Trajectory Space: A Dual Representation for Nonrigid Structure from Motion

Ijaz Akhter, *Student Member*, *IEEE*, Yaser Sheikh, *Member*, *IEEE*, Sohaib Khan, *Member*, *IEEE*, and Takeo Kanade, *Fellow*, *IEEE*

Abstract—Existing approaches to nonrigid structure from motion assume that the instantaneous 3D shape of a deforming object is a linear combination of basis shapes. These bases are object dependent and therefore have to be estimated anew for each video sequence. In contrast, we propose a dual approach to describe the evolving 3D structure in trajectory space by a linear combination of basis *trajectories*. We describe the dual relationship between the two approaches, showing that they both have equal power for representing 3D structure. We further show that the temporal smoothness in 3D trajectories alone can be used for recovering nonrigid structure from a moving camera. The principal advantage of expressing deforming 3D structure in trajectory space is that we can define an *object independent* basis. This results in a significant reduction in unknowns and corresponding stability in estimation. We propose the use of the Discrete Cosine Transform (DCT) as the object independent basis and empirically demonstrate that it approaches Principal Component Analysis (PCA) for natural motions. We report the performance of the proposed method, quantitatively using motion capture data, and qualitatively on several video sequences exhibiting nonrigid motions, including piecewise rigid motion, partially nonrigid motion (such as a facial expressions), and highly nonrigid motion (such as a person walking or dancing).

Index Terms—Nonrigid structure from motion, 3D reconstruction, motion and tracking.

1 Introduction

To what extent is it possible to infer the 3D structure of a deforming object from the motion of its salient features in a video sequence? Johansson, in his famous Moving Light Display experiment [1], demonstrated that if humans recognize an object, they can perceive structure deformations correctly. Johansson's study, for the first time, showed that recovering the 3D structure of a deforming object is possible, provided that the deforming object is recognized. In this paper, we show that the temporal smoothness of points in 3D, in the absence of recognition, is also sufficient to reconstruct the structure of a deforming object from a moving camera.

Temporal smoothness can be exploited to express trajectories as a linear combination of basis trajectories. This representation is illustrated in Fig. 1; each trajectory corresponds to a salient feature on the mouth of a smiling actor. We represent each trajectory by a point in the linear space of trajectories spanned by a trajectory basis. We show that this representation is a dual to the shape basis representation of Bregler et al. [2]. The key idea in [2] is that observed shapes can be represented as a linear combination of a few basis shapes, as illustrated in Fig. 1c. The duality between shape and trajectory basis arises because the trajectory and the shape basis span the column and row space of the same matrix representing nonrigid structure. Compactness in one automatically imposes compactness in the other. We will show that the role of the bases and their coefficients is swapped between these two representations; the shape basis and shape coefficients become the trajectory coefficients and trajectory basis, respectively, in the dual space and vice versa.

Although, both shape and trajectory are alternate ways of looking at the nonrigid structure, there is a key advantage to taking the trajectory approach—the trajectory basis can be predefined in an object independent way. Consider a deformable object being acted upon by a force. The extent of its deformation is limited by the force that can be applied. Hence, a tree swaying in the wind or a person walking cannot arbitrarily and randomly deform; the trajectories of their points are a function of the force of the wind and the flexing of muscles, respectively. Deformations are therefore constrained by the physical limits of the actuation to remain incremental, not random, across time. As this property is largely ubiquitous, a basis can be defined in a trajectory space that is *object independent*.

The incremental nature of the trajectories suggests that they should be well modeled by low-order Markov chains. Moreover, the temporal smoothness in trajectories implies a high auto-correlation in a small time window. It is widely known that for highly correlated data, the Discrete Cosine Transform (DCT) has excellent energy compaction [3]. It has further been shown that for a first-order Markov model, the basis computed by Principal Component Analysis (PCA) approaches DCT when correlation approaches unity or the signal length approaches infinity [4]. Empirically, our experiments on the large CMU motion capture data set show that for motion trajectories, the basis computed by PCA approaches the DCT basis. We conclude from

- *I. Akhter and S. Khan are with the Department of Computer Science, LUMS School of Science and Engineering, Lahore University of Management Sciences, D.H.A, Lahore Cantt, Pakistan, 54792. Email: {akhter, sohaib}@lums.edu.pk.*
- *Y. Sheikh and T. Kanade are with the Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. Email: {yaser, tk}@cs.cmu.edu.*

Manuscript received 26 May 2009; revised 5 Mar. 2010; accepted 15 Sept. 2010; published online 9 Nov. 2010.
Recommended for acceptance by F. Kahl.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2009-05-0333.
Digital Object Identifier no. 10.1109/TPAMI.2010.201.

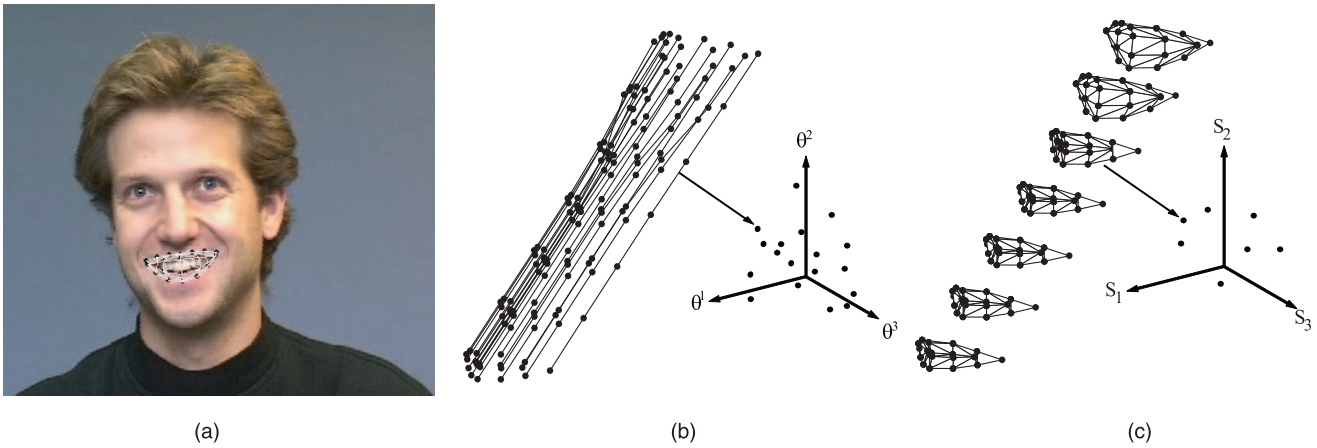

Fig. 1. 3D features on a smiling mouth: A comparison of shape and trajectory space. (a) In approaches that represent the time varying structure in shape space, all 3D points observed at one time instant are projected onto a single point in the shape space. $S_1, S_2, \ldots, S_K$ each represent a shape basis vector. (b) In our approach, we represent the time varying structure in trajectory space, where a 3D point's trajectory over time is projected to a single point in the trajectory space. $\theta^1, \theta^2, \ldots, \theta^K$ each represent a trajectory basis vector. P points observed across F frames are expressed as F projected points in shape space and $3P$ points in trajectory space.

quantitative and qualitative experiments that the DCT basis serves well as an object independent trajectory basis.

Using DCT as an object independent basis for nonrigid structure representation, we are able to propose a new algorithm for recovering structure from motion. An important consequence of structure representation in the space of a priori basis is a significant reduction in the number of unknowns. This allows us to handle more nonrigidity of deformation than state-of-art methods such as [5] and [6]. In fact, most previous results consider deformations which have a large rigid component, such as talking-head videos or the motion of a swimming shark [2], [5], [6], [7], [8], [9], [10], [11]. To the best of our knowledge, we are the first to show reasonable reconstructions of highly nonrigid motions, e.g., a person dancing or a group of people moving, without making object specific assumptions. We observe and empirically demonstrate that the stability of structure estimation is linked to the amount of camera motion. The greater the motion of the camera, the more nonrigid the structure which can be reconstructed.

2 Related Work

The study of the relationship between 2D image projections and underlying 3D scenes has a long history over several centuries in the fields of optics and photogrammetry. The study of multiview geometry was spurred in the computer vision community by Longuet-Higgins in [12]. Almost three decades of section-sequent research is summarized in [13], [14], [15], investigating various constraints between point measurements in different images. While the multilinear relationships described in this body of work greatly expanded conceptual understanding, in practice the most successful line of investigation for 3D reconstruction has been the application of factorization approaches to structure recovery beginning with the seminal work by Tomasi and Kanade in [16]. The same formulation was independently proposed by Kontsevich et al. in [17]. The key observation in [16], [17] was the rank 3 theorem. It stated that a $2F \times P$ measurement matrix $\mathbf{W}$, containing the image coordinates of P points across F frames,

$$\mathbf{W} = \begin{bmatrix} u_{11} & \ldots & u_{1P} \\ v_{11} & \ldots & v_{1P} \\ \vdots & & \vdots \\ u_{F1} & \ldots & u_{FP} \\ v_{F1} & \ldots & v_{FP} \end{bmatrix}, \quad (1)$$

has a rank of 3 if orthographic projection is assumed and image coordinates are taken with respect to a common convention of origin. A numerically stable algorithm was proposed to exploit this rank constraint for structure reconstruction, and the orthonormality of rotations was used to estimate metric structure. The vast majority of research into structure from motion, including factorization approaches, assumes that scene is stationary and equivalently considers either multiple static cameras or a single moving camera. These methods cannot be used to directly recover the structure of dynamic scenes or the scene independent relative motion of two (or more) cameras.

