Nonrigid Structure From Motion via Sparse Representation

Kun Li, Jingyu Yang, and Jianmin Jiang

***Abstract*—This paper proposes a new approach for nonrigid structure from motion with occlusion, based on sparse representation. We address the occlusion problem based on the latest developments on sparse representation: matrix completion, which can recover the observation matrix that has high percentages of missing data and can also reduce the noises and outliers in the known elements. We introduce sparse transform to the joint estimation of 3-D shapes and motions. 3-D shape trajectory space is fit by wavelet basis to achieve better modeling of complex motion. Experimental results on datasets without and with occlusion show that our method can better estimate the 3-D shapes and motions, compared with state-of-the-art algorithms.**

***Index Terms*—Matrix completion, nonrigid structure from motion (NR-SFM), occlusion, sparse representation, wavelet basis.**

I. Introduction

DYNAMIC scene 3-D reconstruction from image sequences is a classic and hot issue in image/video processing and computer vision. Some methods capture dynamic scenes by constructing a multicamera system [1]–[3] or using a depth camera [4]–[6]. However, high cost, complex maintenance, and carrying inconvenience of these systems are problematic for practical applications. A more efficient way is to estimate shapes and motions from a monocular video sequence, and the key technology is nonrigid structure from motion (NR-SFM). Given a set of corresponding 2-D points in a video sequence of a dynamic scene, the goal of NR-SFM is to recover the 3-D shape and relative camera position for each frame. It has great application potentials in various areas, such as object retrieval [7], [8] and object recognition [9].

Manuscript received February 18, 2014; revised June 23, 2014 and August 21, 2014; accepted August 22, 2014. Date of publication September 8, 2014; date of current version July 15, 2015. This work was supported in part by the National Basic Research Program of China under Grant 2013CB329301, in part by the National Natural Science Foundation of China under Grant 61302059, Grant 61372084, Grant 61228104 and Grant 61373103, and in part by the Tianjin Research Program of Application Foundation and Advanced Technology under Grant 13JCQNJC03900. This paper was recommended by Associate Editor L. Wang.

K. Li is with the Tianjin Key Laboratory of Cognitive Computing and Application, School of Computer Science and Technology, Tianjin University, Tianjin 300072, China (e-mail: lik@tju.edu.cn).

J. Yang is with the School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: yjy@tju.edu.cn).

J. Jiang is with the Research Institute of Future Media Computing, Shenzhen University, Shenzhen 518060, China (e-mail: jianmin.jiang@szu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2014.2351831

Rigid SFM [10] has well developed over the past two decades. For dynamic scenes, especially deformable objects, NR-SFM is still a difficult underconstrained problem due to the absence of prior knowledge. The standard matrix factorization method [11] assumes that all the 3-D shapes can be represented as a linear combination of a compact set of 3-D basis shapes, and the problem is solved by singular value decomposition (SVD). This method needs to estimate the shape basis for each sequence in advance. Akhter *et al.* [12] proposed a trajectory space method that represents the time-varying shapes as a linear combination of a set of basis trajectories, instead of a shape space representation. Discrete cosine transform (DCT) basis is used to compactly describe real motions, without estimating any basis vectors during computation. But this method cannot model high-frequency deformation without increasing the rank of the resulting matrix factors, and is hence restricted to structures with slow and smooth deformation. Considering the occlusion, Gotardo and Martinez [13] proposed a column space fitting method which tolerates missing data, and further improve the approach by explicitly modeling K complementary spaces of rank-3 [14].

NR-SFM is essentially an ill-posed problem to decompose a matrix into two, having infinite number of combinations. The inherent occlusion further complicates the NR-SFM. The key to yield correct decomposition is to provide sufficient prior information for the latent matrices. Sparse representation together with latest advances, i.e., low-rank matrix completion, have shown their power in the regularization of ill-posed estimation problem [2]. In previous work on NR-SFM, motion trajectories are represented by DCT basis. However, DCT basis widely used in harmonic analysis of stationary signals cannot provide compact representation for nonstationary motion trajectories in NR-SFM. We use wavelet basis instead of DCT basis to provide sparser representation of complex motion trajectories so that the ill-posed NR-SFM is well-regularized to have correct decomposition. Besides, the occlusion further underdetermines the NR-SFM, i.e., we have to recover complete shapes from incomplete observations that may be also corrupted by noises and outliers due to the errors in feature tracking and projection assumption. Low-rank matrix completion is a powerful model to recover a complete matrix from an incomplete one. The matrix in NR-SFM is inherent low-rank due to the specific structures of NR-SFM problems. This motives us to stand on the low-rank matrix completion to handle the occlusion problem.

In this paper, we propose a new NR-SFM method based on sparse representation. This paper introduces the sparse representation theory (wavelet transform and low-rank matrix completion) into NR-SFM, and obtains excellent results. The main contributions are twofold. On the one hand, we propose a shape trajectory space fitting method based on wavelet transform. 3-D shape trajectory space is fit by wavelet basis, which achieves better modeling of complex motion. On the other hand, we address the occlusion problem based on low-rank matrix completion, which is efficiently solved by nuclear norm minimization. Through this method, not only the matrix that has high percentages of missing data is recovered, but also the noises and outliers in the known elements are reduced. The proposed method is evaluated on both synthetic and motion capture datasets. Experimental results show that our method can better estimate the 3-D shapes and motions, compared with state-of-the-art algorithms.

The remainder of this paper is organized as follows. Section II reviews related work in SFM. The problem of NR-SFM is formulated in Section III. The proposed approach based on sparse representation is described in Section IV. Validation experiments are presented in Section V. The paper is concluded in Section VI.

II. Related Work

A. SFM With Missing Data

In SFM, the situation with missing data is general due to occlusion and matching failure. One group of related methods that address the missing data problem are known as batch algorithms. These methods propose the strategies that combine partial low-rank factorizations obtained for complete sub-blocks of the measurement matrix, which reconstruct the matrix by first building its row null-space [15], [16], column null-space [17], [18], or one of its range spaces [19]. The main problem of these methods is their sub-optimality and the performance degrades in the presence of noise. The second group of related approaches including iterative methods use all the data at once without searching for complete sub-blocks in the measurement matrix. Alternation methods [20]–[22] iteratively solve for subsets of unknowns while the others remain fixed. Alternation guarantees convergence to a local minimum as the cost function is reduced at each iteration. The convergence rate is fast in first few iterations, but becomes slow later. Furthermore, the algorithms are susceptible to flatlining. The third group of related approaches are nonlinear optimization algorithms that directly optimize the cost function. Buchanan and Fitzgibbon [23] propose a Damped-Newton algorithm, which provides faster and more accurate solutions than standard alternation approaches. Chen [24] uses the Levenberg–Marquardt algorithm to solve the missing data problem as a minimization on subspaces. Despite its superior performance, proper initialization remains an open problem.

The blooming of sparse representation inspires the theory foundation of recovering low-rank data from a highly-incomplete set of (possibly corrupted) entries, which is referred to as matrix completion [25], [26]. This brings new opportunities to fully exploit the low-rank structure in SFM. In this paper, we address the occlusion problem based on low-rank matrix completion, which can recover the observation matrix that has high percentages of missing data and can also reduce the noise and outliers in the known elements.

B. NR-SFM

SFM can be classified into rigid SFM and NR-SFM, according to the rigidity of structures. Rigid SFM is considerably matured over the past two decades [27]–[29], even for large-scale scenes [30]–[32], while NR-SFM is still a difficult problem.

To make the NR-SFM problem tractable, recent work has attempted to define new general constraints for 3-D shape deformation. Bregler *et al.* [11] impose this constraint as a linear shape basis. They assume that all the 3-D shapes can be represented as a linear combination of 3-D basis shapes, and recover the structure, the shape basis, and the camera rotations simultaneously by exploiting orthonormality constraints of the rotation matrices. Xiao *et al.* [33] discuss the ambiguity in the solution provided by the orthonormality constraints alone, and introduce additional constraints to remove the ambiguity. In the same way, Torresani *et al.* [34] adopt the rank constraint to assist in tracking and modeling non-rigid objects. Further, they introduce a Gaussian prior on the shape at each time instant [35], and a hierarchical prior by using a probabilistic principal components analysis (PPCA) shape model [22]. Bartoli *et al.* [36] use three kinds of priors to achieve high quality reconstruction: a coarse-to-fine ordering of the deformation modes, temporal smoothness, and spatial smoothness. Different from conventional linear basis interpretation, Rabaud and Belongie [37] propose to learn a smooth manifold of shape configurations from video, which is more intuitive and flexible. Del Bue *et al.* [38] also propose a nonlinear optimization method, which uses an adaptive 2-D point-tracking algorithm based on ranklets. But this kind of methods is time consuming.

For most of the above methods, the shape basis is specific, i.e., the shape basis varies for different sequences. Akhter *et al.* [12] propose a new approach for NR-SFM, which is based on the trajectory basis and is shape agnostic. They represent the time-varying shapes as a linear combination of a set of basis trajectories, instead of a shape space representation. DCT basis is used to compactly describe most real motions, without estimating any basis vectors during computation. But this method cannot model high-frequency deformation without increasing the rank of the resulting matrix factors, and is hence restricted to structures with slow and smooth deformation. Park *et al.* [39] demonstrate that the accuracy of reconstruction is fundamentally limited by the correlation between the trajectory of the point and the trajectory of successive camera centers. Poor correlation leads to good reconstruction, and high correlation leads to poor reconstruction. They also define a criterion, called reconstructibility, to determine the accuracy of reconstruction. Zhu *et al.* [40] show that sequences with poor reconstructibility could be salvaged by injecting rigid keyframes. But this method is specifically for human bodies. Recently, Gotardo and Martinez [13], [14] combine shape and trajectory basis approaches, explicitly modeling K complementary spaces of rank-3. They describe the shape basis coefficients with a DCT basis over time.

