle denotes our results and blue circle denotes the ground truth.

and BMM) with ground truth correspondences. From the table, we observe that, our method, while not requiring any point correspondences as input, achieves comparable or even better shape recovery performance compared with PTA, CSF2 and BMM.

C. Sparse point trajectory results

By fixing the correspondences, our method can be applied to non-rigid reconstruction problem with correspondences. To illustrate the benefit of our ADMM formulation, here we compared our method with the state-of-the-art methods in non-rigid structure from motion factorization. The results are illustrated in Table II, where EM-PPCA denotes the EM-PPCA method proposed in [34], MP denotes metric projection method presented in [23], while KSTA denotes the Kernel Shape Trajectory Approach in [14], . It is observed that our method achieves comparable 3D reconstruction results with state-of-the-art methods. However, compared with fixed-point continuation based BMM implementation, our ADMM based implementation could achieve a speed-up factor of 5.

Additionally, efficient implementation of the ADMM method enables us to solve even dense non-rigid structure from motion problem under the same framework. To show the scalability of our method, we applied our method to dense non-rigid structure from motion problem. We tested it on the synthetic face data set used in [11]. Experimental results are illustrated in Table III. Note that under our formulation, we only use the constraint in deformable shape without using any spatial constraint as in [11]. Nevertheless, our method achieves comparable performance while its implementation is much more efficient.

VI. CONCLUSION

This paper advocates a unified framework to simultaneously recover non-rigid structure and point correspondences. We formulated the problem as searching for deformable shape and a reordering of the feature point positions such that the reshuffled non-rigid shape is low-rank and the reordered point

positions lie in the subspace spanned by the non-rigid shape. The resultant non-convex optimization problem is solved via the ADMM, where for each subproblem there exists efficient solution. Experimental results on real image evidence the performance of our method in simultaneously establishing correspondences and estimating non-rigid 3D shape. In the future, we plan to extend our method to handle camera motion and hence simultaneously dealing with camera motion, non-rigid structure and multi-view correspondences.

REFERENCES

[1] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Nonrigid structure from motion in trajectory space. In *Advances in Neural Information Processing Systems*, pages 41–48, 2008. 4, 5, 7

[2] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd. Coarse-to-fine low-rank structure-from-motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1–8, 2008. 2

[3] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 690–696, 2000. 1, 2

[4] T. Collins and A. Bartoli. Using isometry to classify correct/incorrect 3d-2d correspondences. In *European Conference on Computer Vision*, volume 8692, pages 325–340. 2014. 2

[5] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 2018–2025, 2012. 1, 2, 5

[6] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. *Int. J. Comput. Vision*, 107(2):101–122, Apr. 2014. 3

[7] A. Del Bue. A factorization approach to structure from motion with shape priors. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1–8, 2008. 2

[8] H. Deng and Y. Dai. Pushing the limit of non-rigid structure-from-motion by shape clustering. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1999–2003, March 2016. 1

[9] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION*, pages 1200–1224, 1995. 4

[10] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. 2

[11] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1272–1279, 2013. 2, 6, 7

[12] R. Garg, A. Roussos, and L. Agapito. A variational approach to video registration with subspace constraints. *International Journal of Computer Vision*, pages 1–29, 2013. 2

[13] P. Gotardo and A. Martinez. Computing smooth time-trajectories for camera and deformable shape in structure from motion with occlusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(10):2051–2065, Dec 2011. 4

[14] P. Gotardo and A. Martinez. Kernel non-rigid structure from motion. In *Proc. IEEE Int'l Conf. Computer Vision*, pages 802–809, 2011. 6

[15] P. Gotardo and A. Martinez. Non-rigid structure from motion with complementary rank-3 spaces. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 3065–3072, 2011. 5

[16] O. Hamsici, P. Gotardo, and A. Martinez. Learning spatially-smooth mappings in non-rigid structure from motion. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors, *Proc. European Conf. Computer Vision*, volume 7575 of *Lecture Notes in Computer Science*, pages 260–273. Springer Berlin Heidelberg, 2012. 2

[17] P. Ji, H. Li, M. Salzmann, and Y. Dai. Robust motion segmentation with unknown correspondences. In *European Conference on Computer Vision*, pages 204–219. 2014. 2

TABLE I

QUANTITATIVE EVALUATION ON MOTION CAPTURE DATA WITH UNKNOWN CORRESPONDENCES, WHERE THE METHODS PTA, CSF2 AND BMM TAKE GROUND TRUTH CORRESPONDENCES AS INPUT.

Dataset	correspondence accuracy	reconstruction error	PTA	CSF2	BMM
Yoga	95.55%	0.0330	0.0580	0.0371	0.0334
Pickup	93.73%	0.0499	0.0992	0.0814	0.0497
Stretch	81.32%	0.0574	0.0822	0.0442	0.0456
Drink	96.72%	0.0221	0.0229	0.0215	0.0238

TABLE II

QUANTITATIVE COMPARISON IN NON-RIGID SHAPE RECOVERY OF OUR PROPOSED METHOD WITH THE STATE-OF-THE-ART METHODS ON MOTION CAPTURE DATA (NORMALIZED MEAN 3D ERROR).

