RECOVERING COMPLEX NON-RIGID 3D STRUCTURES FROM MONOCULAR IMAGES BY UNION OF NONLINEAR SUBSPACES

*Yanan Chen, Fei Wang, Xuan Wang**

Institute of Artificial Intelligence and Robotics
Xi'an Jiaotong University
710049, 28 Xianning Road, Xian, China

ABSTRACT

Non-rigid structure from motion (NRSfM) is a well-known challenging task due to its inherent ambiguities. Most existing approaches rely on kinds of low-rank *linear* subspaces assumption to make the problem well-constrained.

In this paper, we make two contributions. First, we empirically present that relying on the assumption, 3D shapes lie on a union of *non-linear* subspaces, can better model the complex non-rigid motion than its *linear* counterparts. Second, we introduce the nonlinear low-rank representation as a regularizer to the objective function for NRSfM and show that it can be solved by alternating direction multiplier method (ADMM). Our experiments demonstrate that our method yields more accurate reconstruction and more reasonable clustering results than state-of-the-art methods, on CMU MoCap and UMPM datasets.

Index Terms— Non-rigid structure-from-motion, kernel methods, subspace clustering

1. INTRODUCTION

This paper aims to solve the problem of recovering the 3D structure of an object, with complex non-rigid deformation, from a monocular image sequence. Such problem is ill-posed, therefore additional priors and assumptions are needed.

Lots of approaches have proposed to recover the non-rigid structures from monocular images in past decades. The seminal factorization method [1], first proposed to handle non-rigid scenario by extending its rigid version [2]. Xiao et al. [3] showed the inherent ambiguity in modeling non-rigid shape and proposed a remedy of "basis constraints" to derive a closed-form solution. Akhter et al. [4] presented a dual approach by modeling 3D trajectories, i.e. "trajectory space". Then it has proved in [5] that even there is an ambiguity in shape bases or trajectory bases, non-rigid shapes can still be solved uniquely. Based on this, a "prior-free" method has proposed to recover the 3D non-rigid structures and rotations by only exploiting the low-rank constraint. Besides shape basis model and trajectory basis model, the shape-trajectory approach [6] combines two models and formulates the problems as revealing trajectory of the shape basis coefficients. Moreover, a Procrustean Normal Distribution (PND) model, where 3D shapes are aligned and fit into a normal distribution, proposed in [7]. Simon et al. [8] proposed to exploit the Kronecker pattern in the shape-trajectory (spatial-temporal) priors. Zhu and Lucey [9] applied the convolutional sparse coding technique to NRSFM using point trajectories. However, the method requires learning an over-complete basis of 3D trajectories, prior to performing the 3D reconstruction. However, most existing approaches, which highly depend on the complexity of the motion, will fail to recover the complex motion.

In a more practical scenario, a relatively longer sequence will usually depict more complex non-rigid motion which can't be well-approximated by single low-rank subspace. The approach in [10] was designed to tackle such case by incorporating the low-rank representation(LRR) which assume that the data lies in a union of linear subspaces. By jointly estimating the 3D structure and affinity matrix, it yields both 3D reconstruction and motion clustering of the deformable object. Note that LRR may not achieve satisfactory results when dealing with the data from nonlinear subspaces [11]. Another method[12] relies on making the union of linear subspaces assumption. In specific, they assume the 3D structure is compressible. Nevertheless, the real world motion is usually nonlinear, especially when the motion is complex. As a consequence, making the linear subspaces assumption is not well-suited to modeling such motions.

To this end, we re-formulate the problem by leveraging the kernel trick. The experiments demonstrate that using nonlinear subspaces provides more reasonable clustering results. It proves that the non-linear representation can better model the complex non-rigid motions. Therefore, our method utiliz-

This work was supported in part by Natural Science Foundation of China (No.61231018), National Science and Technology Support Program (2015BAH31F01) and Program of Introducing Talents of Discipline to University under grant B13043.

*Corresponding author.

ing the non-linear low-rank representation yields more accurate reconstruction than start-of-the-art method.

2. OUR APPROACH

Due to the inherent basis ambiguity of the nonrigid problem, NRSFM remains a difficult topic in the field of computer vision. Dai et al. [13] get good results by exploiting the low-rank nature of non-rigid objects. Recent process [11] facilitate our work by exploring the subspace structure of non-rigid data. We propose a novel strategy for estimating the 3D structure, by combining the low-rank constraint and subspace clustering, from an ensemble of 2D projections. We have motivated our approach based on the insight that complex nonrigid motion, such as sequences contains different types of human actions, can be modeled as a union of subspaces.