The principal challenge in reconstructing dynamic structure from a moving camera is that the problem is ill-posed if the dynamics of the structure is unconstrained, i.e., if the structure at time $t+1$ is independent of the structure at t. The groundbreaking 1973 study by Johansson in [1] demonstrated that in interpreting structure from motion, the human visual system has the ability to reconstruct time-varying structures. Within the computer vision community, the design of algorithms to reconstruct nonrigidly deforming structure began in earnest in the 1980s, investigating constraints like maximal rigidity, isometry, symmetry, and linear representations in low dimensions in [18], [19], [20], [21]. During the 1990s, the success of the factorization-based structure from motion algorithm initiated investigation into leveraging similar ideas toward nonrigid structure from motion, with independently moving rigid objects [22], [23], [24], [25], [26]. A more general model of nonrigid structure from motion was proposed by Bregler et al. in [2]. They modeled the deformations of a nonrigid object using a low-dimensional set of linear basis, which they called the shape basis, reminiscent of the model of [21]. The structure at a time instant t was represented by arranging the 3D locations of the P points in a matrix $S(t) \in \mathbb{R}^{3\times P}$:

$$S(t) = \begin{bmatrix} X_{t1} & \cdots & X_{tP} \\ Y_{t1} & & Y_{tP} \\ Z_{t1} & & Z_{tP} \end{bmatrix}.$$

The complete time varying structure was represented by concatenating these instantaneous structures as

$$\mathbf{S}_{3F\times P} = [S(1)^T S(2)^T \cdots S(F)^T]^T. \quad (2)$$

In [2], each instantaneous shape matrix $S(t)$ was approximated by a weighted sum of K basis shapes,

$$S(t) = \sum_{j=1}^{K} \omega_{tj} B^j, \quad (3)$$

where $B^j \in \mathbb{R}^{3\times P}$ is a basis shape and ω_{tj} is the coefficient of that basis shape. Assuming that K shape basis can capture the variation in the object's deformation, Bregler et al. described a rank $3K$ theorem, analogous to the rank 3 theorem in the case of stationary objects.

While this framework was seminal, the algorithm to estimate the metric structure of the nonrigid object did not have the stability of the rigid factorization approach of Tomasi and Kanade [16]. The difference between the two cases was that in addition to camera motion and structure information, in the approach of Bregler et al. the shape bases had to be estimated anew for each object since they were specific to the observed data. Brand in [8] proposed several optimization strategies to estimate nonrigid structure. Xiao et al. in [5] attributed the fragility of the algorithm to the ambiguity in orthonormality constraints. They proposed additional constraints, called basis constraints, to resolve this ambiguity to get a unique solution. Torresani et al. [6] proposed a Gaussian prior for the shape coefficients and solve the optimization using Expectation Maximization. Bue et al. [27] assumed that nonrigid shape contained a significant number of points which behave in a rigid fashion. Bue in [28] introduced the idea of priors in nonrigid structure from motion. Bartoli et al. [29] used a smoothness prior on structure and an ordering prior on the basis, where the smoothness prior is the closeness of deformations from mean. Yan and Pollefeys [30], [31] and Tresadern and Reid [32] assumed that a nonrigid object is articulated. Paladini et al. [33] proposed an alternating least square approach and manifold projection technique for nonrigid and articulated objects. Rabaud and Belongie in [34] showed that repetitions in the object deformations can be exploited to estimate shape coefficients, which can further provide a good estimate for the remaining unknowns in the final optimization. Akhter et al. in [35] demonstrated that the fundamental problem in nonrigid structure from motion is the optimization not the ambiguity of orthonormality constraints. Investigation has been conducted to extend nonrigid factorization approaches to perspective camera models in [36], [37], [38]; however, robust solutions for handling significant nonrigidity $(K > 2)$ remain elusive. Recently, Rabaud and Belongie [34] proposed nonlinear subspace reduction by imposing locally linear subspace compaction on nonrigid structure [39].

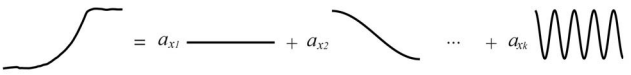

Fig. 2. As described in (4), each trajectory is represented as a linear combination of K predefined basis trajectories. In this paper, we use the Discrete Cosine Transform basis to compactly represent trajectories.

In contrast to this entire corpus of work, which approximates structure by a shape basis, we propose a new representation of time varying structure as a collection of trajectories. An initial framework of the proposed method was published in [40]. In [40], we not only demonstrate that a compact trajectory space can be defined, but also that the basis of this trajectory space can be predefined, removing a large number of unknowns from the estimation process altogether. The duality of spatial and temporal representations has been hinted at earlier in the literature. Carlsson and Weinshall [41] discussed the duality between 3D points and projective cameras. Shashua [42] discussed the duality of the *joint image space* and the *joint point space* in the context of multiview geometry. Zelnik-Manor and Irani [43] have exploited a similar duality for an alternate approach to segmenting video sequences. Ours is the first paper to use this dual representation in the structure from motion problem and to note that a generic basis can be defined in trajectory space, which compactly represents most real trajectories.

We exploit temporal smoothness of trajectories to predefine the basis. In contrast to earlier constraints imposed on nonrigid structure from motion, we note that temporal smoothness is physically motivated and is a limitation imposed by the actuators causing the nonrigid motion in deforming object. Torresani et al. proposed a linear dynamical system to exploit temporal smoothness. Olsen and Bartoli [44] used temporal smoothness to handle missing data. However, ours is the first paper to treat temporal smoothness as a sufficient condition to solve nonrigid structure from motion. Another distinction of our paper is that it gives reasonable reconstructions not only for cases for which specialized methods were proposed, like multirigid and articulated objects, but also for the cases which have never been reported before, like a person dancing or multiple people walking.

3 Duality in 3D Structure Representation

The shape basis representation as given in (3) is one way of imposing compactness on the nonrigid structure $\mathbf{S}$. Another way is to look across time and impose compactness on trajectories. We define the 3D trajectory of the ith point as $T(i) = [T_x(i)^T T_y(i)^T T_z(i)^T]^T$, where $T_x(i) = [X_{1i}, \ldots, X_{Fi}]^T$, $T_y(i) = [Y_{1i}, \ldots, Y_{Fi}]^T$, $T_z(i) = [Z_{1i}, \ldots, Z_{Fi}]^T$ are the X, Y, and Z coordinates of the ith trajectory. We assume that the trajectory components can be approximated by a linear combination of a small number of trajectory basis (see Figs. 1b and 2) as

$$T_x(i) = \sum_{j=1}^{K} a_{xj}(i)\theta^j,$$
$$T_y(i) = \sum_{j=1}^{K} a_{yj}(i)\theta^j, \quad (4)$$
$$T_z(i) = \sum_{j=1}^{K} a_{zj}(i)\theta^j,$$

where $\theta^j \in \mathbb{R}^F$ is a basis trajectory. Although (3) and (4) show two apparently different approaches to approximate the nonrigid structure, they are both related. In the following, we explore the relationship between shape basis and trajectory basis representation of nonrigid structure.

3.1 Duality Theorem

Let us assume that the nonrigid structure $\mathbf{S}$ can be compactly expressed in terms of K basis shapes, as given in (3). This relationship can also be written as a matrix multiplication by vectorizing each of the $3 \times P$ instantaneous shapes $S(t)$ as a $1 \times 3P$ row vector. The nonrigid structure $\mathbf{S}$ can be reformulated as a $F \times 3P$ matrix $\mathbf{S}^*$ as follows:

$$\mathbf{S}^* = \begin{bmatrix} X_{11} & Y_{11} & Z_{11} & & X_{1P} & Y_{1P} & Z_{1P} \\ & \vdots & & \cdots & & \vdots & \\ X_{F1} & Y_{F1} & Z_{F1} & & X_{FP} & Y_{FP} & Z_{FP} \end{bmatrix}. \quad (5)$$

The row space of $\mathbf{S}^*$ corresponds to the shape space, which, according to [2], can be compactly represented by K basis vectors. As the row and column space of a matrix are of equal dimensions, it follows that the columns of $\mathbf{S}^*$ are also spanned by K vectors. We call the column space of this matrix the *trajectory space* and note that it enjoys a dual relationship with the shape space. Specifically, if the time varying shape of an object can be expressed by a minimum of K shape basis, then there exist exactly K trajectory basis vectors that can represent the same time varying shape. This leads us to the following theorem between the shape and the trajectory basis.

Theorem 3.1. *The shape basis and trajectory basis are dual to each other.*

Proof. In order to prove the theorem, let us assume that the nonrigid structure lies in a shape space of dimension K, then we can express $\mathbf{S}^*$ as a linear combination of shape basis as follows:

$$\mathbf{S}^* = \mathbf{\Omega}^* \mathcal{B}^*, \quad (6)$$

where

$$\mathbf{\Omega}^* = \begin{bmatrix} \omega_{11} & \cdots & \omega_{1K} \\ \vdots & \cdots & \vdots \\ \omega_{F1} & \cdots & \omega_{FK} \end{bmatrix}$$

and

$$\mathcal{B}^* = \begin{bmatrix} b_{x1}(1) & b_{y1}(1) & b_{z1}(1) & & b_{x1}(P) & b_{y1}(P) & b_{z1}(P) \\ & \vdots & & \cdots & & \vdots & \\ b_{xK}(1) & b_{yK}(1) & b_{zK}(1) & & b_{xK}(P) & b_{yK}(P) & b_{zK}(P) \end{bmatrix}.$$

Each row in $\mathcal{B}^*_{K\times 3P}$ represents a shape basis and c_{ij} is the ith shape coefficient of jth shape basis. In order to relate this representation to the X, Y, and Z coordinates of the ith trajectory, we have to multiply $\mathbf{\Omega}^*$ with the ith, $P+i$th, and $2P+i$th column of $\mathcal{B}^*$ as follows:

$$\begin{bmatrix} T_x(i) T_y(i) T_z(i) \end{bmatrix} = \begin{bmatrix} \omega_{11} & \cdots & \omega_{1K} \\ \vdots & \cdots & \vdots \\ \omega_{F1} & \cdots & \omega_{FK} \end{bmatrix} \begin{bmatrix} b_{x1}(i) & b_{y1}(i) & b_{z1}(i) \\ \vdots & \vdots & \vdots \\ b_{xK}(i) & b_{yK}(i) & b_{zK}(i) \end{bmatrix}. \quad (7)$$

By denoting the jth column in $\mathbf{\Omega}^*$ as ω^j, the above equation can be written as

$$T_x(i) = \sum_{j=1}^{K} b_{xj}(i)\omega^j,$$
$$T_y(i) = \sum_{j=1}^{K} b_{yj}(i)\omega^j, \quad (8)$$
$$T_z(i) = \sum_{j=1}^{K} b_{zj}(i)\omega^j.$$

By comparing (4) and (8), we conclude that the shape coefficients ω^j are serving as the trajectory basis and the components of the shape basis $b_{xj}(i)$, $b_{yj}(i)$, and $b_{zj}(i)$ have become the corresponding trajectory coefficients. The converse can also be shown in a similar way, that is, (3) can be derived from (8). Hence, the shape and trajectory representations are dual to each other and have equal compactness. □

A consequence of the Duality Theorem is the dual relationship between coefficients and basis in trajectory and shape spaces. Specifically, the shape basis and trajectory basis define two dual spaces and basis in one become coefficients in the other and vice versa.