Dataset	EM-PPCA	MP	PTA(K)	CSF2(K)	KSTA(K)	BMM(K)	ADMM(K)
Drink	0.3393	0.4604	0.0250(13)	0.0223(6)	0.0156(12)	0.0266(12)	0.0263(12)
Pick-up	0.5822	0.4332	0.2369(12)	0.2277(3)	0.2322(6)	0.1731(12)	0.1731(12)
Yoga	0.8097	0.8039	0.1625(11)	0.1464(7)	0.1476(7)	0.1150(10)	0.1148(10)
Stretch	1.1111	0.8549	0.1088(12)	0.0685(8)	0.0674(12)	0.1034(11)	0.1036(11)
Dance	0.9839	0.2639	0.2958(5)	0.1983(7)	0.2504(4)	0.1864(10)	0.1864(10)
Face	0.0329	0.0357	0.0444(5)	0.0314(5)	0.0339(4)	0.0303(7)	0.0424(7)
Walking	0.4917	0.5607	0.3954(2)	0.1035(5)	0.1029(5)	0.1298(8)	0.1307(8)
Shark	0.0501	0.1571	0.1796(9)	0.0444(5)	0.0160(3)	0.2311(4)	0.2270(3)

TABLE III

QUANTITATIVE EVALUATION ON 4 SYNTHETIC FACE SEQUENCES IN [11]. (AVERAGE RMS 3D RECONSTRUCTION ERROR.)

Dataset	PTA [1]	MP [23]	DV [11]	ADMM
Seq1	0.2431	0.2575	0.0531	0.0884
Seq2	0.0988	0.0644	0.0457	0.0506
Seq3	0.0596	0.0682	0.0346	0.0367
Seq4	0.0877	0.0772	0.0379	0.0409

[18] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistic Quarterly*, 2:83–97, 1955. 4

[19] M. Lee, J. Cho, C.-H. Choi, and S. Oh. Procrustean normal distribution for non-rigid structure from motion. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1280–1287, 2013. 2

[20] M. Lee, C.-H. Choi, and S. Oh. A procrustean markov process for non-rigid structure recovery. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1550–1557, June 2014. 2

[21] M. Marques, M. Stosic, and J. Costeira. Subspace matching: Unique solution to point matching with geometric constraints. In *Proc. IEEE Int'l Conf. Computer Vision*, pages 1288–1294, 2009. 3

[22] R. Oliveira, J. Costeira, and J. Xavier. Optimal point correspondence through the use of rank constraints. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 2, pages 1016–1021 vol. 2, 2005. 3

[23] M. Paladini, A. D. Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito. Factorization for non-rigid and articulated structure using metric projections. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 2898–2905, 2009. 6, 7

[24] H. S. Park, T. Shiratori, I. Matthews, and Y. Sheikh. 3D reconstruction of a moving point from a series of 2D projections. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Proc. European Conf. Computer Vision*, volume 6313 of *Lecture Notes in Computer Science*, pages 158–171. Springer Berlin / Heidelberg, 2010. 2

[25] M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. *International Journal of Computer Vision*, 95(2):124–137, 2011. 1

[26] S. Ricco and C. Tomasi. Video motion for every visible point. In *The IEEE International Conference on Computer Vision (ICCV)*, 2013. 2

[27] M. Salzmann, F. Moreno-Noguer, V. Lepetit, and P. Fua. Closed-form solution to non-rigid 3d surface registration. In *Proc. European Conf. Computer Vision*, pages 581–594, 2008. 1

[28] J. Sánchez-Riera, J. Östlund, P. Fua, and F. Moreno-Noguer. Simultaneous pose, correspondence and non-rigid shape. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1189–1196, 2010. 1

[29] A. Shaji, A. Varol, L. Torresani, and P. Fua. Simultaneous point matching and 3d deformable surface reconstruction. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1221–1228, 2010. 1

[30] T. Simon, J. Valmadre, I. Matthews, and Y. Sheikh. Separable spatiotemporal priors for convex reconstruction of time-varying 3D point clouds. In *European Conference on Computer Vision*, pages 204–219. 2014. 2

[31] L. Tao and B. Matuszewski. Non-rigid structure from motion with diffusion maps prior. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 1530–1537, 2013. 2

[32] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int'l J. Computer Vision*, 9(2):137–154, 1992. 2

[33] M. Torki and A. Elgammal. One-shot multi-set non-rigid feature-spatial matching. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 3058–3065, 2010. 1, 2

[34] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(5):878–892, 2008. 2, 4, 6

[35] L. Torresani, V. Kolmogorov, and C. Rother. Feature correspondence via graph matching: Models and global optimization. In *Proc. European Conf. Computer Vision*, pages 596–609, 2008. 1, 2

[36] Q.-H. Tran, T.-J. Chin, G. Carneiro, M. Brown, and D. Suter. In defence of ransac for outlier rejection in deformable registration. In *Proc. European Conf. Computer Vision*, volume 7575, pages 274–287. Springer Berlin Heidelberg, 2012. 2

[37] J. Xiao, J. Chai, and T. Kanade. A closed-form solution to non-rigid shape and motion recovery. In *Proc. European Conf. Computer Vision*, volume 3024, pages 573–587, 2004. 1, 2

[38] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios. Dense non-rigid surface registration using high-order graph matching. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 382–389, June 2010. 1, 2

[39] Z. Zeng, T.-H. Chan, K. Jia, and D. Xu. Finding correspondence from multiple images via sparse and low-rank decomposition. In *Proc. European Conf. Computer Vision*, volume 7576, pages 325–339. 2012. 2, 3

[40] F. Zhou and F. D. la Torre. Deformable graph matching. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2013. 1, 2

[41] Y. Zhu, D. Huang, F. De La Torre, and S. Lucey. Complex non-rigid motion 3d reconstruction by union of subspaces. June 2014. 1, 2