2.1. FORMULATIONS

By combining the low-rank constraint of shape and nonlinear subspace constraint, we have

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{X},\mathbf{E}} \quad & \|\mathbf{Z}\|_* + \lambda_1 \|\mathbf{X}\|_* + \lambda_2 \|\mathbf{E}\|_1 \\ & + \lambda_3 \|\Phi(\mathbf{X})(\mathbf{I}-\mathbf{Z})\|_F^2 \\ s.t. \quad & \mathbf{W} = \mathbf{R}\mathbf{X}^{\#} + \mathbf{E}, \end{aligned} \tag{1}$$

where λ_1, λ_2 and λ_3 are penalty parameters for $\|\mathbf{X}\|_*$, $\|\mathbf{E}\|_1$ and $\|\Phi(\mathbf{X})\mathbf{H}\|_F^2$ respectively. $\mathbf{Z} \in \mathbb{R}^{F\times F}$ is a coefficients matrix corresponding to subspace clustering. $\mathbf{X} \in \mathbb{R}^{3P\times F}$ is the 3D shape data which stacks all the F frames of measurements and all the P points in a matrix form. $\mathbf{X}^{\#} \in \mathbb{R}^{3F\times P}$ is a re-arrangement of $\mathbf{X}$. $\mathbf{W} \in \mathbb{R}^{2F\times P}$ is the image measurements. $\mathbf{R} = blkdiag(\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_F) \in \mathbb{R}^{2F\times 3F}$ denotes the camera motion matrix, where $\mathbf{R}_i \in \mathbb{R}^{2\times 3}$ is the first two rows of the i-th frame camera rotation. $\|\mathbf{E}\|_1$ is the norm on the reprojection error where $\|\cdot\|_1$ denotes the convex approximation to sparse error. $\mathbf{I} \in \mathbb{R}^{F\times F}$ denotes the identity matrix, and $\Phi(\mathbf{X})$ denotes the kernel-induced mapping from $\mathbf{X}$ to $\Phi(\mathbf{X})$.

Low-rank Reconstruction. The term $\lambda_1 \|\mathbf{X}\|_*$ and the first constrain $\mathbf{W} = \mathbf{R}\mathbf{X}^{\#} + \mathbf{E}$ correspond to the classic reconstruction process [13]. $\lambda_1 \|\mathbf{X}\|_*$ is meant for exploiting the low rank nature of non-rigid object. The constrain term $\mathbf{W} = \mathbf{R}\mathbf{X}^{\#} + \mathbf{E}$ is meant for penalizing re-projection error under orthographic projection. The use of nuclear norm $\|\cdot\|_*$ is a convex approximation to rank. One thing in particular to note is that the nuclear norm of $\mathbf{X} \in \mathbb{R}^{3P\times F}$ is being minimized, rather than $\mathbf{X}^{\#} \in \mathbb{R}^{3F\times P}$ as described in [13]. This is done because the rank of $\mathbf{X}$ is bound by $min(F, 3N)$ whereas the rank of $\mathbf{X}^{\#}$ is bound by $min(3F, N)$. Minimizing the rank of $\mathbf{X}$ is preferable as it attempts to directly learn redundancies between frames.

Subspace Clustering. The coefficients matrix $\|\mathbf{Z}\|_*$ and term $\lambda_3 \|\Phi(\mathbf{X})(\mathbf{I}-\mathbf{Z})\|_F^2$ correspond to the subspace clustering constraint. Given a set of data samples (vectors) approximately drawn from a union of multiple subspaces, our goal is to cluster the samples into their respective subspaces. Low-Rank Representation (LRR) [14] seeks for the low-rank solution of $\mathbf{Z}$ to reveal the intrinsic structure of data. However, LRR may not achieve satisfactory results when dealing with the data from nonlinear subspaces, since it is originally designed to handle the data from linear subspaces in the input space. Meanwhile, the kernel-based methods deal with the nonlinear data by mapping it from the original input space to a new feature space through a kernel-induced mapping. By introducing kernel-induced mapping ($\mathbf{X} \rightarrow \Phi(\mathbf{X})$), the subspace clustering error becomes $\|\Phi(\mathbf{X})(\mathbf{I}-\mathbf{Z})\|_F^2$ respectively. The kernel trick makes us model the nonlinear non-rigid motion data better.