3.2 An Illustration of the Duality for SVD Basis

For the illustration of duality between shape and trajectory representations, we give the example of SVD basis. This basis has been extensively used in structure from motion literature because of its effectiveness for structure recovery. In Fig. 3, we illustrate the duality between trajectory and shape representation using an example. We compute the SVD trajectory basis and shape basis for a time-varying 3D structure, generated through motion capture and normalize the trajectory coefficients and shape coefficients to unit norm. The coefficients of one space exactly overlap the basis of the other, illustrating the validity of duality for this data set.

The dual relationship between SVD trajectory and shape basis, shown to exist empirically in Fig. 3, can also be proven formally. In order to reaffirm the Duality Theorem, we will derive this relationship autonomously. For this purpose, we consider the covariance matrices for the trajectory and shape data as $\mathbf{\Sigma}_t = \mathbf{S}^* \mathbf{S}^{*T}$ and $\mathbf{\Sigma}_s = \mathbf{S}^{*T} \mathbf{S}^*$, respectively. We denote $\mathbf{\Phi}_t$ and $\mathbf{\Lambda}_t$ as the matrices of eigenvectors and eigenvalues of the covariance matrix and $\mathcal{A}^*$ as the coefficients of the trajectory data $\mathbf{S}^*$. Similarly, we denote $\mathbf{\Phi}_s$, $\mathbf{\Lambda}_s$, and $\mathbf{\Omega}^*$ for the shape data $\mathbf{S}^{*T}$. Matrices $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_s$ are assumed to be diagonal. The eigenvalue equation for the shape and the trajectory covariance matrices can be written as

$$\mathbf{\Sigma}_t \mathbf{\Phi}_t = \mathbf{\Phi}_t \mathbf{\Lambda}_t, \quad (9)$$

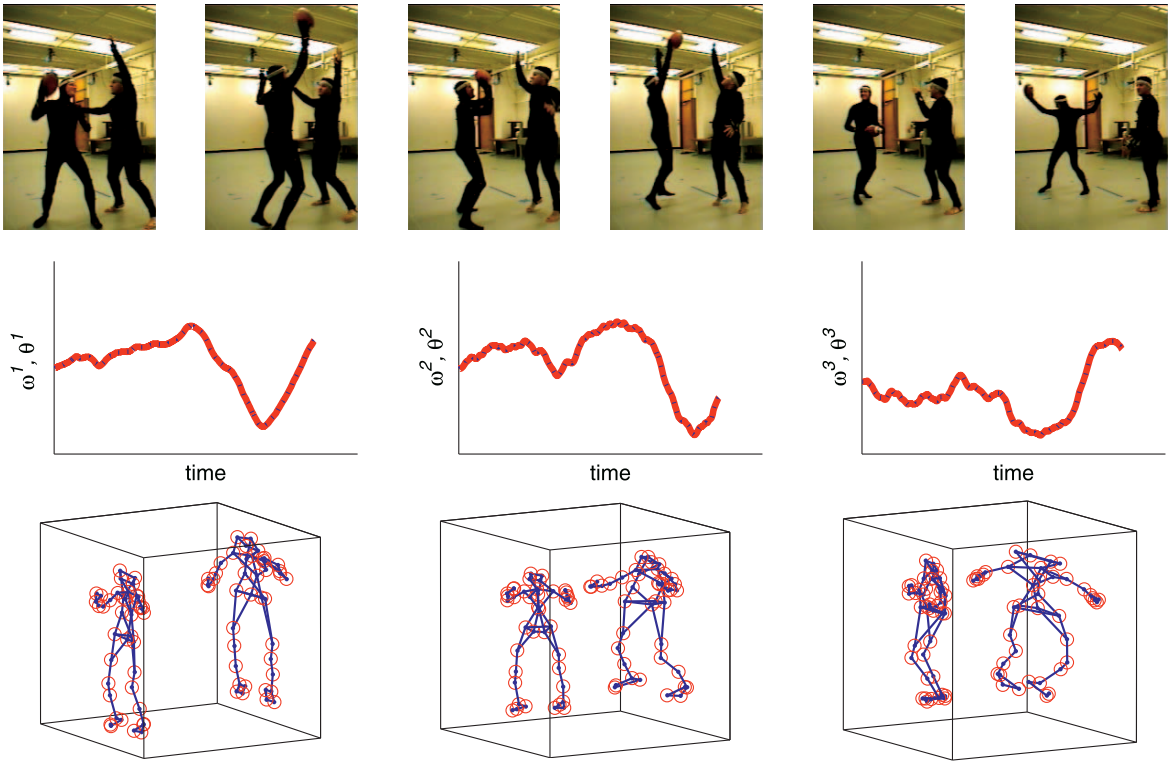

Fig. 3. Duality between SVD shape and trajectory representation: The shape bases are equal to the trajectory coefficients as the trajectory bases are equal to shape coefficients up to a scale. The top row shows sample frames in the data set. The second row shows the three trajectory bases (red solid) and shape coefficients (blue dots) superimposed on one another. The third row shows the shape basis (blue dots) and trajectory coefficients (red circle). The trajectory and shape coefficients were appropriately scaled for better visualization.

$$\mathbf{\Sigma}_s \mathbf{\Phi}_s = \mathbf{\Phi}_s \mathbf{\Lambda}_s. \tag{10}$$

The shape and trajectory eigenvalues and eigenvectors are related by the following equations [45]:

$$\mathbf{\Phi}_s = \mathbf{S}^{*T} \mathbf{\Phi}_t, \tag{11}$$

$$\mathbf{\Lambda}_s = \mathbf{\Lambda}_t. \tag{12}$$

Once, $\mathbf{\Phi}_t$ is known, the trajectory coefficients can be found as $\mathcal{A}^* = \mathbf{\Phi}_t{}^T \mathbf{S}^*$. Using (11), this can be simplified as

$$\mathcal{A}^* = \mathbf{\Phi}_s{}^T. \tag{13}$$

Similarly, the shape coefficients can be found as $\mathbf{\Omega}^{*T} = \mathbf{\Phi}_s{}^T \mathbf{S}^{*T}$. Using (11), this can be written as

$$\begin{aligned} \mathbf{\Omega}^* &= \mathbf{S}^* \mathbf{\Phi}_s = \mathbf{S}^* (\mathbf{S}^{*T} \mathbf{\Phi}_t), \\ &= \mathbf{\Sigma}_t \mathbf{\Phi}_t. \end{aligned}$$

Using (9), the above equation can be simplified as

$$\mathbf{\Omega}^* = \mathbf{\Phi}_t \mathbf{\Lambda}_t. \tag{14}$$

Equations (13) and (14) show that SVD basis and coefficients in trajectory representation are the same as SVD coefficients and basis for shape representation, respectively. However, in order to make the norm of eigenvectors equal to unity, appropriate normalization will be needed. This confirms the Duality Theorem for SVD basis.

3.3 Dual Representation of the Structure

Given the duality between the shape and trajectory bases, it is easy to see that nonrigid structure representation using trajectory basis will also be dual to the one using shape basis as given in [2], [5]. That is, the trajectory coefficients and trajectory basis will take the role of the shape basis and shape coefficients, respectively. Hence, the structure matrix can be written as a multiplication of an inverse projection matrix containing trajectory basis and trajectory coefficient matrix as

$$\mathbf{S}_{3F \times P} = \mathbf{\Theta}_{3F \times 3K} \mathcal{A}_{3K \times P}, \tag{15}$$

where

$$\mathbf{\Theta} = \begin{bmatrix} \theta_{11} I & \dots & \theta_{1K} I \\ \vdots & \ddots & \vdots \\ \theta_{F1} I & \dots & \theta_{FK} I \end{bmatrix}, \quad \mathcal{A} = \begin{bmatrix} \mathbf{a}_1(1) & \dots & \mathbf{a}_1(P) \\ & \vdots & \\ \mathbf{a}_K(1) & \dots & \mathbf{a}_K(P) \end{bmatrix}.$$

I is a 3×3 identity matrix and $\mathbf{a}_i(j) = [a_{xi}(j), a_{yi}(j), a_{zi}(j)]^T$. Equation (15) can also be derived directly from (4) and should be considered a dual of the shape basis representation of structure given in (6).

Although the representation power of the shape basis and trajectory basis is equal, the principal benefit of the trajectory space representation is that the trajectory basis can be predefined independent of the observed data. This follows from the observation that most trajectories are smooth and continuous in nature. In the next section, we address the question of what basis should be used for predefining natural trajectories?