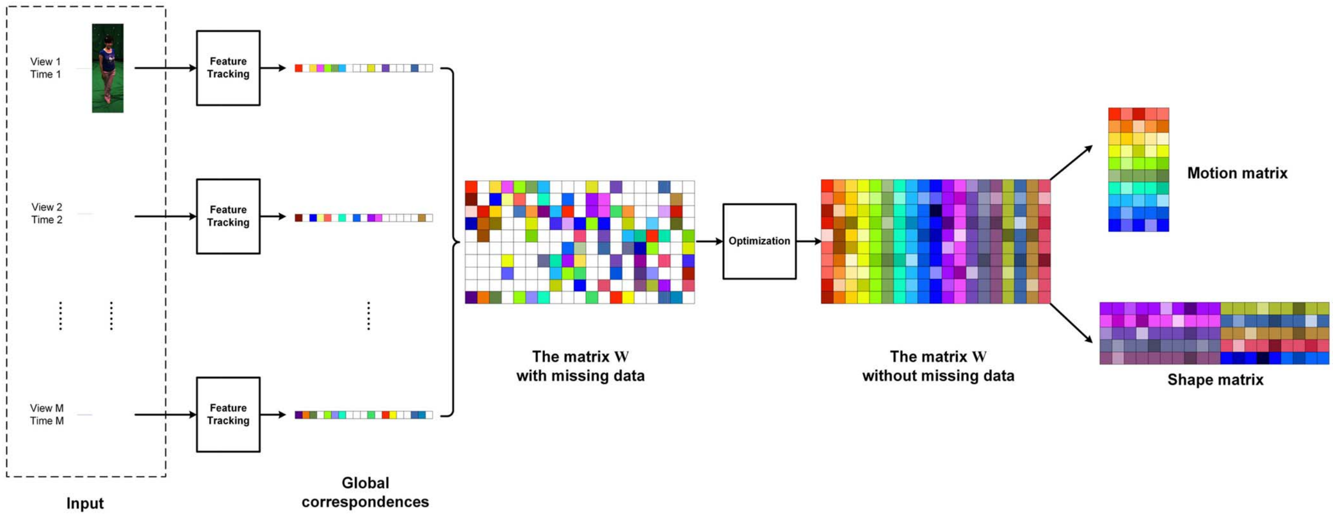

Fig. 1. Workflow of our method.

DCT basis is powerful in representing stationary signals, but is difficult to accurately represent the trajectories of very complex motions that are nonstationary signals. On the contrary, wavelet basis is quite suitable in analyzing nonstationary signals due to its excellent capability in capturing isolated singularities. With careful design, wavelet transform can give very sparse representations for a wide range of nonstationary signals. In this paper, we use sparse transform (wavelet transform) to accurately fit the space of complex motion trajectories in nonrigid structure from motion.

III. Problem Formulation

Fig. 1 illustrates the workflow of our proposed method. The input data is a single video captured by a camera moving across multiple views. N global correspondences $E = \{e_n\}_{1 \le n \le N}$ for M images are established by feature tracking methods, e.g., the Kanade–Lucas–Tomasi (KLT) feature tracker [41]. Each global correspondence e_n is represented as $\{(\gamma_i, \mathbf{P}^n_{\gamma_i})\}_{1 \le i \le V_n}$, where γ_i denotes the associated view index, $\mathbf{P}^n_{\gamma_i}$ is the pixel location (x, y) of e_n in the image γ_i, and V_n is the number of associated cameras in correspondence e_n. Given the global correspondences for all the views, an observation matrix $\mathbf{W} \in \mathbb{R}^{2M \times N}$ can be formed by stacking these correspondences together. Each two rows of $\mathbf{W}$ correspond to the same view, and each column of $\mathbf{W}$ corresponds to the same global correspondence. Considering the case with occlusion, each global correspondence e_n has less than M associated cameras ($V_n = M$), i.e., the matrix $\mathbf{W}$ is incomplete with missing data. We recover the matrix based on rank minimization (Section IV-A), and jointly determine the dynamic 3-D shapes and the relative camera motions with sparse transform (Section IV-B).

The matrix $\mathbf{W}$ is known to have a low-rank $r \le \min(2M, N)$ [28], and can be factorized by standard matrix factorization methods into two parts

$$\mathbf{W} = \widetilde{\mathbf{M}}\widetilde{\mathbf{S}}, \qquad \widetilde{\mathbf{M}} \in \mathbb{R}^{2M \times r},\ \widetilde{\mathbf{S}} \in \mathbb{R}^{r \times N} \tag{1}$$

where $\widetilde{\mathbf{S}}$ contains K ($r = 3K+1$) basis shapes, and $\widetilde{\mathbf{M}}$ includes the camera motions and the shape coordinates in terms of basis $\widetilde{\mathbf{S}}$. More generally, an additional mean column $\mathbf{t} \in \mathbb{R}^{2M}$ is first calculated and the model in (1) can be rewritten as

$$\mathbf{W} = \mathbf{MS} + \mathbf{t1}^T \tag{2}$$

where $\mathbf{t} \in \mathbb{R}^{2M}$ is the mean column of $\mathbf{W}$, $\mathbf{M} \in \mathbb{R}^{2M \times 3K}$, $\mathbf{S} \in \mathbb{R}^{3K \times N}$, and $\mathbf{1} \in \mathbb{R}^N$ is a vector of all ones.

IV. NR-SFM via Sparse Representation

In this section, we first give a solution for the case with occlusion. The complete observation matrix is recovered via low-rank matrix completion. Then, 3-D shapes and motions are jointly estimated with wavelet basis.

A. Recovering Observation Matrix via Matrix Completion

For the problem with occlusion, the matrix $\mathbf{W}$ has missing data, and the known entries in $\mathbf{W}$ may be corrupted by noises due to errors in feature tracking and projection assumption. This is common in most practical applications.

To recover the complete matrix, the only information available about $\mathbf{W}$ is a sampled set of entries W_{ij}, $(i, j) \in \Omega$, where Ω is a subset of the complete set of entries $[2M] \times [N]$ ($[n]$ denotes the list $\{1, \ldots, n\}$). Considering the noises and outliers, the following observation model is employed:

$$W_{ij} = W^0_{ij} + N_{ij}, \qquad (i, j) \in \Omega \tag{3}$$

where $\{N_{ij} : (i, j) \in \Omega\}$ denotes additive white Gaussian noise. Equation (3) can also be expressed as

$$\mathcal{P}_\Omega(\mathbf{W}) = \mathcal{P}_\Omega\left(\mathbf{W}^0\right) + \mathcal{P}_\Omega(\mathbf{N}) \tag{4}$$

where $\mathbf{N}$ is a $2M \times N$ matrix with entries N_{ij} for $(i, j) \in \Omega$ (the values of entries outside Ω are irrelevant), and $\mathcal{P}_\Omega(\mathbf{W}) : \mathbb{R}^{2M \times N} \to \mathbb{R}^{2M \times N}$ is a sampling operator defined by

$$[\mathcal{P}_\Omega(\mathbf{W})]_{ij} = \begin{cases} W_{ij} & (i, j) \in \Omega \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Given that the latent matrix is low-rank, the problem can be formulated as

$$\begin{aligned} &\text{minimize} && \operatorname{rank}\left(\widetilde{\mathbf{W}}\right) \\ &\text{subject to} && \left\|\mathcal{P}_\Omega\left(\widetilde{\mathbf{W}} - \mathbf{W}\right)\right\|_F \le \delta \end{aligned} \tag{6}$$

where $\widetilde{\mathbf{W}} \in \mathbb{R}^{2M\times N}$ is the desired complete matrix, and $\|\mathbf{A}\|_F$ represents the Frobenious norm of a matrix $\mathbf{A}$.

Analogous to the intractability of ℓ_0-minimization in sparse signal recovery, this rank-minimization problem is NP-hard. But Candès and Recht [25] prove that most low-rank matrices can be perfectly recovered by solving the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \|\widetilde{\mathbf{W}}\|_* \\ \text{subject to} \quad & \|\mathcal{P}_\Omega(\widetilde{\mathbf{W}} - \mathbf{W})\|_F \le \delta \end{aligned} \tag{7}$$

where the functional $\|\widetilde{\mathbf{W}}\|_*$ is the nuclear norm of the matrix $\widetilde{\mathbf{W}}$, i.e., the sum of its singular values. This optimization problem is convex and can be recast as a semidefinite programming [42]. Candès *et al.* [26] prove that the recovery is accurate under the assumption that $\|\mathcal{P}_\Omega(\mathbf{N})\|_F \le \delta$ for some $\delta > 0$.

Instead of solving (7) directly, we solve its Lagrangian version

$$\text{minimize} \quad \frac{1}{2}\left\|\mathcal{P}_\Omega(\widetilde{\mathbf{W}} - \mathbf{W})\right\|_F^2 + \mu\left\|\widetilde{\mathbf{W}}\right\|_*. \tag{8}$$

The value of μ is picked large enough to threshold away the noises (keep the variance low), and small enough not to overshrink the original matrix (keep the bias low). In our method, we set $\mu = (\sqrt{M} + \sqrt{N})\sqrt{p}\delta$ as suggested in [26], where p is the ratio of the number of known entries over the total number of entries in the matrix. We use the fixed point iterative algorithm [43] to solve the nuclear norm minimization problem in (8).