Kernel-induced Mapping. For the given data $\{\mathbf{x}_i\}_{i=1}^F$ in the input space, where $\mathbf{x}_i \in \mathbb{R}^{3P}, \forall i = 1, ..., F$, we define $\mathbf{X} \in \mathbb{R}^{3P\times F}$ as $\mathbf{X} = [x_1, ..., x_F]$. We introduce $\mathbf{K} \in \mathbb{R}^{F\times F}$ to denote the kernel matrix, and introduce $ker(\mathbf{x}, \mathbf{y})$ to denote the kernel function. The $(i, j)th$ element of $\mathbf{K}$ is calculated as follows:

$$\mathbf{K}_{ij} = ker(\mathbf{x}_i, \mathbf{x}_j) \quad \forall i, j = 1, ..., F. \tag{2}$$

The kernel function $ker(\mathbf{x}, \mathbf{y})$ induces a mapping $\mathbf{X} \rightarrow \Phi(\mathbf{X})$. Namely, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{3P}$, we have $ker(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^{'}\phi(\mathbf{y})$. We define $\Phi(\mathbf{X})$ as $\Phi(\mathbf{X}) = [\phi(\mathbf{x}_1), ..., \phi(\mathbf{x}_F)]$, so that we have

$$\mathbf{K} = \Phi(\mathbf{X})^{'}\Phi(\mathbf{X}). \tag{3}$$

By defining a new variable $\mathbf{H} = \mathbf{I} - \mathbf{Z} \in \mathbb{R}^{F\times F}$, we can rewrite $\|\Phi(\mathbf{X})(\mathbf{I}-\mathbf{Z})\|_F^2$ in (1) as $\|\Phi(\mathbf{X})(\mathbf{I}-\mathbf{Z})\|_F^2 = \|\Phi(\mathbf{X})\mathbf{H}\|_F^2 = tr(\mathbf{H}^{'}\mathbf{K}\mathbf{H})$. $tr(\cdot)$ denote the trace operation. Finally, We convert (1) to the following equivalent problem:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{X},\mathbf{E},\mathbf{H}} \quad & \|\mathbf{Z}\|_* + \lambda_1\|\mathbf{X}\|_* + \lambda_2\|\mathbf{E}\|_1 + \lambda_3 tr(\mathbf{H}^{'}\mathbf{K}\mathbf{H}) \\ s.t. \quad & \mathbf{W} = \mathbf{R}\mathbf{X}^{\#} + \mathbf{E},\ \mathbf{H} = \mathbf{I} - \mathbf{Z}, \end{aligned} \tag{4}$$

In summary, we enforce the nuclear norm minimization of $\mathbf{Z}$ matrix to reveal the subspace structure of non-linear 3D data. And, we enforce a global shape constraint for compact representation of non-rigid objects by penalizing the rank of entire non-rigid shape. Note that, when using the linear kernel $\mathbf{K} = \mathbf{X}'\mathbf{X}$, the problem in **Eq.** (4) is reduced to the form in [10], so [10] is a special case of our proposed method when using the linear kernel.

2.2. Optimization

Eq. (4) can be solved using the alternating direction multiplier method (ADMM) efficiently. In particular, we introduce the corresponding Lagrange multiplier $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ and operate on

the augmented Lagrange function $\mathcal{L}$ as follows:

$$\begin{aligned}\min_{\mathbf{Z},\mathbf{X},\mathbf{E},\mathbf{H}} \mathcal{L} = &\|\mathbf{Z}\|_* + \lambda_1 \|\mathbf{X}\|_* + \lambda_2 \|\mathbf{E}\|_1 \\ &+\lambda_3 tr(\mathbf{H}^{'}\mathbf{K}\mathbf{H}) \\ &+<\mathbf{\Gamma}_1, \mathbf{W} - \mathbf{R}\mathbf{X}^{\#} - \mathbf{E}> \\ &+\tfrac{\mu_1}{2} \left\|\mathbf{W} - \mathbf{R}\mathbf{X}^{\#} - \mathbf{E}\right\|_F^2 \\ &+<\mathbf{\Gamma}_2, \mathbf{H} - \mathbf{I} + \mathbf{Z}> \\ &+\tfrac{\mu_2}{2} \left\|\mathbf{H} - \mathbf{I} + \mathbf{Z}\right\|_F^2,\end{aligned} \tag{5}$$