4 Predefining the Trajectory Basis

The smoothness in 3D trajectories is an inherent property of most natural deforming objects. This smoothness can be

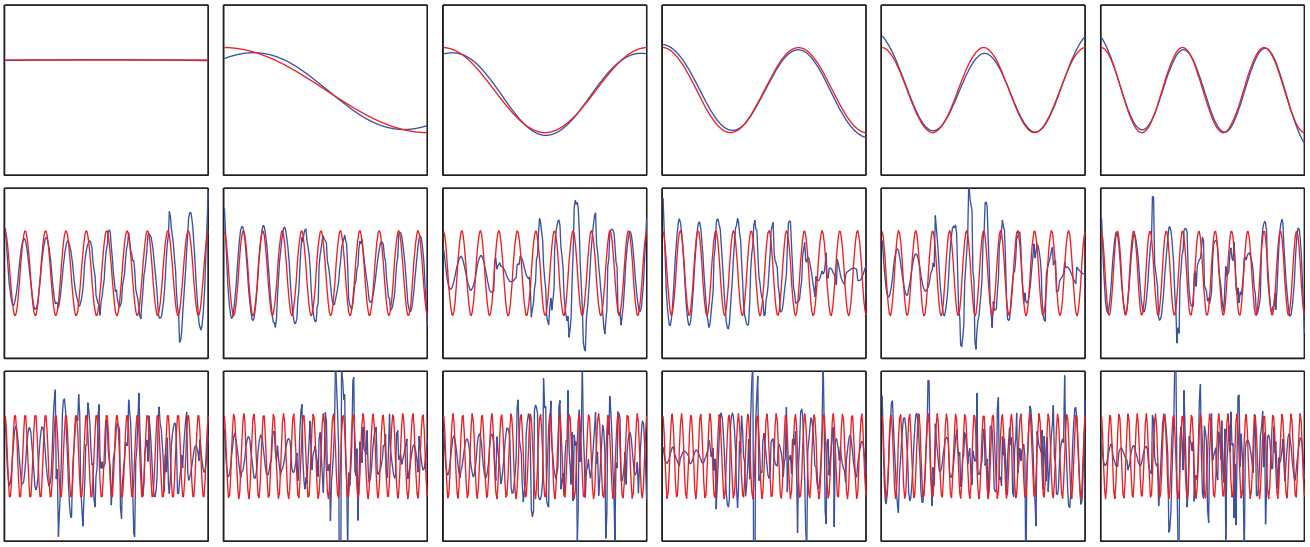

Fig. 4. The comparison of PCA (blue) and DCT (red) as the trajectory basis for the CMU motion capture data. Here, we plot the 1st-6th, 21st-26th, and 41st-46th PCA and DCT basis. The plot shows the close resemblance between the two, especially for initial PCA basis. Some of the basis have been multiplied by -1 for better visual comparison.

exploited to predefine the trajectory basis. There exist a number of predefined bases which can approximate smooth signals compactly. A few examples are the Hadamard Transform basis, the Discrete Sine Transform (DST) basis, DCT basis, and the Discrete Wavelet Transform (DWT) basis. It is generally hard to decide which predefined basis will be optimal for a specific problem. However, for certain problems, their optimal predefined basis is known. For example, it has been shown that DST and DCT are optimal for poorly correlated and highly correlated first-order Markov data, respectively [3]. Moreover, it is also possible to find a sinusoidal transform as a good substitute for PCA for higher order stationary random fields with arbitrary cross correlations [3].

The optimality of DCT for Markov models has been exploited in numerous applications, such as transform coding of speech signals [46], speech recognition [47], spectral document ranking [48], image compression [3], and face recognition [49]. In [50], Li and Wang model the motion capture data sets with a first-order Markov chain for synthesizing human motion. Our experiments on motion capture data sets also show the close resemblance of PCA and DCT as trajectory bases. This resemblance suggests that motion capture trajectories are well modeled by Markov chains.

To empirically demonstrate the suitability of the DCT basis for representing human motion, we perform an experiment to compare them to the PCA basis estimated from the CMU motion capture database [51]. It consists of almost 4,000 examples from different actors with different actions. The length and number of points in each data set differ from others. In our experiment, we first make all of the data sets zero mean by taking the world origin at the center of the object. Then, we divide the data into smaller nonoverlapping partitions of 256 frames. After this, we concatenate the (X, Y, Z) components of the trajectories of all the partitions into a single data matrix as column vectors. The number of columns of the data matrix is about six million. Finally, we compute the PCA basis of the data matrix. Since the frame-length is taken to be 256, we obtain 256 basis vectors, which are then compared to the DCT basis. Fig. 4 shows that the initial PCA basis, which contains the most energy, very closely resembles the DCT basis.[1] In order to test the representation power of DCT, we choose about 7,000 random 3D trajectories from the training data set. Then, we project these trajectories on DCT and PCA basis and reconstruct them using a varying number of bases (K). We also do the same comparison with other orthonormal transforms: DST, Hadamard, and Haar. In Figs. 5a and 5b, we plot the mean square error of PCA versus other orthonormal transforms as the number of basis vectors, K, varies. The plots show the close match of the DCT-based reconstruction with that using the PCA, whereas other transforms do not perform as well as the DCT. In Fig. 5c, we plot the mean square error of PCA and DCT basis for a motion capture data set of face. This plot also shows the close match between DCT-based reconstruction and PCA-based reconstruction.

In Section 3, we discussed the equivalence of basis and coefficients with coefficients and basis in the dual spaces and illustrated the equivalence for SVD basis. Applying the same idea to DCT, we conclude that DCT coefficients can also be thought of as shape bases as well. The illustration of this equivalence is given in Fig. 6, where we plot the dual shape basis or, equivalently, the trajectory coefficients corresponding to the DCT trajectory basis for a motion capture data set. The figure provides another example of the duality between trajectory and shape basis.

1. As the CMU Motion Capture database has not been postprocessed for removing errors in the data, there exist a lot of unnatural discontinuities in the trajectories which need to be removed. To clean the erroneous discontinuous trajectories, we compute the difference between consecutive values of all (X, Y, Z) components of the trajectory for each data set and compute their mean (μ) and standard deviation (σ). Then, we discard those trajectories for which the difference lies outside the interval of $[\mu \pm 10\sigma]$. We notice that the standard deviation threshold of 10 still leaves some discontinuous trajectories, but they are very small as compared to the total number of trajectories present in the training data.

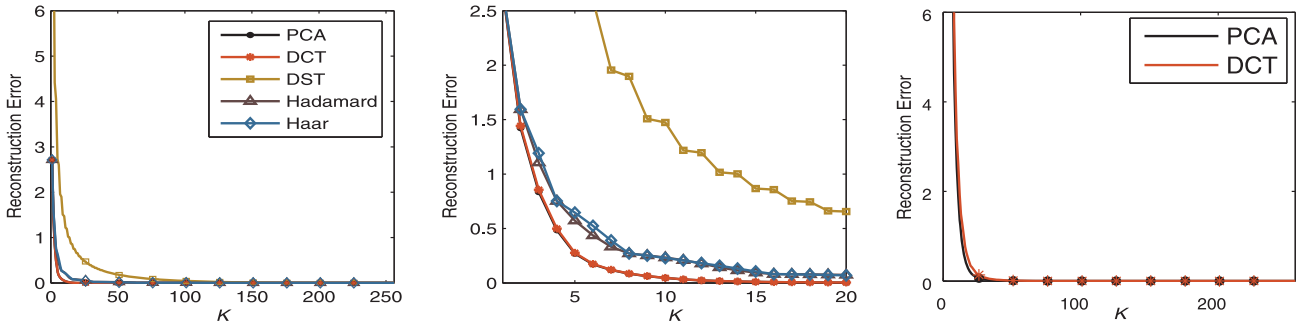

Fig. 5. (a) The reconstruction accuracy (mean square error) on a random subset of motion capture trajectories using a different number of bases (K) for different orthonormal transforms. (b) Zoomed-in view of the plot of Fig. 5a. The plots show that DCT is close to optimal for representing human motion. (c) The reconstruction accuracy (mean square error) on a motion capture face data set using different number of bases (K) for the PCA and DCT basis.

5 Structure Recovery in Trajectory Space

In this section, we will show that by using predefined trajectory basis, the nonrigid structure can be estimated. In order to find the time-varying structure $\mathbf{S}$, defined in (2), the input data is the image observation matrix $\mathbf{W}$, given in (1). $\mathbf{W}$ can be decomposed as $\mathbf{W} = \mathcal{R}\mathbf{S}$, where $\mathcal{R}$ is a $2F \times 3F$ matrix,

$$\mathcal{R} = \begin{bmatrix} \mathbf{R}_1 & & \\ & \ddots & \\ & & \mathbf{R}_F \end{bmatrix},$$

and $\mathbf{R}_t$ is a 2×3 orthographic projection matrix. Using (15), $\mathbf{W}$ can be factorized as

$$\mathbf{W} = \mathcal{R}\boldsymbol{\Theta}\mathcal{A} = \boldsymbol{\Lambda}\mathcal{A}, \tag{16}$$

where $\boldsymbol{\Lambda} = \mathcal{R}\boldsymbol{\Theta}$ is a $3F \times 3K$ matrix. The rank of $\mathbf{W}$ will be at most $3K$ if K basis vectors in each of the X, Y, and Z dimensions are used to construct $\boldsymbol{\Theta}$. This is a dual property to the rank constraint defined by [2]. Using SVD, $\mathbf{W}$ can be factorized as

$$\mathbf{W} = \hat{\boldsymbol{\Lambda}}\hat{\mathcal{A}},$$

where dimensions of $\hat{\boldsymbol{\Lambda}}$ and $\hat{\mathcal{A}}$ are $2F \times 3K$ and $3K \times P$, respectively. In general, $\hat{\boldsymbol{\Lambda}}$ and $\hat{\mathcal{A}}$ will not be equal to $\boldsymbol{\Lambda}$ and $\mathcal{A}$, respectively, because this factorization is not unique: For any invertible $3K \times 3K$ matrix $\mathbf{Q}$, $\hat{\boldsymbol{\Lambda}}\mathbf{Q}$ and $\mathbf{Q}^{-1}\mathcal{A}$ are also valid factorizations. Hence, to recover metric structure, we need to estimate the rectification matrix $\mathbf{Q}$ satisfying the following constraints:

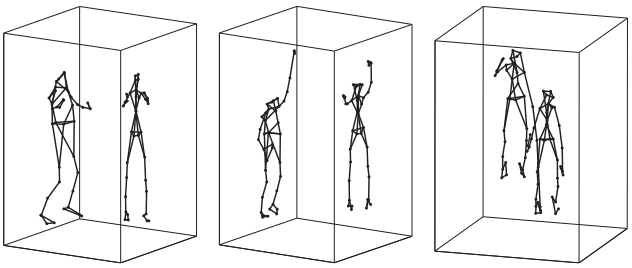

Fig. 6. The illustration of duality for DCT as a trajectory basis. In this figure, we plot the first three DCT coefficients corresponding to a motion capture data set consisting of two people. The plots show that (DCT) trajectory coefficients can be thought of as a shape basis.

$$\boldsymbol{\Lambda} = \hat{\boldsymbol{\Lambda}}\mathbf{Q}, \qquad \mathcal{A} = \mathbf{Q}^{-1}\hat{\mathcal{A}}. \tag{17}$$

To compute the rectifying transform $\mathbf{Q}$, consider the structure of $\boldsymbol{\Lambda}$, which can be written as

$$\boldsymbol{\Lambda} = \begin{bmatrix} \theta_{11}\mathbf{R}_1 & \dots & \theta_{1K}\mathbf{R}_1 \\ & \vdots & \\ \theta_{F1}\mathbf{R}_F & \dots & \theta_{FK}\mathbf{R}_F \end{bmatrix}. \tag{18}$$

Equation (18) shows that instead of estimating the complete matrix $\mathbf{Q}$, it is sufficient to estimate only three columns of $\mathbf{Q}$. In order to see this, let us define $\mathbf{Q}_{|||}$ to be the first column triple of the matrix $\mathbf{Q}$. In (17), multiplying $\hat{\boldsymbol{\Lambda}}$ with $\mathbf{Q}_{|||}$ instead of $\mathbf{Q}$ gives

$$\hat{\boldsymbol{\Lambda}}\mathbf{Q}_{|||} = \begin{bmatrix} \theta_{11}\mathbf{R}_1 \\ \vdots \\ \theta_{F1}\mathbf{R}_F \end{bmatrix}. \tag{19}$$

Equation (19) shows that camera rotations can be estimated if $\mathbf{Q}_{|||}$ is known. These rotations can be used to form the matrix $\mathcal{R}$. Once $\mathcal{R}$ is known, it can be multiplied with the (known) trajectory basis matrix $\boldsymbol{\Theta}_{3F\times 3K}$ to recover the matrix $\boldsymbol{\Lambda}_{2F\times 3K} = \mathcal{R}_{2F\times 3F}\boldsymbol{\Theta}_{3F\times 3K}$. Finally, the coefficients $\hat{\mathcal{A}}$ can be estimated by solving the following overconstrained linear system of equations:

$$\boldsymbol{\Lambda}_{2F\times 3K}\hat{\mathcal{A}}_{3K\times P} = \mathbf{W}_{2F\times P}. \tag{20}$$

Therefore, instead of estimating the whole matrix $\mathbf{Q}$, only three columns are enough for estimating nonrigid structure. Although, this approach reduces the number of unknowns in the upgrading step from $9K^2$ to $9K$, this is no longer a Maximum Likelihood approach and is suboptimal. However, our experiments show that it provides a reliable solution.

In order to estimate $\mathbf{Q}_{|||}$, orthonormality constraints of camera rotations $\mathbf{R}_i$ can be exploited, following an approach similar to [16]. Specifically, if $\hat{\boldsymbol{\Lambda}}_{2i-1:2i}$ denotes the two rows of matrix $\hat{\boldsymbol{\Lambda}}$ at positions $2i-1$ and $2i$, then we have

$$\hat{\boldsymbol{\Lambda}}_{2i-1:2i}\mathbf{Q}_{|||}\mathbf{Q}_{|||}^T\hat{\boldsymbol{\Lambda}}_{2i-1:2i}^T = \theta_{i1}^2 I_{2\times 2}, \tag{21}$$

where $I_{2\times 2}$ is an identity matrix, giving three independent constraints for each image i. Therefore, for F frames, we have $3F$ constraints and $9K$ unknowns in $\mathbf{Q}_{|||}$. Hence, at

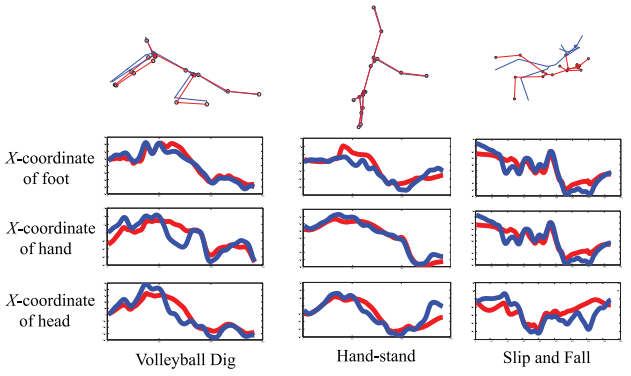

Fig. 7. Reconstruction accuracy for three actors. The X-coordinate trajectories for three different points on the actors are shown. The approximation error introduced by DCT projection has a smoothing impact on the reconstruction. Red lines indicate ground truth data and blue lines indicate reconstructed data.

least $3K$ nondegenerate[2] images are required to estimate $\mathbf{Q}_{|||}$. Once $\mathbf{Q}_{|||}$ has been computed using a nonlinear minimization routine (e.g., Levenberg Marquardt), we can estimate the rotation matrices, and therefore $\mathcal{R}$, using (19).

The constraints given in (21) are dual to the orthonormality constraints considered by Bregler et al. [2], and are computationally equivalent if the trajectory basis is unknown. However, knowing the trajectory basis beforehand leads to important advantages over [2]. First, the number of available constraints generated by each image increases to three, rather than two, since θ_{i1} is known. Hence, the total number of constraints given by (21) are $3F$ rather than $2F$, resulting in more stable estimation of the $9K$ unknowns in $\mathbf{Q}_{|||}$. Second, due to the predefined basis, the ordering of image observations in the measurement matrix $\mathbf{W}$ becomes important and is implicitly encoded in the optimization process, whereas in the traditional approaches using the shape basis, the ordering of image observations in $\mathbf{W}$ can be changed arbitrarily. Hence, the predefined trajectory basis better exploit the inherent constraints of the problem, allowing the same basis to work for many different types of natural motions. Finally, the availability of predefined basis improves the numerical stability of structure estimation. If the basis and the coefficients were to be estimated, the reconstructed structure, which is a product of the coefficients times the basis, will contain the accumulated numerical error of both. In our approach only the coefficients are being estimated, hence allowing us to reconstruct more complicated nonrigid motions with less error.

It is also pertinent to discuss whether the constraints given in (21) are sufficient to obtain a unique solution of the 3D structure. According to Xiao and Kanade [52], orthonormality constraints are inherently ambiguous and cannot be used to recover nonrigid structure in the absence of additional constraints. They linearized the quadratic terms in the orthonormality constraints by considering $\mathbf{G} = \mathbf{Q}_{|||}\mathbf{Q}_{|||}^T$ and showed that there exist ambiguous solutions of $\mathbf{G}$. Akhter et al. observed that $\mathbf{G}$ should be of rank 3 because $\mathbf{Q}_{|||}$ is of rank 3. They derived the solution space of $\mathbf{G}$ under the rank 3 constraint, and showed that all solutions lying in the solution space give unique structure recovery up to a 3×3 rotation.

2. Degenerate images are the images of an object which do not span the complete $3D$ space.

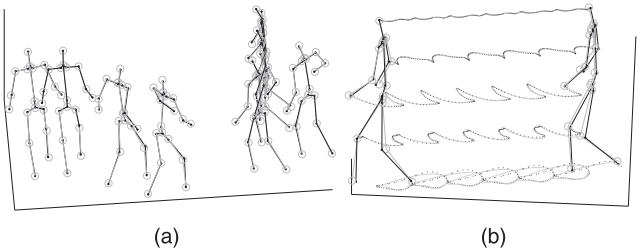

Fig. 8. Structure recovery for multiple walk data set containing eight people with different walking styles. (a) The recovered structure (gray circles) and ground truth (black dots) for one frame. (b) The recovered 3D trajectories (gray solid line) and ground truth trajectories (black dotted line) of some points of a walk in the data set. It also shows the recovered structure for the starting and ending frame.

6 Results

The proposed algorithm has been validated quantitatively on motion capture data over different actions and qualitatively on video data. We have tested the approach extensively on highly nonrigid human motion, such as walking, dancing, handstands, volleyball digs, and karate moves. Figs. 7 and 8 show a few sample reconstructions for different actors. As mentioned earlier, we choose the DCT basis for the trajectory space representation for all experiments in this paper. In subsequent experiments, we compare our approach with [6] and [52] (code kindly provided by the respective authors). The results, data, and the code used to produce the results are all shared at http://cvlab.lums.edu.pk/nrsfm.

For quantitative evaluation of the reconstruction accuracy, we use five different actions from the CMU motion capture database (drinking, pickup, yoga, stretch, and dance actions), shark, and face sequence available on the project website of Torresani et al. [7] and pants sequence by White et al. [53]. We also generate a multiple rigid body sequence by simulating points on three rigidly moving cubes. We generate synthetic camera rotations and project 3D data using these rotations to generate image observations. The camera rotation is 5 degrees per frame around the z-axis, while the overall camera motion is oscillatory with a pan of ± 45 degrees. We do not rotate the camera for the dance, shark, and face sequences since the object itself is rotating in these sequences. In our experiments on highly nonrigid objects, including handstands, karate moves, and volleyball digs, we simulate random camera motions to generate images. We normalize the structure, so that the average standard deviation of the structure matrix $\mathbf{S}$ is unity, to make comparison of error across data sets more meaningful.