The data term reflects the fidelity of the reconstructed matrix, and ensures the accuracy of the reconstruction. The regularization term takes the low-rankness into account, and ensures the smoothness of the reconstruction. By bridging the data and regularization terms with a penalization parameter and then minimizing it, not only the unknown elements are recovered, but also the noises and outliers in the known elements are reduced.

B. Solving for Shape Trajectory With Wavelet Basis

In this section, we incorporate wavelet basis representation into the shape trajectory estimation problem to achieve better modeling of complex motions. In previous work on NR-SFM [12]–[14], DCT is used to fit shape trajectories in matrix-factorization- based methods. Equipping with global sinusoids of different frequencies, DCT can concentrate most signal energy into few coefficients, which thus provides a sparse representation. However, DCT has poor analysis resolution in time domain. When signals contain nonstationary features such as isolate singularities and burst spikes, the representation would spread over many coefficients. With scaled and translated local basis functions, wavelet transforms have stronger capabilities in capturing nonstationary characteristics of signals due to their good balance in time-frequency localization. For this merit, wavelet transforms have been exploited with great success across various fields, e.g., signal processing [44], image compression [45], and computer vision [46].

The matrix $\mathbf{M}$ in (2) is composed of two parts: a block-diagonal rotation matrix $\mathbf{D} \in \mathbb{R}^{2M\times 3M}$ and a shape coordinate matrix $\mathbf{C} \in \mathbb{R}^{M\times K}$, i.e., $\mathbf{M} = \mathbf{D}(\mathbf{C} \otimes \mathbf{I}_3)$, where $\mathbf{I}_3$ is a 3×3 identity matrix and $\otimes$ is the Kronecker product. The diagonal block-elements of $\mathbf{D}$ are M rotation matrices of size 2×3. Each row of $\mathbf{C}$ has the coordinates of 3-D shape in the corresponding image with respect to the shape basis in $\mathbf{S}$, and hence can be considered as a single point in shape space [13]. The 3-D shape trajectory over time can be represented with wavelet basis

$$\mathbf{C} = \mathbf{\Phi}\mathbf{X}, \qquad \mathbf{\Phi} \in \mathbb{R}^{M\times M},\ \mathbf{X} \in \mathbb{R}^{M\times K} \tag{9}$$

where $\mathbf{\Phi}$ is the wavelet basis matrix with J-level decomposition. Each column of $\mathbf{X}$ has the coordinates for the corresponding column of $\mathbf{C}$ as represented in the M-dimensional space spanned by $\mathbf{\Phi}$. For more efficient computation, the wavelet basis matrix $\mathbf{\Phi}$ can be truncated to be a $M \times T$ $(T < M)$ matrix, keeping T wavelet basis functions, i.e., the coefficient matrix $\mathbf{X} \in \mathbb{R}^{T\times K}$.

Assuming $\widetilde{\mathbf{W}}$ is complete, $\mathbf{t}$ is estimated as the mean column of $\widetilde{\mathbf{W}}$, and $\mathbf{D}$ is computed by iteratively using the trajectory space method [12] until no improvement is obtained in the average camera orthonormality. The coefficient matrix $\mathbf{X}$ is initialized with $\begin{bmatrix} \mathbf{I}_K \\ \mathbf{0} \end{bmatrix}$ and calculated by minimizing the following energy function with Damped-Newton optimization method:

$$\min f(\mathbf{X}) = \frac{1}{2}\sum_{j=1}^{N}\left(\mathbf{P}^\perp\left(\widetilde{\mathbf{w}}_j - \mathbf{t}\right)\right)^T\left(\mathbf{P}^\perp\left(\widetilde{\mathbf{w}}_j - \mathbf{t}\right)\right) \tag{10}$$

where $\widetilde{\mathbf{w}}_j$ is the jth column of $\widetilde{\mathbf{W}}$ and

$$\mathbf{P}^\perp = \mathbf{I} - (\mathbf{D}(\mathbf{\Phi}\mathbf{X} \otimes \mathbf{I}_3))(\mathbf{D}(\mathbf{\Phi}\mathbf{X} \otimes \mathbf{I}_3))^\dagger$$

where $\dagger$ denotes the Moore–Penrose pseudo-inverse. Specifically, at each iteration, the current estimate of $\mathbf{X}$ is updated by computing an adjustment matrix $\triangle\mathbf{X}$ in vectorized form $\text{vec}(\triangle\mathbf{X})$. The vector $\text{vec}(\triangle\mathbf{X})$ is solved using the gradient vector $\mathbf{g}$ and Hessian matrix $\mathbf{H}$ of f by matrix differential calculus [47]. Then, $\mathbf{M}$ and $\mathbf{S}$ are solved by $\mathbf{M} = \mathbf{D}(\mathbf{\Phi}\mathbf{X} \otimes \mathbf{I}_3)$ and $\mathbf{S} = \mathbf{M}^\dagger(\widetilde{\mathbf{W}} - \mathbf{t}\mathbf{1}^T)$, respectively. The recovered shapes are $\widehat{\mathbf{S}} = (\mathbf{\Phi}\mathbf{X} \otimes \mathbf{I}_3)\mathbf{S}$.

To verify the potential of wavelets in shape trajectory representation, we compare the representation efficiency between DCT and discrete wavelet transform (DWT) on dataset *Face2*. As shown in Fig. 2(a), the shape coordinates over time present nonstationary time-varying characteristics with burst jumps. We apply DCT and DWT to the shape coordinate matrix $\mathbf{C}$ along the column (i.e., the temporal). Fig. 2(b) shows the energy compaction efficiency in terms of normalized energy with respect to the percentage of retained largest coefficients. The curves in Fig. 2(b) indicate that the wavelet transform has significantly higher approximation power than DCT in representing complex motion trajectories. This justifies the use of wavelet transforms in NR-SFM.

V. Experimental Results

In this section, we evaluate the performance of the proposed method on the datasets without missing data (Section V-A) and datasets with missing data (Section V-B). We discuss

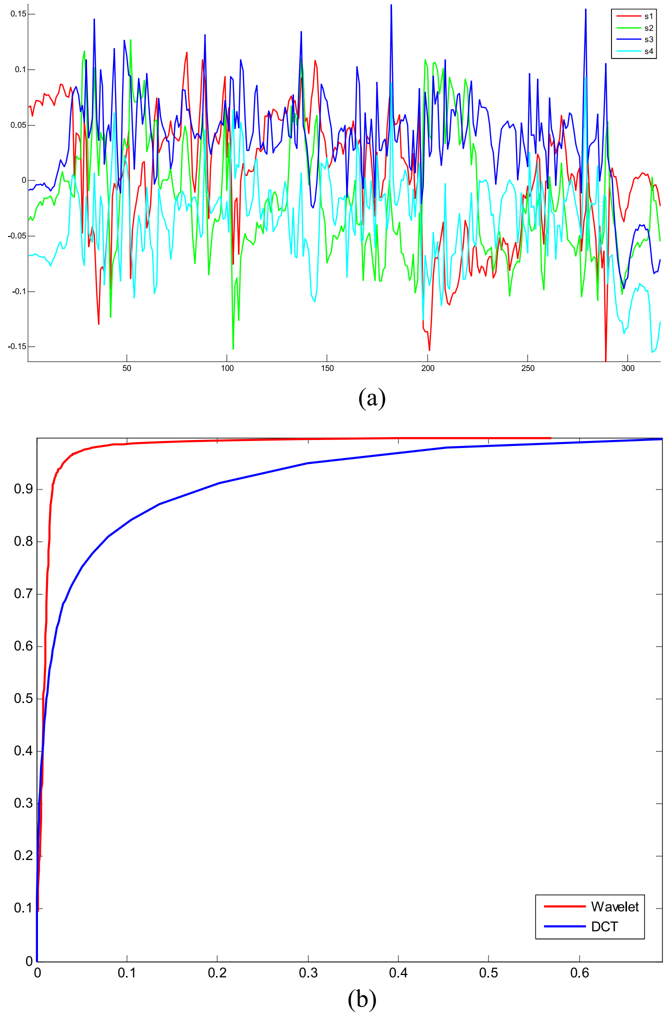

Fig. 2. Performance comparison of DCT and DWT on dataset *Face2*. (a) Coordinates of 3-D shape at each time instant with respect to four shape bases in **S**. (b) Energy compaction efficiency in terms of normalized energy with respect to the percentage of retained largest coefficients.

the performances of using various wavelet transforms in Section V-C.

A. NR-SFM Results Without Missing Data

We first evaluate the performance of our method on two synthetic datasets: *Shark1* (240/91) [22] and *Shark2* (240/91) [12], and four motion capture datasets: *Face1* (74/37) [21], *Face2* (316/40) [22], *Walking* (260/55) [22], and *Dance* (264/75) [12]. (M/N) after each dataset's name indicates the number of images (M) and the number of 3-D points (N). Six state-of-the-art methods are compared: PTA [12], CSF1 [13], CSF2 [14], LSSM1 [48], LSSM2 [48], and BMM [49]. The source codes of all the methods are provided by the corresponding authors. In our experiments, we employ Daubechies 10 wavelet basis to fit the 3-D shape trajectory space. The decomposition levels are set to be 4, 3, 2 for *Shark1/2, Dance, Walking/Face1/2,* respectively.