where μ_1 and μ_2 are positive penalty parameters. We split the above problem into four subproblems. The ADMM iteratively updates the individual variable so as to minimize $\mathcal{L}$ while the other variables are fixed. The variables $\mathbf{X}, \mathbf{Z}, \mathbf{E}, \mathbf{H}$ are solved by the following subproblems:

$$\begin{aligned}\mathbf{X}^{k+1} &= \arg\min_{\mathbf{X}} \quad \mathcal{L}(\mathbf{X}^k, \mathbf{Z}, \mathbf{E}, \mathbf{H}), \\ \mathbf{Z}^{k+1} &= \arg\min_{\mathbf{Z}} \quad \mathcal{L}(\mathbf{X}, \mathbf{Z}^k, \mathbf{E}, \mathbf{H}), \\ \mathbf{E}^{k+1} &= \arg\min_{\mathbf{E}} \quad \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{E}^k, \mathbf{H}), \\ \mathbf{H}^{k+1} &= \arg\min_{\mathbf{H}} \quad \mathcal{L}(\mathbf{X}, \mathbf{Z}, \mathbf{E}, \mathbf{H}^k),\end{aligned} \tag{6}$$

where k is the index of iterations. $\mathbf{X}^{k+1}$ and $\mathbf{Z}^{k+1}$ can be solved via the Singular Value Thresholding (SVT) operator [15], and $\mathbf{Z}^{k+1}$ is relaxed following the techniques developed in [16]. $\mathbf{E}^{k+1}$ is solved using the Shrinkage Operator [17]. The sub-problem for $\mathbf{H}^{k+1}$ reaches a least squares problem. The closed-form solution of $\mathbf{H}^{k+1}$ can be derived by setting it's derivative to zero. The Lagrange multipliers $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are updated as:

$$\begin{aligned}\mathbf{\Gamma}_1^{k+1} &= \mathbf{\Gamma}_1^k + \mu_1(\mathbf{W} - \mathbf{R}\mathbf{X}^{\#} - \mathbf{E}), \\ \mathbf{\Gamma}_2^{k+1} &= \mathbf{\Gamma}_2^k + \mu_2(\mathbf{H} - \mathbf{I} + \mathbf{Z}).\end{aligned} \tag{7}$$

Finally, the penalty parameters are updated as:

$$\begin{aligned}\mu_1^{k+1} &= min(\mu_{max}, \rho\mu_1^k), \\ \mu_2^{k+1} &= min(\mu_{max}, \rho\mu_2^k),\end{aligned} \tag{8}$$

where $\rho > 1$ is a scaled factor and μ_{max} limit the max value of μ.

3. EXPERIMENTS

3.1. Setup

In this section, we compare our method against the baseline method: Dai et al.'s approach [13]. To facilitate the comparison, we use the same error metrics as reported in [13].

Extensive experiments are conducted to test the performance of the proposed method. The sequences we have tested are from the 3D CMU Motion Capture (MoCap) dataset [1] and Utrecht Multi-Person Motion (UMPM) benchmark [18]. Using the sequences from these two dataset, the synthetic 2D projections are generated by a randomly rotating orthographic camera, which rotates with the y-axis of camera at a steady speed and is always pointing at the center of the moving object. The proposed method and the baseline method are employed to reconstruct the 3D motions from the synthetic 2D projections.