Table 1 shows a quantitative comparison of our method with the shape basis approach of Torresani et al. [6] and Xiao and Kanade [52]. Torresani et al. proposed two algorithms in [6], named EM-PPCA and EM-LDS. We use EM-LDS, because it exploits both shape compactness and trajectories smoothness. This table shows both the camera rotation estimation error and structure reconstruction error. The estimated structure is valid up to a 3D rotation and translation, and the estimated rotations also have a 3D

TABLE 1
The Quantitative Comparison of the Proposed Algorithm with the Techniques Described in Xiao and Kanade [52] and Torresani et al. [6]

		Trajectory Basis			Torresani's EM-LDS			Xiao's Shape Basis	
DATASET	FRAMES	E_{rot}[1]	E_{Δ}[2]	K	E_{rot}[1]	E_{Δ}[2]	K	E_{rot}[1]	E_{Δ}[2]
DRINK	1102	0.0076	0.0254	13	0.352	0.3677	12	0.3640	0.6907
PICKUP	357	0.1559	0.2369	12	0.5027	0.5646	4	0.5793	8.4194
YOGA	307	0.1141	0.1604	11	0.9778	0.8421	4	1.2794	7.5566
STRETCH	370	0.0425	0.0602	12	0.4863	1.0161	8	0.7352	1.684
MULTIRIGID	400	2.40E-08	0.0173	4	0.0825	0.4876	3	0.0855	0.6177
DANCE	264	NA	0.2958	5	NA	0.4102	3	NA	2.9962
SHARK	240	NA	0.3121	2	NA	0.1086	2	NA	0.4565
FACE	316	NA	0.0444	5	NA	0.0529	2	NA	0.0541
PANTS	201	0.059	0.117	10	-	-	-	0.0838	0.1728

[1] E_{rot} *is the average Frobenius difference between original rotations and aligned estimated rotations.*
[2] E_{Δ} *is the average distance between original 3D points and aligned reconstructed points.*

rotation ambiguity. We therefore align them for error measurement using the procrustes method. The error measure for camera rotations is the average Frobenius norm difference between the original camera rotations and the estimated camera rotations. For structure evaluation, we compute the per frame per point euclidean distance between original 3D points and the estimated 3D points. Table 1 also shows the values of K used in our method and Torresani et al.'s method. We choose K by exhaustively trying out different numeric values between 2 and 13, and selecting the best one. We are not reporting K for Xiao and Kanade because their method automatically estimates the number of bases. Figs. 8, 9, 10, and 11 show the results on multiple walks, Stretch, Dance, and Pants sequences.

Finally, to test the proposed approach on real data, we use a face sequence from [52], a sequence from the movie "The Matrix," a sequence capturing two rigidly moving cubes, a sequence of a toy dinosaur moving nonrigidly, and a cloth sequence. For the last four sequences, the image points are tracked in a semiautomatic manner, using the approach proposed in [54], supplemented with manual correction. We show the resulting reconstructions in Figs. 12 and 13 and compare against the reconstructions obtained using the implementations of Torresani et al. [6] and Xiao and Kanade [52]. These figures show that Xiao and Kanade's method works reasonably well on the face and dinosaur sequence. Torresani's EM-LDS works on matrix sequence and to some extent on face sequence, where it only recovers the rigid component. On the other hand, proposed trajectory basis works reasonably well on all of these sequences. The value of K in EM-LDS is equal to 2 for all of the sequences. We observe that, for K greater than 2, the results get worse. In the trajectory basis approach, we use the values of K as 10, 12, 3, 2, 2 in the cloth, dinosaur, matrix, face, and cubes sequences, respectively.

7 DISCUSSION

In general, nonrigid structure from motion is an ill-posed problem with more unknowns in the structure matrix, $\mathbf{S}$, than the observations in measurement matrix, $\mathbf{W}$. To make the problem tractable and numerically stable, the solution of 3D structure $\mathbf{S}$ is constrained to lie in a compact subspace, spanned by a small number of basis shapes in previous literature. Additional constraints have also been proposed by several researchers as discussed in Section 2.

We have taken a dual view of the problem by constraining the 3D structure to a subspace spanned by basis trajectories. Since smoothness of trajectories is a ubiquitous property of natural motions, it allows us to use the *same* subspace for a variety of sequences. The resulting factorization problem is optimal up to an affine transform in $3K$ space. This is because if the basis is unknown, as is usually the case with the shape basis, the nonrigid structure

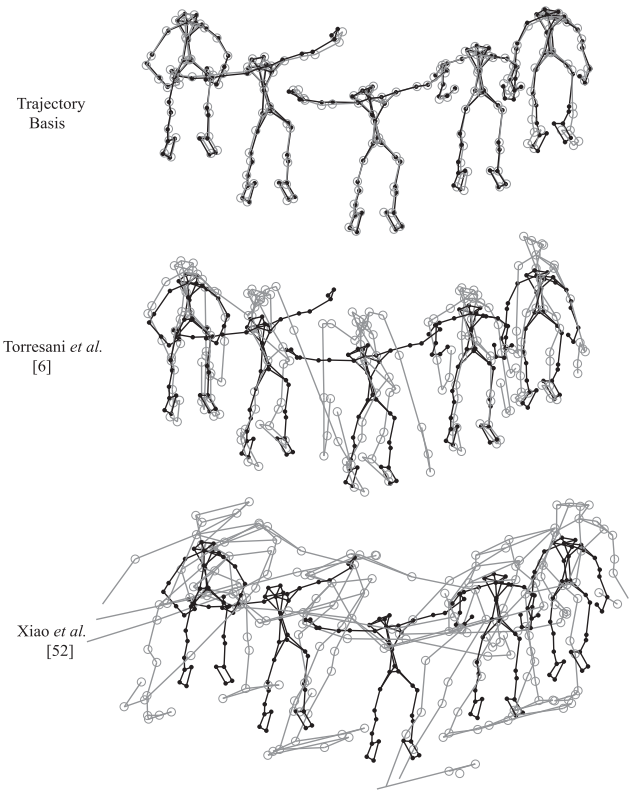

Fig. 9. Structure recovery for the Stretch data set: The black dots are the ground truth points, while the gray circles are the reconstructions by the three methods, respectively.

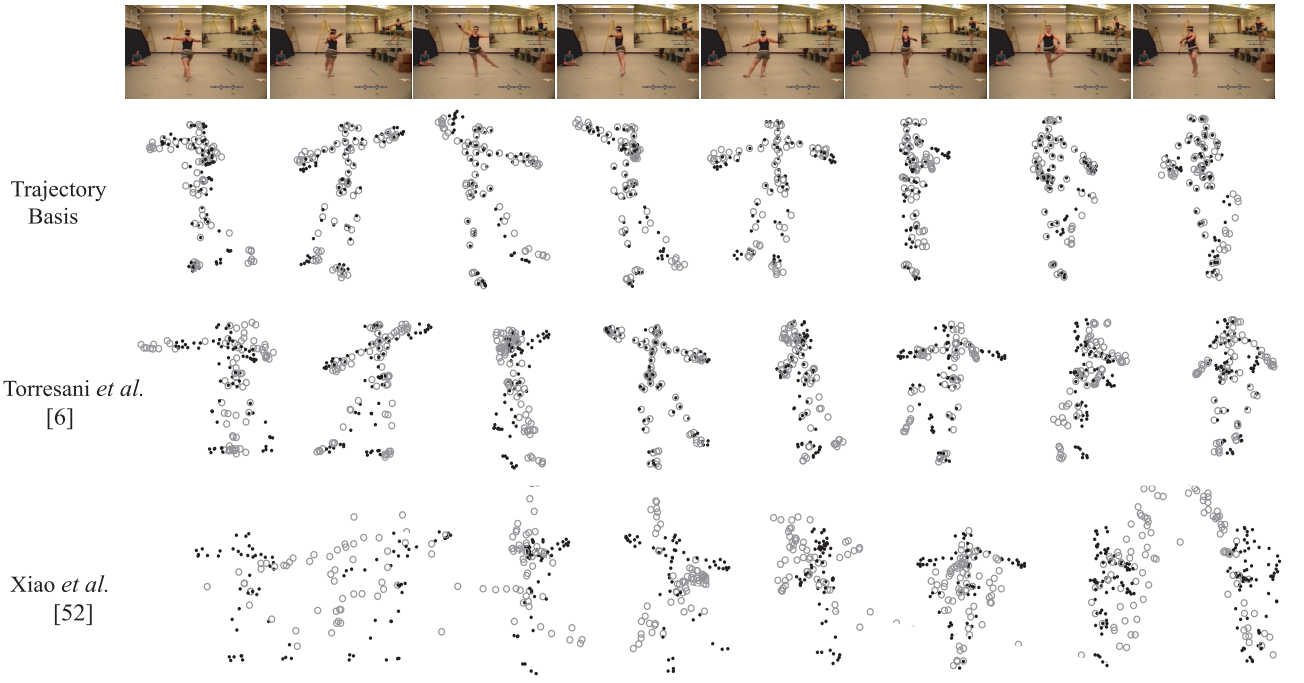

Fig. 10. The dance sequence from the CMU motion capture database. The black dots are the ground truth points, while the gray circles are the reconstructions by the three methods, respectively.

from motion problem is a trilinear problem—variable interactions occur between rotations, coefficients, and basis. However, if the basis is known, as in the DCT trajectory space, the nonrigid structure from problem is bilinear—variable interactions occur only between rotations and coefficients. Current methods solve a trilinear estimation problem by successively solving two bilinear problems, by grouping two sets of variables together, applying SVD, and then applying further optimization to the subblock grouped variables. SVD, which inherently is a bilinear decomposition, gives us the optimal decomposition for a bilinear problem, which is what our approach considers, whereas using SVD to solve a trilinear decomposition (in two steps) is a relaxation and not optimal.

The key relationship which determines successful reconstruction is the one between the amount of camera motion and the degree of deformation of the object, the latter being measured by the number of basis K needed to approximate it. This relationship is analogous to our understanding of stereo, where numerical stability increases with increasing baseline. In nonrigid structure from motion, we have found the solution to be more stable as the amount of per-frame camera motion increases. In other words, more complicated nonrigid motions (those that need a higher K for reasonable representation) can be reconstructed if the average per-frame camera motion is larger.