Table I gives the quantitative evaluation results. In the table, K is the number of basis shapes, RMSE represents the root mean square error between the reconstructed matrix $\widetilde{\mathbf{M}}\widetilde{\mathbf{S}}$ and the input matrix $\mathbf{W}$, e_{2D} is the maximum 2-D reprojection error in pixel units, and e_{3D} is the normalized mean 3-D error over all points and images computed as

$$e_{3D} = \frac{1}{\overline{\sigma} MN} \sum_{m=1}^{M} \sum_{n=1}^{N} e_{mn} \tag{11}$$

TABLE I
QUANTITATIVE EVALUATION OF NR-SFM METHODS ON SIX DATASETS

Dataset	Method	K	RMSE	e_{2D}	e_{3D}
Shark1	PTA	9	0.15192	3.61823	0.17958
	CSF1	3	0.03120	1.01626	0.00808
	CSF2	5	0.02996	1.01520	0.04490
	LSSM1	3	0.36536	4.06580	0.13295
	LSSM2	27	0.09039	1.16210	0.12680
	BMM	4	0.99629	5.38578	0.52708
	Ours	3	**0.00013**	**0.00128**	**0.00001**
Shark2	PTA	2	0.18585	0.99712	0.31208
	CSF1	2	0.07053	0.60373	0.25387
	CSF2	3	0.00302	0.07426	0.00520
	LSSM1	3	0.02680	0.34932	0.11224
	LSSM2	27	0.00662	0.20844	0.31994
	BMM	4	1.28826	4.18561	0.62280
	Ours	3	**0.000004**	**0.00004**	**0.00018**
Face1	PTA	2	2.88624	13.32130	0.10825
	CSF1	5	0.69651	7.40142	0.06368
	CSF2	5	0.82459	8.72875	**0.05200**
	LSSM1	5	1.04472	5.97012	0.06789
	LSSM2	27	0.63446	4.28733	0.06850
	BMM	7	1.36008	4.24464	0.05857
	Ours	5	**0.52413**	**3.02690**	0.06248
Face2	PTA	5	0.96073	13.83020	0.04414
	CSF1	3	0.69681	7.52201	0.03627
	CSF2	5	0.59967	7.97482	0.03192
	LSSM1	5	0.51858	7.23383	0.02301
	LSSM2	27	0.14076	7.60027	0.03487
	BMM	7	0.45498	**2.14363**	0.03027
	Ours	4	**0.00097**	7.22362	**0.02144**
Walking	PTA	2	44.54224	370.94500	0.68234
	CSF1	2	19.07181	163.40800	0.18630
	CSF2	5	6.65805	83.74850	0.10495
	LSSM1	5	7.91027	92.68730	0.14794
	LSSM2	27	1.76707	34.2073	0.26665
	BMM	8	**1.76492**	**4.87614**	0.12985
	Ours	4	6.04000	74.74300	**0.10271**
Dance	PTA	5	0.09306	0.68469	0.29584
	CSF1	2	0.15109	0.86895	0.2705
	CSF2	7	0.02936	0.17881	0.19483
	LSSM1	7	0.04030	0.22394	**0.17639**
	LSSM2	27	0.00919	0.17146	0.21806
	BMM	10	2.29392	6.56087	0.18639
	Ours	5	**0.00822**	**0.15688**	0.18591

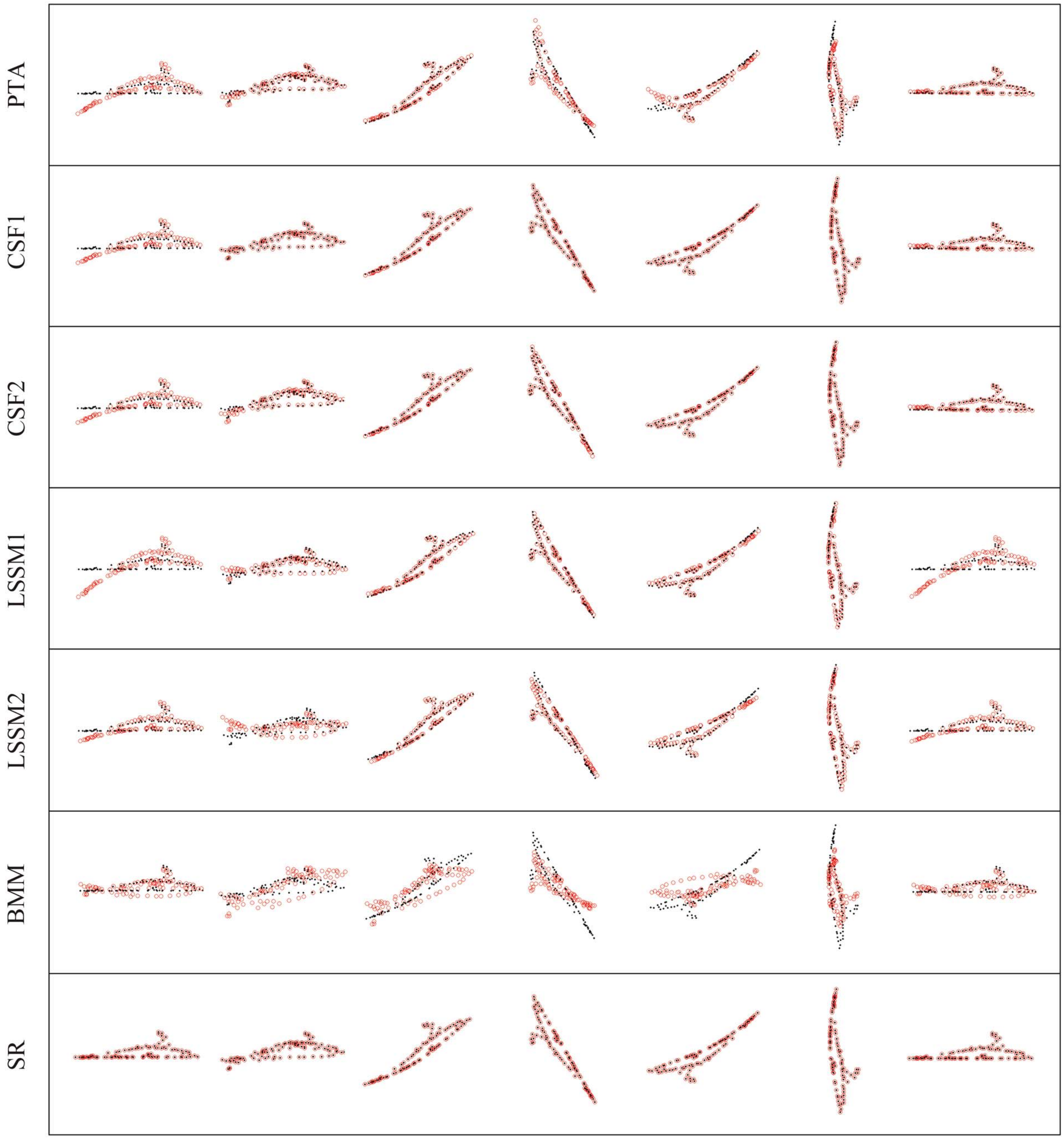

Fig. 3. 3-D shapes of *Shark1* at t_1, t_{40}, t_{80}, t_{120}, t_{160}, t_{200}, and t_{240} recovered by PTA [12], CSF1 [13], and CSF2 [14], LSSM1 [48], LSSM2 [48], BMM [49], and our method (SR).

with

$$\overline{\sigma} = \frac{1}{M}\sum_{m=1}^{M}\sigma\left(\mathbf{S}_m\right) \tag{12}$$

where e_{mn} is the reconstruction error (i.e., Euclidean distance) for the nth 3-D point of image m, and $\mathbf{S}_m \in \mathbb{R}^{3\times N}$ is the original 3-D shape in image m. Let σ_r be the standard deviation of the available entries in the rth row of $\sigma(\mathbf{S}_m)$

$$\sigma\left(\mathbf{S}_m\right) = \frac{1}{3}\sum_{r=1}^{3}\sigma_r. \tag{13}$$

As shown in Table I, our method significantly outperforms PTA, CSF1, and LSSM2 methods, shows better performance than BMM method with smaller 3-D error and lower rank for *Face2* and *Walking* datasets, and gives better results than CSF2 and LSSM1 methods in terms of maximum 2-D reprojection error and RMSE for *Face1* and *Dance* datasets. Especially for *Shark1*, our method achieves nearly perfect reconstruction ($e_{3D} = 0.00001$) with only 10% basis functions, while CSF1 method has $e_{3D} = 0.00004$ with a full DCT basis [13]. Overall, our method achieves better comprehensive performance in term of 3-D error, 2-D reprojection error, and matrix rank.