[1] http://mocap.cs.cmu.edu

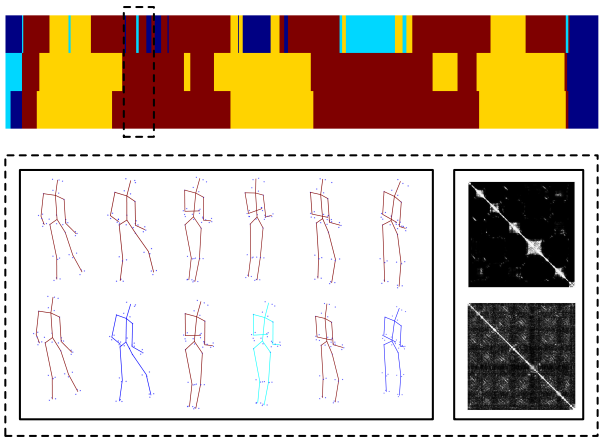
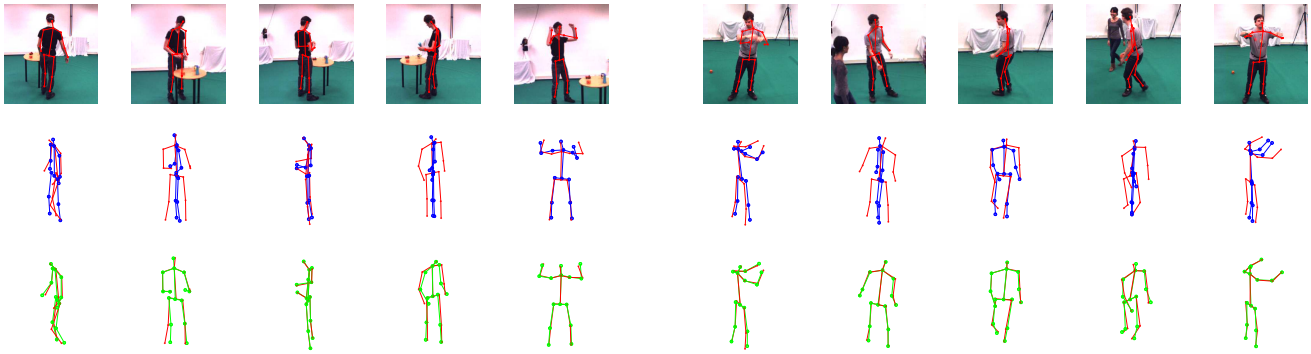


Fig. 1. Subspace analysis on 3D human motions. On the top of this figure, three color bars represent the subspace clustering results. In each bar, the same color means same cluster. From the top down, the clustering results are obtained from the affinity matrices yielded by LRR [14], KLRR [11] and our method respectively. For the first two rows, the LRR and KLRR are applied to the 3D motions 'p1_grab_3' from the UMPM benchmark [18]. For the third row, the affinity matrix produced from the reconstruction process is exploited. On the bottom, two rows of 3D motions and the clustering in a short and consecutive interval are visualized in a dash-line box. The first row is produced by KLRR and the second one is from LRR. The corresponding affinity matrices are shown on the right side.

3.2. Influence of Subspace Structure

Our first experiment is inspired by the motivation that whether the 3D motion data can be fitted better in the kernel space. To illustrate the effectiveness of utilizing LRR [14] and KLRR [11] to recover the low-rank structures of non-rigid object motions, we select the motion sequence 'p1_grab_3' from the UMPM benchmark.

Firstly, we apply subspace clustering on the ground-truth 3D motion data. Fig. 1 shows the affinity matrix $\mathbf{A} = |\mathbf{Z}| + |\mathbf{Z}^T|$ obtained from a human which is undergoing non-rigid deformation. The obtained sparse solution clearly shows that we can get more clear subspace structure in the Gauss-kernel space. And the proposed method can get nearly the same solution as in the Gauss-kernel space when only providing 2D



(a) (b)

Fig. 2. Comparison of the 3D reconstruction results on the sequence of (a) 'p1_grab_3' and (b) 'p3_ball_12'. The first row is the original 2D tracking points. The second and third row denote the reconstruction results by Dai et al.s method and our method respectively. Different colors denote different reconstruction results.

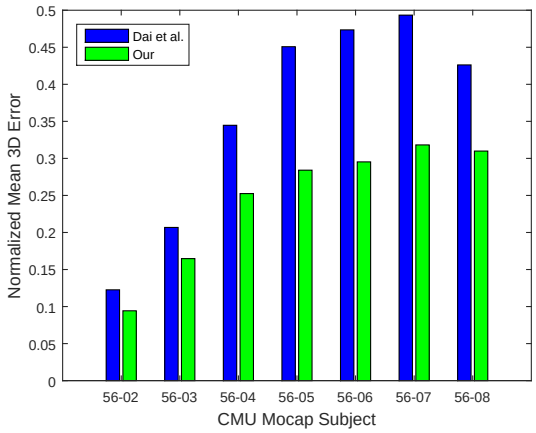


Fig. 3. Quantitative comparison of our proposed method versus Dai et al.s method on the sequences of subject 56 in CMU Mocap database. We generated the 2D points by a moving orthographic camera. The 3D points are reconstructed by Dai et al.s approach and our approach with assuming known the ground truth camera projection matrix.

tracking points.