As a practical demonstration of the impact of per-frame camera motion on reconstruction stability, we synthetically constructed various magnitudes of per-frame camera rotation and constructed $\mathbf{\Lambda}$ for different values of K. The first row of Fig. 14 shows the reconstruction stability, measured by the condition number of $\mathbf{\Lambda}^T\mathbf{\Lambda}$, as K is

Fig. 11. Pants sequence: The first row shows the ground truth structure in four selected frames, each with two views of the object, and the second row is our reconstruction.

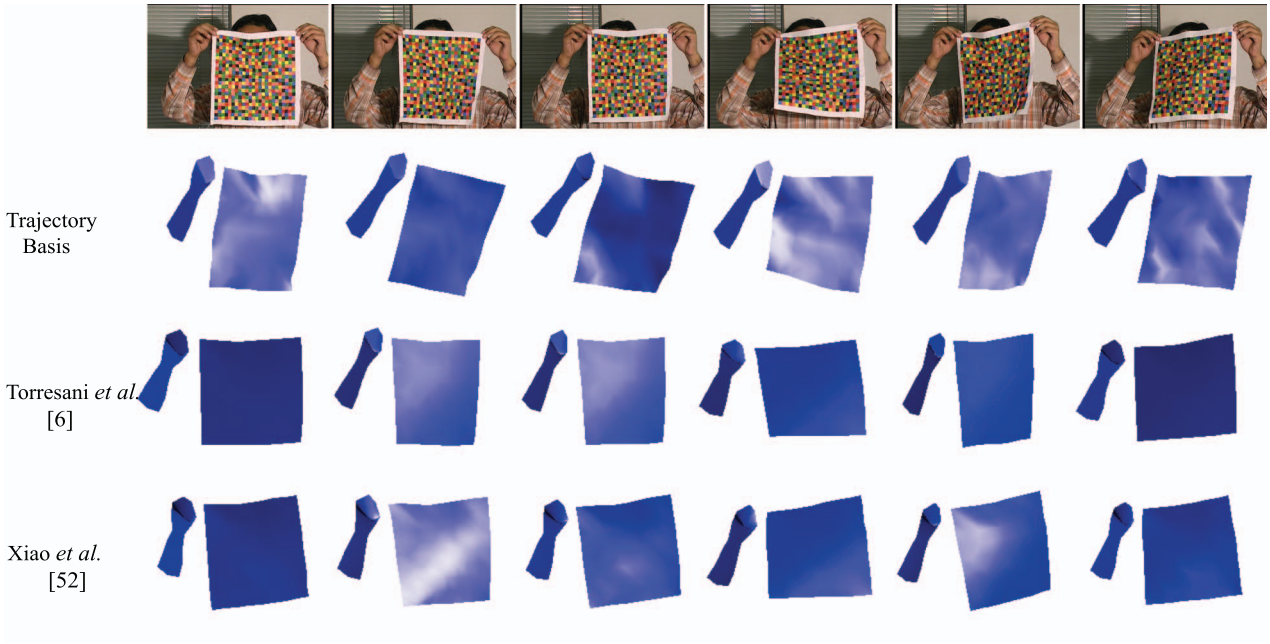

Fig. 12. Results on the cloth sequence. The first row shows a few frames from the sequence. The rest of the rows are the structure reconstruction.

varied between 2 and 8, for 200, 400, and 800 frames, at different angular velocities per frame. The plots confirm intuition—the smaller the degree of object deformation and the larger the camera motion, the better the condition number tends to be. For further verification, we computed the reconstruction error for a motion capture walk sequence. We assumed the camera rotations to be known and estimated the nonrigid structure using (18), (20), and (15), as discussed in Section 5. The reconstruction error, shown in the second row of Fig. 14, generally follows the same trend as that of condition number of $\mathbf{\Lambda}^T\mathbf{\Lambda}$, for a range of appropriate values of K, except that we also see that for a larger value of K, we get lower reconstruction error if optimization is stable. This understanding is also expected to be valid for shape basis since the role of coefficients and basis is interchanged if ideal rotations and coefficients were known and shape basis were to be estimated, exactly similar results would be generated due to the duality theorem. This hints at the possibility that this may be a fundamental limitation of the nonrigid structure from motion problem. The more the object deforms, the more the motion needed to constrain the solution of the structure. In the limiting case of no deformation (rigid), only two views with small disparity would be enough.

It is also instructive to study the impact of varying K as an independent variable, on the estimation of camera rotations and structure. Fig. 15 shows the effect of changing K for four different motion capture data sets. For certain K, the nonlinear optimization may not always converge, as indicated in Figs. 15a and 15e. However, in general, neighboring values of K do not make much of a difference on the estimation of camera rotation, and also result in similar structure reconstruction error if the per-frame camera motion is sufficiently large (Fig. 15b, 15c, 15d, 15f, 15g). For slow camera motion, the reconstruction error gets progressively worse as K increases, as shown in Figs. 15f and 15g. This is presumably because the increase in allowable nonrigidity for larger K and the corresponding increase in unknowns needs to be matched by more stringent constraints imposed by faster camera motion.

It is also relevant to mention here that while the bilinear relationship between unknowns results in an optimal factorization through SVD, the nonlinear solution of the metric upgrade using only first three columns of $\mathbf{Q}$ is nonoptimal. It should be possible to formulate a maximum likelihood solution involving all terms of $\mathbf{Q}$. However, in our approach, we chose to prefer the reduced number of unknowns in $\mathbf{Q}_{|||}$, which contains just $9K$ unknowns rather than $9K^2$ in the full $\mathbf{Q}$ matrix.

Finally, using a predefined basis may not result in a highly compact representation on every sequence compared to data-dependent basis. Hence, there may be a need to increase the number of basis, K, for certain sequences to achieve a reasonable representation. While a higher K implies lesser numerical stability, it is counterbalanced in our approach by the reduction in unknowns due to the availability of predefined basis. Quantitative and qualitative evaluation of our method shows that the advantage of the latter is significant and having a predefined basis improves the numerical stability of the solution. This is especially apparent in sequences which have a high nonrigidity.

8 Conclusions

We show that structure recovery from motion information is possible for smoothly deforming objects without requiring prior knowledge of the object. We propose the trajectory basis to exploit this smoothness and show that our approach is dual to the traditional approach of expressing nonrigid structure as a linear combination of basis shapes. Unlike the traditional approach and its variants, which require the learning of object-specific shape basis, for our approach trajectory basis can be predefined. We demonstrate that the DCT basis can compactly represent the motion of a wide variety of natural deformations, giving near optimal compaction for human motion. Using DCT as

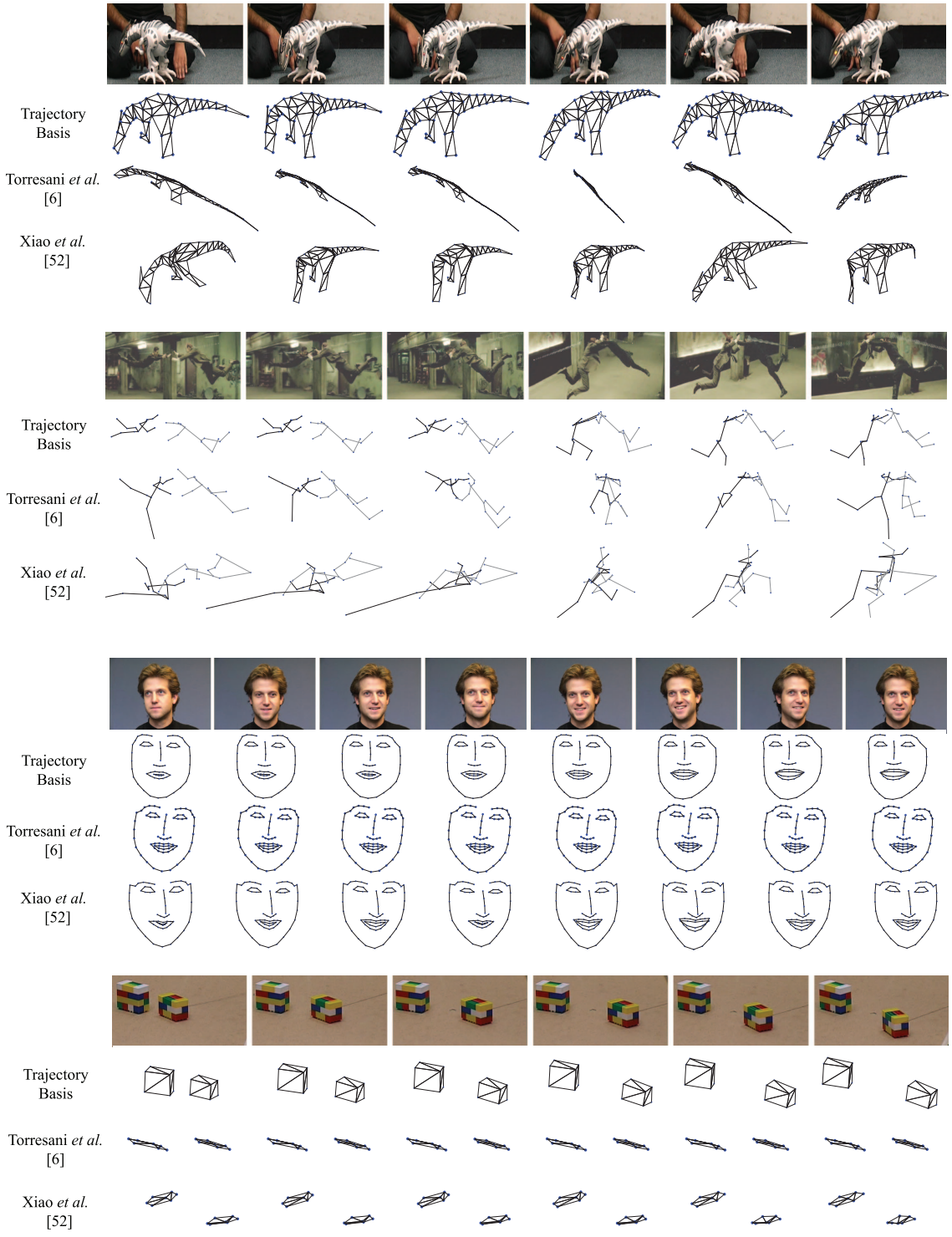

Fig. 13. Results on the Dinosaur, Matrix, face, and Cubes sequences. K was set to 12, 3, 2, and 2, respectively.

an object independent basis has the advantage of a smaller number of unknowns and more stable estimation. We report satisfactory results on dynamic data sets, such as piecewise rigid motion, facial expressions, actors dancing, walking, and performing yoga. Our experiments demonstrate the inherent relationship between camera motion, degree of object deformation, and reconstruction stability. Reconstruction stability increases as the camera motion increases or the degree of deformation decreases. Future directions of research are to explore better techniques of optimization and developing a synergistic approach to use both the shape and trajectory bases concurrently.