3-D shapes of *Shark1* at seven time instants recovered by seven approaches are shown in Fig. 3. The reconstructed 3-D shapes are illustrated as red circles, and the ground truth 3-D

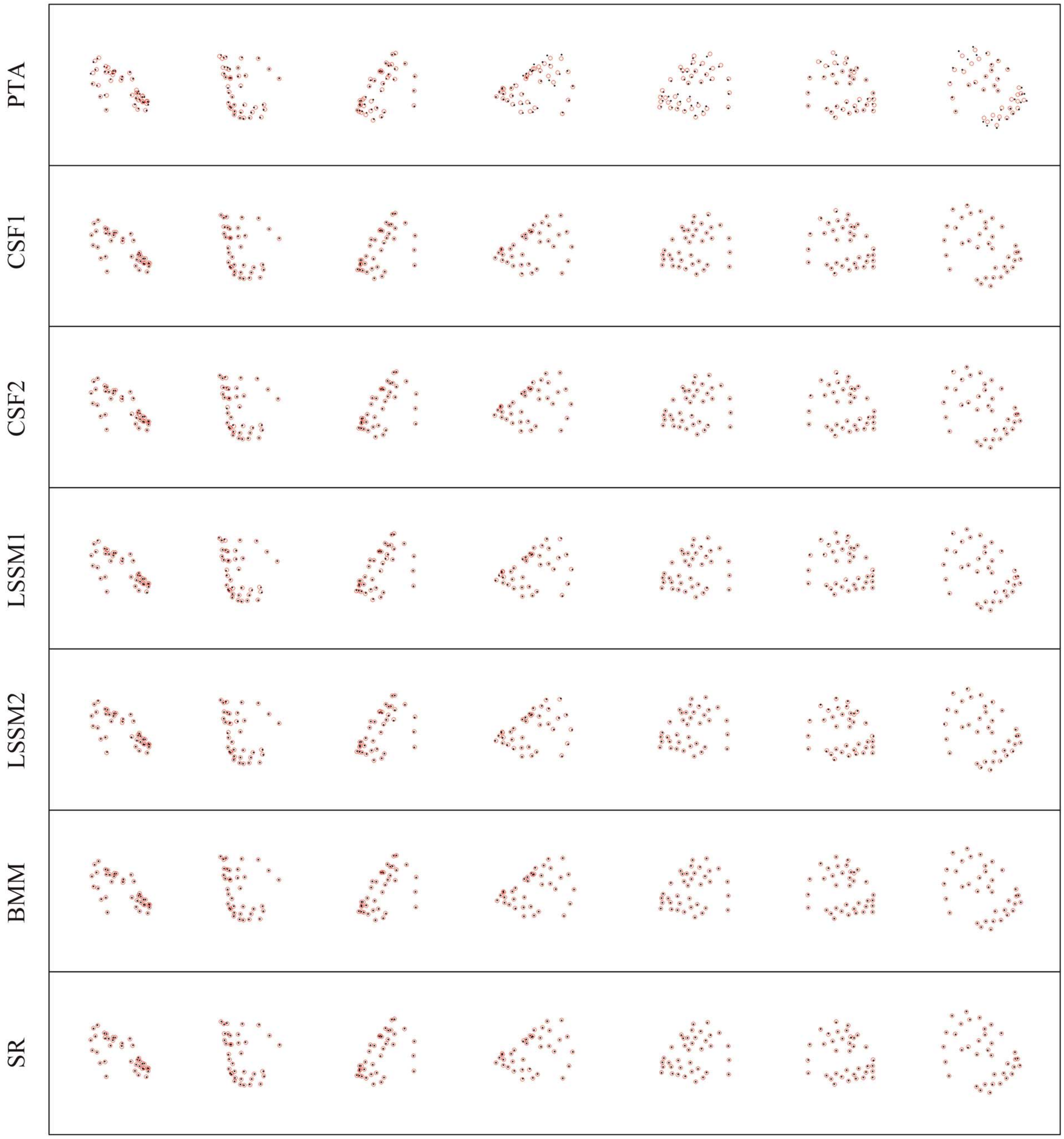

Fig. 4. 3-D shapes of *Face1* at t_1, t_{13}, t_{25}, t_{37}, t_{49}, t_{61}, and t_{74} recovered by PTA [12], CSF1 [13], and CSF2 [14], LSSM1 [48], LSSM2 [48], BMM [49], and our method (SR).

data are shown as dark dots. Only 10% of M basis functions are used for CSF1, CSF2, LSSM1, LSSM2, and our method (SR). It can be seen that our method achieves better reconstruction than the other methods, which benefits from the ingenious use of wavelet basis. The excellent capability in capturing isolated singularities of wavelet basis makes it can accurately fit the space of complex motion trajectories in NR-SFM. Similar phenomena can also be observed in Fig. 4, which gives the reconstructed shapes at seven time instants for *Face1*. The results suggest that our method has better visual quality of reprojection than the other methods, although CSF2 method has smaller 3-D error, which demonstrates that our method accurately recovers not only the 3-D shape but also the camera motions.

We also apply our method to the real dataset *Cubes* (200/14) [22]. Fig. 5 gives the results of six time instants at two views. Because there is no 3-D ground truth data available for this dataset, we only show overlays of the results given by CSF1 [13] and our method. With $K = 2$, the solution of our method has the mean (maximum) 2-D reprojection error of 0.3436 (1.65017) pixel. The solution of CSF1 has an error of 0.4958 (2.0672) pixel also with $K = 2$. It can been seen that our method can better model the changes in speed of the cube being pulled by a string. This demonstrates the effectiveness

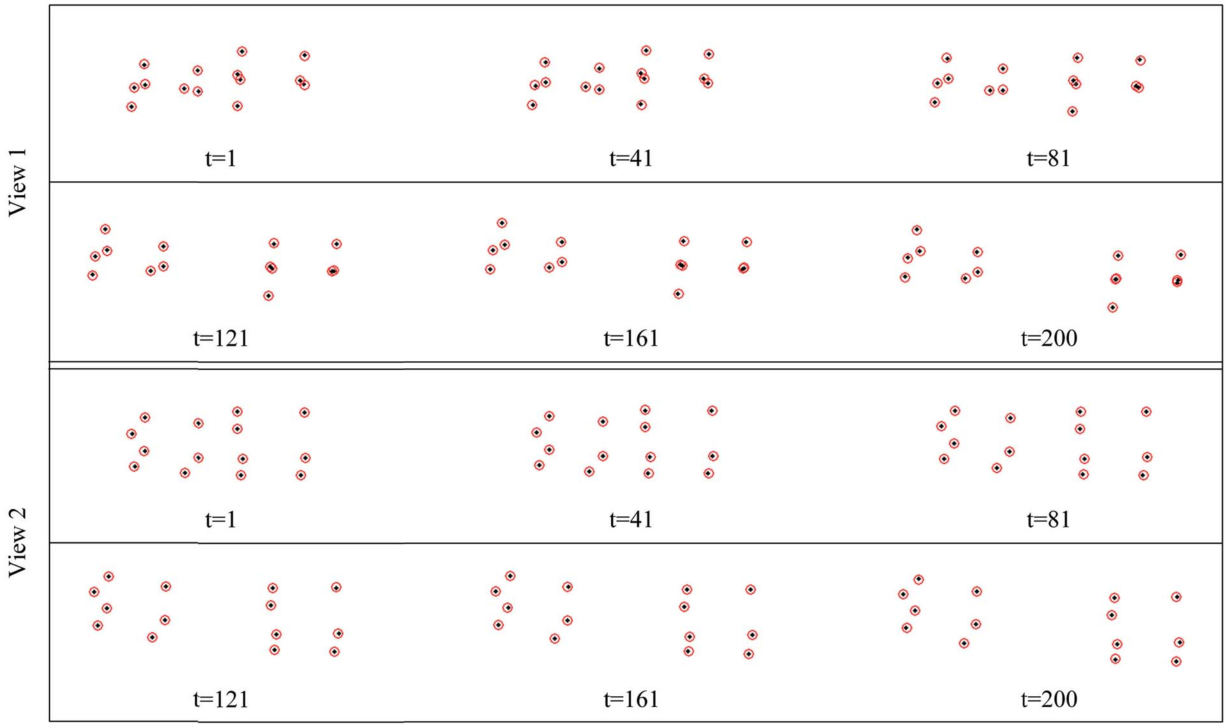

Fig. 5. 3-D shapes of *Cubes* at t_1, t_{41}, t_{81}, t_{121}, t_{161}, and t_{200} recovered by CSF1 [13] (dark dots) and our method (red circles).

TABLE II
QUANTITATIVE EVALUATION OF NR-SFM WITH MISSING DATA FOR *Shark1*

Missing data	Method	K	RMSE	e_{2D}	e_{3D}
25%	CSF1	3	0.03632	1.04442	0.00999
	CSF2	5	0.03981	1.10334	0.08714
	Ours	3	**0.00015**	**0.00162**	**0.00002**
50%	CSF1	3	0.03179	1.11822	0.00924
	CSF2	5	0.04030	1.31023	0.08217
	Ours	3	**0.00022**	**0.00415**	**0.00002**
75%	CSF1	3	0.02209	1.18051	0.00905
	CSF2	5	0.03759	0.91823	0.10494
	Ours	3	**0.02183**	**0.90602**	**0.00106**

TABLE III
QUANTITATIVE EVALUATION OF NR-SFM WITH MISSING DATA FOR *Face2*

Missing data	Method	K	RMSE	e_{2D}	e_{3D}
25%	CSF1	3	0.94335	10.10703	0.04998
	CSF2	5	0.72534	10.56260	0.06207
	Ours	4	**0.24609**	**2.88132**	**0.04769**
50%	CSF1	3	0.74176	9.18713	0.05217
	CSF2	5	0.68770	7.36124	0.06673
	Ours	4	**0.26714**	**2.58031**	**0.05196**
75%	CSF1	3	0.48475	5.96312	0.06845
	CSF2	5	0.56316	5.14008	0.08622
	Ours	4	**0.37072**	**4.17591**	**0.06759**

of the proposed shape trajectory space fitting method based on wavelet transform.

B. NR-SFM Results With Missing Data

To evaluate the performance with occlusion, we simulate the missing data in *Shark1* by randomly discarding 25%, 50%, and 75% of the elements in $\mathbf{W}$. The quantitative evaluation results are shown in Table II, compared with CSF1 [13] and CSF2 [14] because other methods are not proposed for the case with missing data. It can been seen that our method achieves the most accurate results. Assuming that $\mathbf{W}_g$ is the complete ground-truth matrix and $\widetilde{\mathbf{W}}$ is the recovered matrix, the relative recovery error $\|\mathbf{W}_g - \widetilde{\mathbf{W}}\|_F / \|\mathbf{W}_g\|_F$ is $4.62e^{-6}$, $8.11e^{-6}$, and $1.35e^{-3}$ for 25%, 50%, and 75% cases, respectively. The small errors demonstrate that the proposed method for NR-SFM with occlusion recovers nearly the same matrix as the ground-truth. Moreover, the average 3-D errors of our method are still smaller than those of the other methods on the complete data. Fig. 6 gives an example of an occlusion pattern in $\mathbf{W}$ and the corresponding 3-D shape reconstructions on the *Shark1* dataset with 25%, 50%, and 75% missing data, respectively. The left-most plot shows the available entries (in red) of the observation matrix. The other plots show the reconstructed 3-D shapes (red circles) against the original ground-truth data (dark dots). The recovered 3-D shapes are accurate and visually similar to those obtained from the original complete $\mathbf{W}$ (see Fig. 3).