Then, we apply spectral clustering on the affinity matrix $\mathbf{A}$. The results are shown in the color bars on the top of Fig. 1. As the continuity nature of human actions, the clustering results should have temporal continuity. We can see that the clustering bar produced in Gauss space has better time continuity than the one produced in Linear space. Our method jointly estimates the 3D motions and affinity matrix from the 2D projection. And, the proposed method gets the nearly same clustering result as on the original 3D data in the kernel space (Gauss). The bottom of Fig. 1 show some skeleton examples of clustering results. The similar actions are clustered together in the Gauss-kernel space. As the non-linear nature of human motion, it's not surprising to get this result.

3.3. Compare all methods on all real sequences

In this subsection, we provide experimental results of the methods we are benchmarking.

To illustrate the performance of our method against Dai et al.s method, we visualize the 3D reconstruction on the sample sequences from the Utrecht Multi-Person Motion (UMPM) benchmark [18] in Fig. 2. The ground-truth 3D skeletons are in red, and respectively overlapped with the reconstructed 3D skeletons using Dai et al.'s method (blue) and our method (green). Dai et al.'s method achieves poor performance in some cases, and our approach reconstructs the most accurate 3D nonrigid motion in most scenes.

Fig. 3 summarizes our main results on the sequences of subject 56 in CMU Mocap database, where the shape reconstruction error (mean 3D error) are provided (whenever the ground truth are available). Clearly, our proposed method achieves the best performance in shape recovery.

4. CONCLUSIONS

This paper proposed a novel strategy for long sequence NRSFM which explore the 3D data's subspace structure to improve the reconstruction. We have motivated our approach based on the insight that the 3D motion sequences can be fitted better in the kernel space. We demonstrated the superiority of our approach in comparison to Dai et al.'s state-of-the-art NRSFM approach for reconstructing long non-rigid motion sequences. In future, we plan to extend this work to multi-body NRSFM and aim to handle more complex scenes.

5. REFERENCES

[1] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3d shape from image streams," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2000, pp. 690 – 696.

[2] Carlo Tomasi and Takeo Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[3] Jing Xiao, Jinxiang Chai, and Takeo Kanade, "A closed-form solution to non-rigid shape and motion recovery," *International Journal of Computer Vision*, vol. 67, no. 2, pp. 233–246, 2006.

[4] Ijaz Akhter, Yaser Sheikh, Sohaib Khan, and Takeo Kanade, "Nonrigid structure from motion in trajectory space," in *Advances in neural information processing systems*, 2009, pp. 41–48.

[5] Ijaz Akhter, Yaser Sheikh, and Sohaib Khan, "In defense of orthonormality constraints for nonrigid structure from motion," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1534–1541.

[6] Paulo FU Gotardo and Aleix M Martinez, "Non-rigid structure from motion with complementary rank-3 spaces," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 3065–3072.

[7] Minsik Lee, Jungchan Cho, Chong-Ho Choi, and Songhwai Oh, "Procrustean normal distribution for non-rigid structure from motion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1280–1287.

[8] Tomas Simon, Jack Valmadre, Iain Matthews, and Yaser Sheikh, "Separable spatiotemporal priors for convex reconstruction of time-varying 3d point clouds," in *European Conference on Computer Vision*. Springer, 2014, pp. 204–219.

[9] Yingying Zhu and Simon Lucey, "Convolutional sparse coding for trajectory reconstruction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 3, pp. 529–540, 2015.

[10] Yingying Zhu, Dong Huang, Fernando De La Torre, and Simon Lucey, "Complex non-rigid motion 3d reconstruction by union of subspaces," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1542–1549.

[11] Shijie Xiao, Mingkui Tan, Dong Xu, and Zhao Yang Dong, "Robust kernel low-rank representation," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 11, pp. 2268–2281, 2016.

[12] Chen Kong and Simon Lucey, "Prior-less compressible structure from motion," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4123–4131.

[13] Yuchao Dai, Hongdong Li, and Mingyi He, "A simple prior-free method for non-rigid structure-from-motion factorization," *International Journal of Computer Vision*, vol. 107, no. 2, pp. 101–122, 2014.

[14] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

[15] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[16] Kim-Chuan Toh and Sangwoon Yun, "An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems," *Pacific Journal of Optimization*, vol. 6, no. 615-640, pp. 15, 2010.

[17] Zhouchen Lin, Minming Chen, and Yi Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[18] NP Van der Aa, Xinghan Luo, Geert-Jan Giezeman, Robby T Tan, and Remco C Veltkamp, "Umpm benchmark: A multi-person dataset with synchronized video and motion capture data for evaluation of articulated human motion and interaction," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1264–1269.