Another example of NR-SFM with missing data for the *Face2* dataset is given in Table III and Fig. 7. The relative

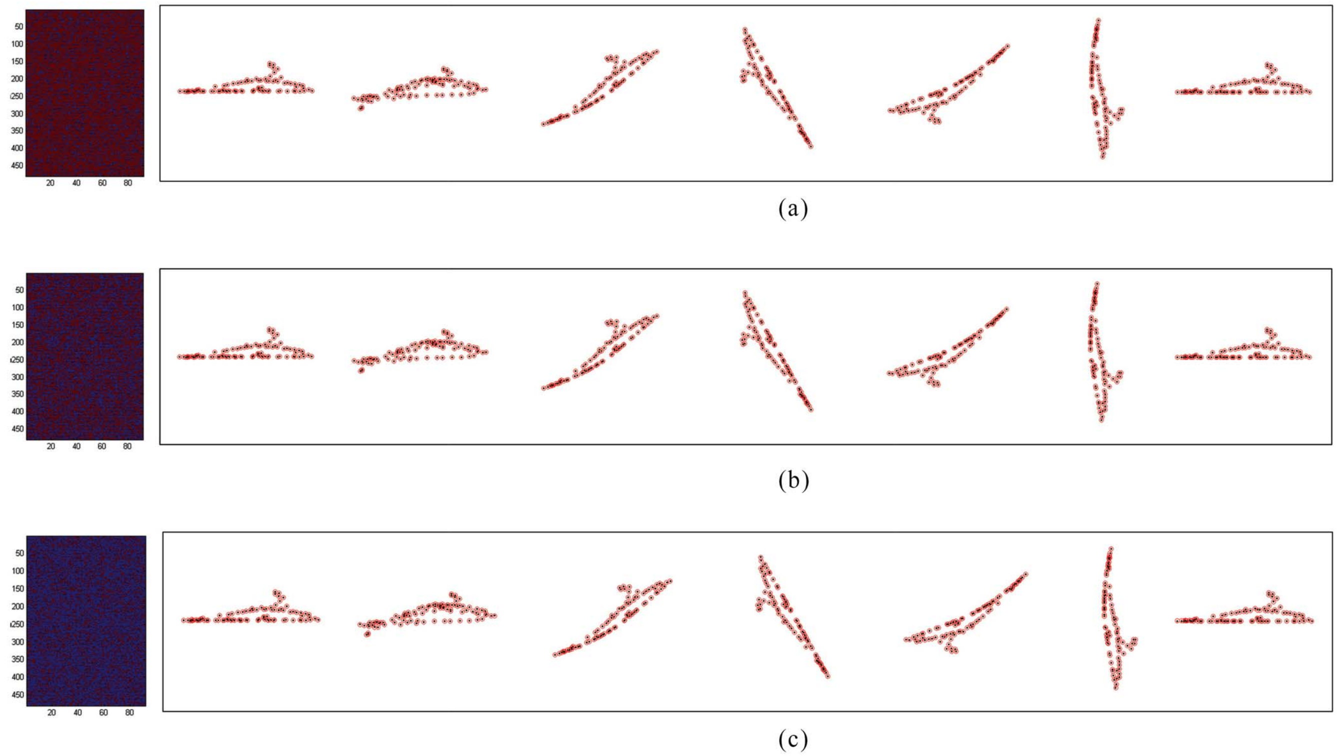

Fig. 6. 3-D shapes of *Shark1* at t_1, t_{40}, t_{80}, t_{120}, t_{160}, t_{200}, and t_{240} recovered by our method with (a) 25%, (b) 50%, and (c) 75% missing data.

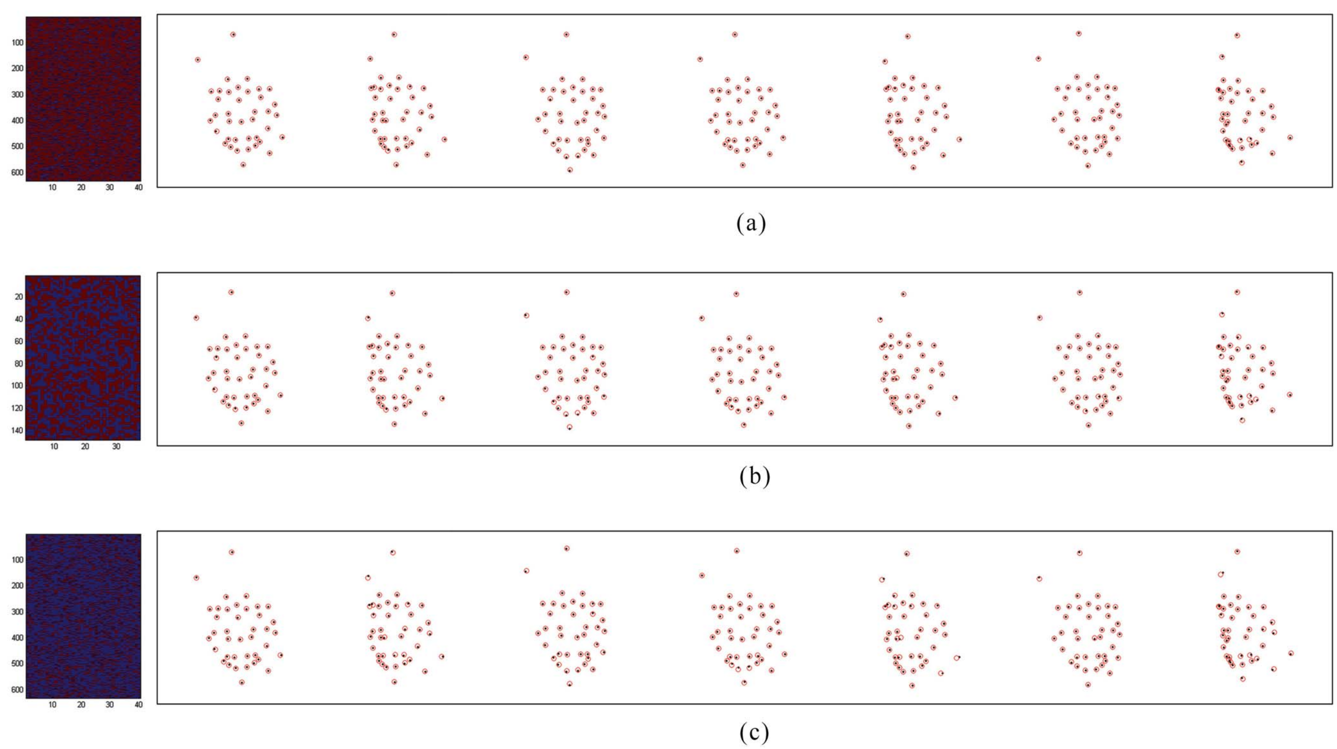

Fig. 7. 3-D shapes of *Face2* at t_1, t_{54}, t_{107}, t_{160}, t_{213}, t_{266}, and t_{316} recovered by our method with (a) 25%, (b) 50%, and (c) 75% missing data.

recovery error $\|\mathbf{W} - \widetilde{\mathbf{W}}\|_F / \|\mathbf{W}\|_F$ is $1.42e^{-3}$, $2.58e^{-3}$, and $7.29e^{-3}$ for 25%, 50%, and 75% cases, respectively. The recovered 3-D shapes are almost coincident with the ground-truth. All the results suggest that our method achieves excellent performance for all the cases. This appealing property attributes to the elegant design of low-rank matrix completion model, which consists of the data and regularization terms. The data term reflects the fidelity of the reconstructed matrix, and plays an important role in efficiently using the observation information. Simultaneously,

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT WAVELETS

Wavelets	Length of Filter	*Shark1*		*Shark2*		*Face1*		*Face2*		*Walking*	
		e_{3D}	RMSE	e_{3D}	RMSE	e_{3D}	RMSE	e_{3D}	RMSE	e_{3D}	RMSE
sym2	4	0.02157	0.25258	0.14035	0.01432	0.06248	0.52413	0.02947	0.46410	0.20653	7.51968
sym5	10	0.00118	0.01466	0.18811	0.02764	0.06248	0.52413	0.02947	0.46410	0.20654	7.51968
sym7	14	0.00024	0.00229	0.00475	0.00007	0.06248	0.52413	0.02947	0.46410	0.20643	7.51968
sym10	20	0.00001	0.00012	0.23175	0.02207	0.06113	0.52749	0.15820	0.49582	0.20653	7.51968
sym15	30	0.00000	0.00000	0.16756	0.00803	0.06114	0.52749	0.02947	0.46410	0.20657	7.51968
db2	4	0.02157	0.25258	0.14035	0.01432	0.06248	0.52413	0.02947	0.46410	0.20653	7.51968
db5	10	0.00203	0.01404	0.00160	0.00046	0.06249	0.52413	0.02947	0.46410	0.10267	6.04000
db7	14	0.00018	0.00196	0.00281	0.00007	0.06248	0.52413	0.02944	0.46410	0.10279	6.04000
db10	20	0.00001	0.00013	0.00018	0.00000	0.06248	0.52413	0.02144	0.00097	0.10271	6.04000
db15	30	0.00000	0.00000	0.17885	0.00680	0.06249	0.52413	0.02947	0.46410	0.10270	6.04000
DT	8	0.00000	0.00000	0.11958	0.00001	0.06227	0.53091	0.02946	0.46672	0.10294	6.04295
DCT		0.00808	0.03120	0.25387	0.07053	0.06368	0.69651	0.03627	0.69681	0.18630	19.07181

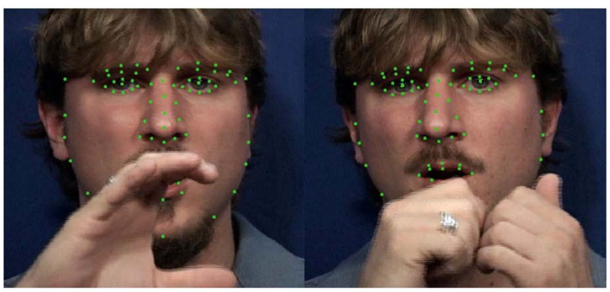

Fig. 8. Two frames of ASL dataset with annotated facial landmarks in green [14].

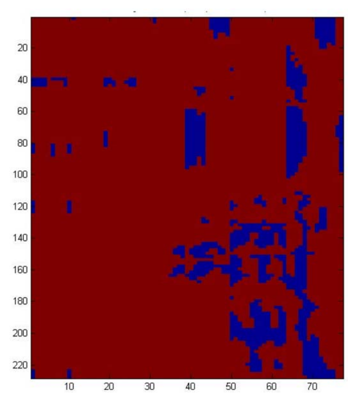

Fig. 9. Distribution of missing data (blue elements).

the regularization term guarantees the low-rankness of the reconstructed result. By bridging the data and regularization terms with a penalization parameter and then minimizing it, accurate tracking features are reserved while ensuring the low-rankness of matrix. Through this method, not only the matrix is recovered, but also the noises and outliers in the known elements are reduced.

Finally, an application of our method is in the interpretation of the facial expression component of sign languages from video [50]. We test our method on a 114-image (4 s long) face close-up video of an American sign language (ASL) sentence, in which facial landmarks are manually annotated in each image when visible [14], [50]. In this case, head rotation and hand gesticulation often cause the occlusion of facial features (see Fig. 8), which leads to incomplete 2-D point tracks. The observation matrix $\mathbf{W} \in \mathbb{R}^{228\times 77}$ misses 11.5% of its data and has small magnitude annotation errors (noise) due to annotation errors and motion blur in the images. The distribution of missing data is shown in Fig. 9. We recover the complete matrix with the proposed method, and the relative error on known elements is $2.317e^{-3}$. Fig. 10 gives one view of 3-D shapes at all time instants recovered by our method with $K = 4$. The RMSE is 1.1582. It can be seen that the pose and the deformation of mouth and eyes are correctly recovered despite the occlusions. This attributes to the proposed matrix recovery method and shape trajectory space fitting method. The former addresses the occlusion problem based on low-rank matrix completion while the latter accurately fits the space of complex motion trajectories due to its excellent capability in capturing isolated singularities.

C. Effects of Different Wavelets

The NR-SFM results of our method with different wavelets are shown in Table IV. We test three types of wavelet filters: Symlet wavelets (sym*N*), Daubechies orthogonal wavelets (db*N*), and dual-tree complex wavelets (DT) [51]. For sym*N* and db*N* wavelets, N is the order of the scaling function, and is half of the filter length. Five different filter lengths, i.e., 4, 10, 14, 20, and 30, are tested. For the DT wavelet, the filter length is 8. For comparison, we also present results generated by CSF1 [13] using DCT. As shown in Table IV, almost all the tested wavelets provide better results than DCT, which justifies the effectiveness of wavelets in fitting shape trajectories. With the same length of filters, Symlet wavelets have similar performance to the Daubechies wavelets. The DT wavelet

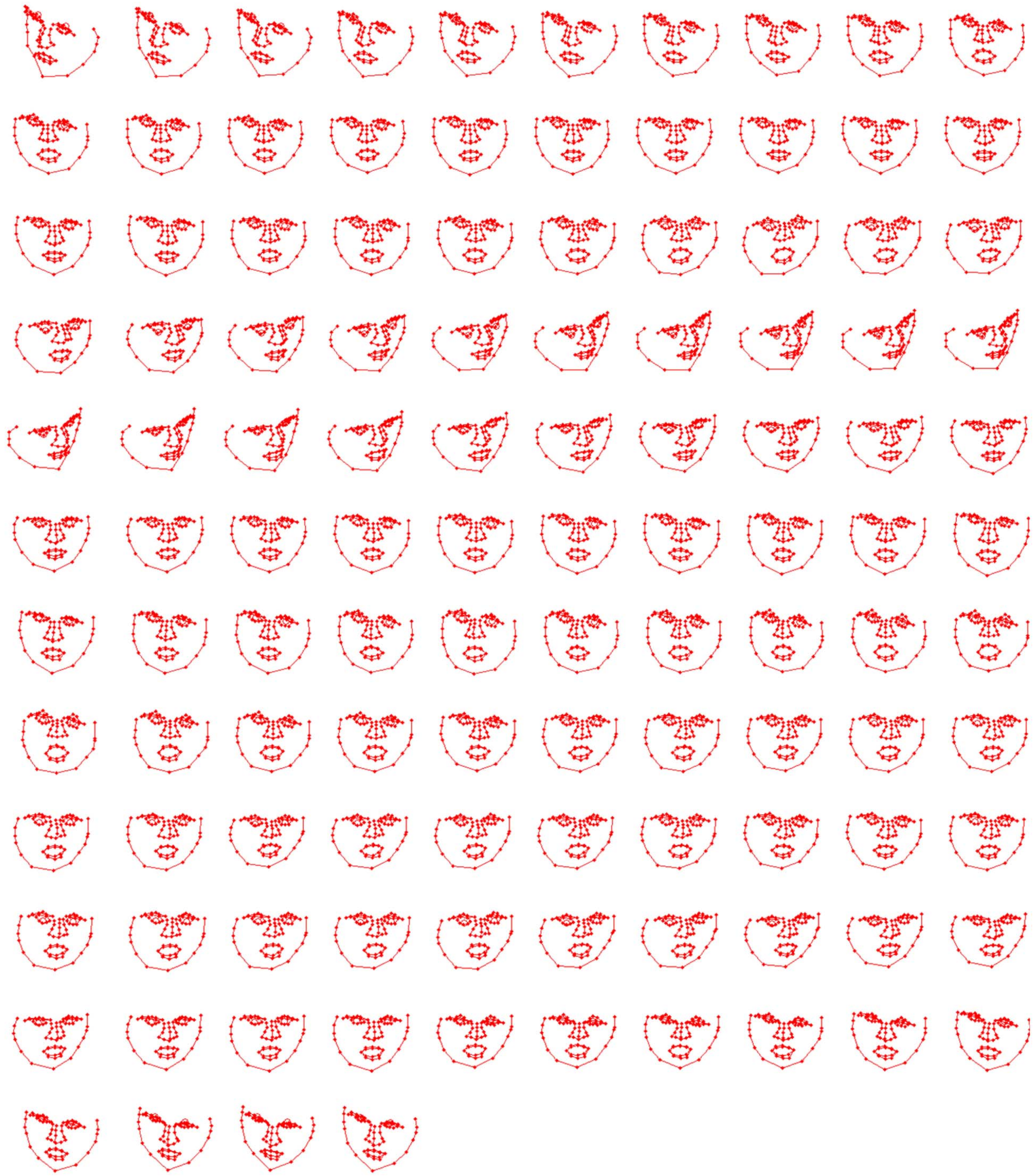

Fig. 10. One view of recovered 3-D shapes of ASL dataset at all time instants.

is also comparable to Symlet and Daubechies wavelets of the same length. However, the behaviors of wavelets with different filter lengths fall into two categories for the test datasets. For *Shark1* and *Shark2*, the performance significantly varies with the filter length, and the middle or long filters (of length 14 and beyond) are preferred. It is quite remarkable that sym15, db15, and DT can even achieve perfect recovery for *Shark1*. For *Face1*, *Face2*, and *Walking*, all the wavelets have quite stable performance over different wavelet types and filter lengths. The reason may be that, for very complicate motions such as in *Face1*, *Face2*, and *Walking*, wavelets of different filter lengths have the same representation efficiency, and hence have very close performance.

VI. Conclusion

This paper presents a novel NR-SFM method based on sparse representation. We address the problem with occlusion based on matrix completion, which recovers the observation matrix that has high percentages of missing data and also reduces the noises and outliers in the known elements. Sparse transform is employed to jointly estimate 3-D shapes and motions. Experimental results show that our method outperforms the state-of-the-art NR-SFM algorithms. In future work, we will improve the performance of our method by learning dictionary for sparser representation of complex motions than wavelet transforms, and deal with larger scale datasets.

Acknowledgment

The authors would like to thank Dr. P. F.U. Gotardo and Dr. Y. Dai for helpful discussions and the distribution of datasets and codes.

References

[1] G. Ye *et al.*, "Free-viewpoint video of human actors using multiple handheld Kinects," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1370–1382, Oct. 2013.

[2] K. Li, Q. Dai, W. Xu, J. Yang, and J. Jiang, "Three-dimensional motion estimation via matrix completion," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 539–551, Apr. 2012.

[3] A. Y. Mulayim, U. Yilmaz, and V. Atalay, "Silhouette-based 3-D model reconstruction from multiple images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 33, no. 4, pp. 582–591, Aug. 2003.

[4] A. Barmpoutis, "Tensor body: Real-time reconstruction of the human body and avatar synthesis from RGB-D," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1347–1356, Oct. 2013.

[5] M. Camplani, T. Mantecon, and L. Salgado, "Depth-color fusion strategy for 3-D scene modeling with Kinect," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1560–1571, Dec. 2013.

[6] M. Liao, Q. Zhang, H. Wang, R. Yang, and M. Gong, "Modeling deformable objects from a single depth camera," in *Proc. 12th IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep./Oct. 2009, pp. 167–174.

[7] Y. Gao *et al.*, "Less is more: Efficient 3-D object retrieval with query view selection," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 1007–1018, Oct. 2011.

[8] Y. Gao *et al.*, "Camera constraint-free view-based 3-D object retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2269–2281, Apr. 2012.

[9] Y. Gao, M. Wang, D. Tao, R. Ji, and Q. Dai, "3-D object retrieval and recognition with hypergraph analysis," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4290–4303, Sep. 2012.

[10] H. V. Le, "A structure-from-motion method for 3-D reconstruction of moving objects from multiple-view image sequences," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 1955–1958.

[11] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3D shape from image streams," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Hilton Head Island, SC, USA, 2000, pp. 690–696.

[12] I. Akhter, Y. A. Sheikh, S. Khan, and T. Kanade, "Nonrigid structure from motion in trajectory space," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 41–48.

[13] P. F. U. Gotardo and A. M. Martinez, "Computing smooth time-trajectories for camera and deformable shape in structure from motion with occlusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 2051–2065, Oct. 2011.

[14] P. F. U. Gotardo and A. M. Martinez, "Non-rigid structure from motion with complementary rank-3 spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2011, pp. 3065–3072.

[15] D. Jacobs, "Linear fitting with missing data for structure-from-motion," *Comput. Vis. Image Underst.*, vol. 82, no. 1, pp. 57–81, 2001.

[16] H. Jia and A. Martinez, "Low-rank matrix fitting based on subspace perturbation analysis with applications to structure from motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 841–854, May 2009.

[17] N. Guilbert, A. Bartoli, and A. Heyden, "Affine approximation for direct batch recovery of Euclidian structure and motion from sparse data," *Int. J. Comput. Vis.*, vol. 69, no. 3, pp. 317–333, 2006.

[18] S. Olsen and A. Bartoli, "Implicit non-rigid structure-from-motion with priors," *J. Math. Imaging Vis.*, vol. 31, no. 2, pp. 233–244, 2008.

[19] J. Tardif, A. Bartoli, M. Trudeau, N. Guilbert, and S. Roy, "Algorithms for batch matrix factorization with application to structure-from-motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[20] R. Hartley and F. Schaffalitzky, "PowerFactorization: 3D reconstruction with missing or uncertain data," in *Proc. Aust. Jpn. Adv. Workshop Comput. Vis.*, vol. 74, 2003, pp. 76–85.

[21] M. Paladini *et al.*, "Factorization for non-rigid and articulated structure using metric projections," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Miami, FL, USA, Jun. 2009, pp. 2898–2905.

[22] L. Torresani, A. Hertzmann, and C. Bregler, "Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 5, pp. 878–892, May 2008.

[23] A. Buchanan and A. Fitzgibbon, "Damped Newton algorithms for matrix factorization with missing data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Jun. 2005, pp. 316–322.

[24] P. Chen, "Optimization algorithms on subspaces: Revisiting missing data problem in low-rank matrix," *Int. J. Comput. Vis.*, vol. 80, no. 1, pp. 125–142, 2008.

[25] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[26] E. J. Candès and Y. Plan, "Matrix completion with noise," *Proc. IEEE*, vol. 98, no. 6, pp. 925–936, Jun. 2010.

[27] J. Fortuna and A. M. Martinez, "Rigid structure from motion from a blind source separation perspective," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 404–424, 2012.

[28] P. Chen and D. Suter, "Recovering the missing components in a large noisy low-rank matrix: Application to SFM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1051–1063, Aug. 2004.

[29] A. Buchanan and A. Fitzgibbon, "Damped Newton algorithms for matrix factorization with missing data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Washington, DC, USA, 2005, pp. 316–322.

[30] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *Int. J. Comput. Vis.*, vol. 80, no. 2, pp. 189–210, Nov. 2008.

[31] S. Agarwal, N. Snavely, I. Simon, S. Seitz, and R. Szeliski, "Building Rome in a day," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 72–79.

[32] C. Wu, S. Agarwal, B. Curless, and S. Seitz, "Multicore bundle adjustment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2011, pp. 3057–3064.

[33] J. Xiao, J. Chai, and T. Kanade, "A closed-form solution to non-rigid shape and motion recovery," *Int. J. Comput. Vis.*, vol. 67, no. 2, pp. 233–246, 2006.

[34] L. Torresani, D. B. Yang, E. J. Alexander, and C. Bregler, "Tracking and modeling non-rigid objects with rank constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2001, vol. 1, pp. 493–500.

[35] L. Torresani, A. Hertzmann, and C. Bregler, "Learning non-rigid 3D shape from 2D motion," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 1–8.

[36] A. Bartoli *et al.*, "Coarse-to-fine low-rank structure-from-motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[37] V. Rabaud and S. Belongie, "Re-thinking non-rigid structure from motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Anchorage, AK, USA, 2008, pp. 1–8.

[38] A. Del Bue, F. Smeraldi, and L. Agapito, "Non-rigid structure from motion using ranklet-based tracking and non-linear optimization," *Image Vis. Comput.*, vol. 25, no. 3, pp. 297–310, 2007.

[39] H. Park, T. Shiratori, I. Matthews, and Y. Sheikh, "3D reconstruction of a moving point from a series of 2D projections," in *Proc. Eur. Conf. Comput. Vis.*, Heraklion, Greece, 2010, pp. 158–171.

[40] Y. Zhu, M. Cox, and S. Lucey, "3D motion reconstruction for real-world camera motion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, 2011, pp. 1–8.

[41] J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, USA, 1994, pp. 593–600.

[42] M. Fazel, H. Hindi, and S. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proc. Amer. Control Conf.*, Denver, CO, USA, 2003, pp. 2156–2162.

[43] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Math. Program.*, vol. 128, no. 1, pp. 321–353, 2011.

[44] J. Gao, H. Sultan, J. Hu, and W. Tung, "Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison," *IEEE Signal Process. Lett.*, vol. 17, no. 3, pp. 237–240, Mar. 2010.

[45] J. Yang, Y. Wang, W. Xu, and Q. Dai, "Image coding using dual-tree discrete wavelet transform," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1555–1569, Sep. 2008.

[46] Y. Xu, X. Yang, H. Ling, and H. Ji, "A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, Jun. 2010, pp. 161–168.
[47] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*. New York, NY, USA: Wiley, 1988.
[48] O. C. Hamsici, P. F. Gotardo, and A. M. Martinez, "Learning spatially-smooth mappings in non-rigid structure from motion," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 260–273.
[49] Y. Dai, H. Li, and M. He, "A simple prior-free method for non-rigid structure-from-motion factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2018–2025.
[50] L. Ding and A. M. Martinez, "Modelling and recognition of the linguistic components in American Sign Language," *Image Vis. Comput.*, vol. 27, no. 12, pp. 1826–1844, 2009.
[51] N. G. Kingsbury, "The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters," in *Proc. IEEE Digit. Signal Process. Workshop*, vol. 86. 1998, pp. 120–131.

Kun Li received the B.E. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2006, the master's and the Ph.D. degrees from Tsinghua University, Beijing, in 2011.

She is currently an Assistant Professor with the School of Computer Science and Technology, Tianjin University, Tianjin, China. She visited École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 2012. Her current research interests include image/video processing, image-based modeling, dynamic scene 3-D reconstruction, and multicamera imaging.

Jingyu Yang received the B.E. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2003, and the Ph.D. (Hons.) degree from Tsinghua University, Beijing, in 2009.

Since 2009, he has been with the Faculty of Tianjin University, Tianjin, China, and is currently an Associate Professor with the School of Electronic Information Engineering, Tianjin University. He visited Microsoft Research Asia (MSRA), Beijing, within the MSRAs young scholar supporting program in 2011. He also visited the Signal Processing Laboratory with École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 2012. He was selected into the program for New Century Excellent Talents in University from the Ministry of Education of China, in 2011, and the Elite Scholar Program of Tianjin University, in 2012. His current research interests include image/video processing, 3-D imaging, and computer vision.

Jianmin Jiang received the Ph.D. degree from the University of Nottingham, Nottingham, U.K., in 1994.

He joined Loughborough University, Loughborough, U.K, as a Lecturer in computer science. From 1997 to 2001, he was a Full Professor of Computing with the University of Glamorgan, Wales, U.K. In 2002, he joined the University of Bradford, Bradford, U.K, as a Chair Professor of Digital Media, and Director of Digital Media and Systems Research Institute. In 2014, he moved to Shenzhen University, Shenzhen, China, to carry on holding the same professorship. He is also an Adjunct Professor with the University of Surrey, Guildford, U.K. His current research interests include image/video processing in compressed domain, computerized video content understanding, stereo image coding, medical imaging, computer graphics, machine learning, and AI applications in digital media processing, retrieval, and analysis. He has published over 400 refereed research papers.

Prof. Jiang is a Chartered Engineer, a member of EPSRC College, and EU FP$-6/7$ evaluation expert. In 2010, he was elected as a scholar of One-Thousand-Talent-Scheme funded by the Chinese Government and joined Tianjin University, Tianjin, China, to hold the One-Thousand-Talent-Scheme professorship.