Acknowledgments

This research was partly supported by a grant from the Higher Education Commission of Pakistan. The authors acknowledge Fernando De La Torre for useful discussions.

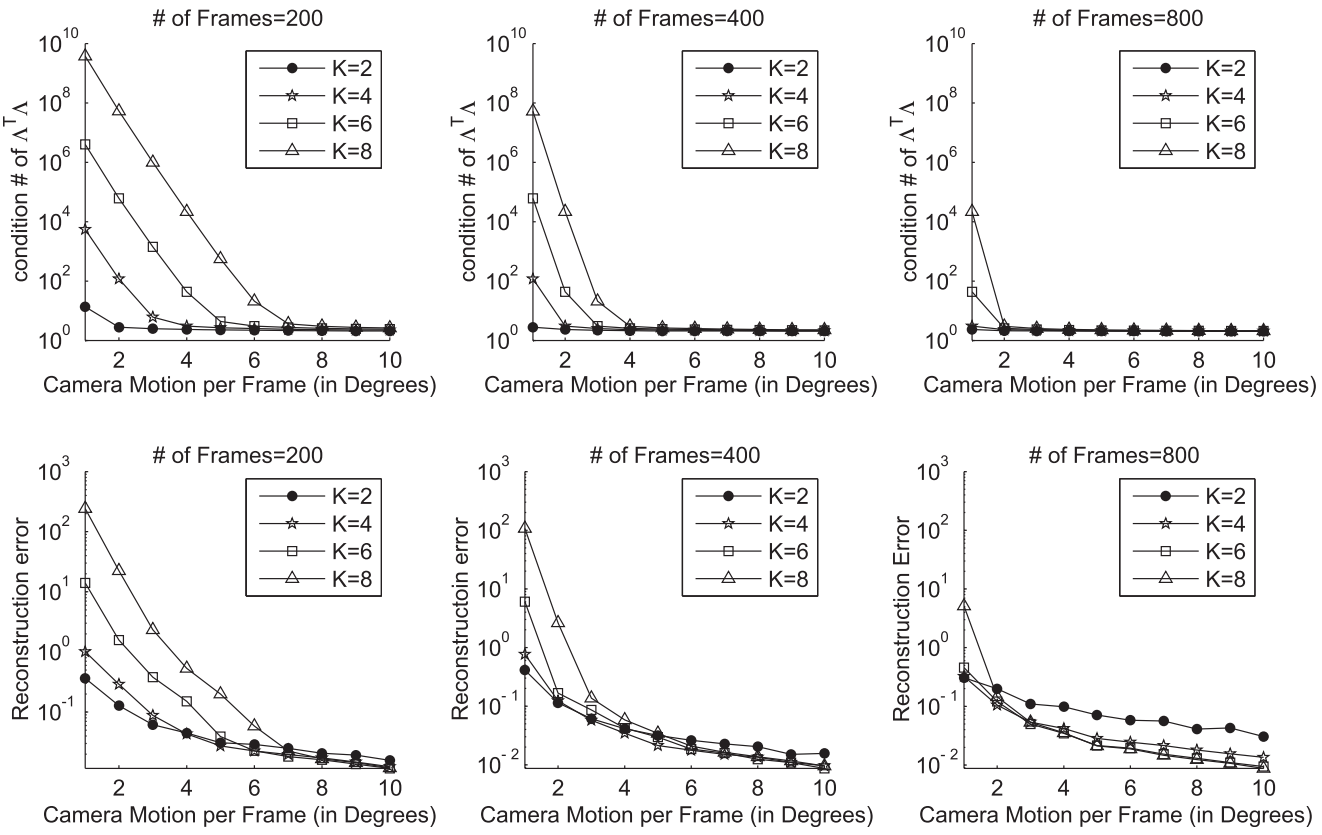

Fig. 14. The effect of increasing camera motion on reconstruction stability. First row: Reconstruction stability is measured in terms of the condition number of matrix $\mathbf{\Lambda}^T\mathbf{\Lambda}$ with different values of K and different values of F. Second row: The reconstruction error on a walk data set using known rotations with different values of K and different values of F. Synthetic rotations were generated by revolving the camera around the z-axis and camera motion was measured in terms of the angle the camera moved per frame.

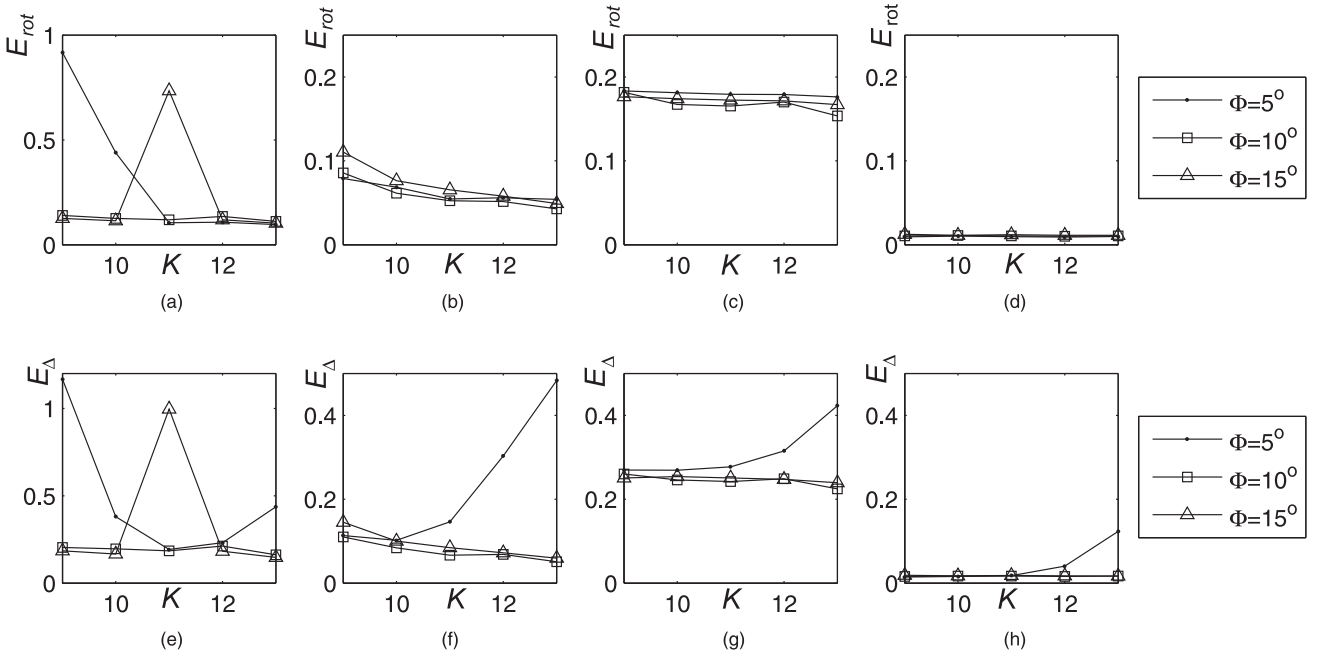

Fig. 15. The estimation error in camera rotations (E_{rot}) and structure recovery (E_Δ) for different values of K per frame camera motion (denoted by ϕ here). (a)-(d) Rotation estimation and (e)-(h) structure estimation error for the yoga, stretch, pickup, and drink data sets, respectively.

The authors also thank Jing Xiao, Lourdes Agapito, Iain Matthews, and Lorenzo Torresani for making their code or data available. The motion capture data used in this project were obtained from http://mocap.cs.cmu.edu. Yaser Sheikh was supported by US National Science Foundation grant IIS-0916272.

REFERENCES

[1] G. Johansson, "Visual Perception of Biological Motion and a Model for Its Analysis," *Perception and Psychophysics,* vol. 14, pp. 201-211, 1973.

[2] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering Non-Rigid 3D Shape from Image Streams," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 690-696, 2000.

[3] A.K. Jain, *Fundamentals of Digital Image Processing.* Prentice Hall, 1989.

[4] K. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications.* Academic, 1990.

[5] J. Xiao, J. Chai, and T. Kanade, "A Closed Form Solution to Non-Rigid Shape and Motion Recovery," *Int'l J. Computer Vision,* vol. 67, pp. 233-246, 2006.

[6] L. Torresani, A. Hertzmann, and C. Bregler, "Nonrigid Structure-from-Motion: Estimating Shape and Motion with Hierarchical Priors," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 5, pp. 878-892, May 2008.

[7] L. Torresani, A. Hertzmann, and C. Bregler, "Learning Non-Rigid 3D Shape from 2D Motion," *Proc. 19th Ann. Conf. Neural Information Processing Systems,* 2005.

[8] M. Brand, "A Direct Method for 3D Factorization of Nonrigid Motion Observed in 2D," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, 2005.

[9] A.D. Bue, F. Smeraldi, and L. Agapito, "Non-Rigid Structure from Motion Using Ranklet-Based Tracking and Non-Linear Optimization," *Image and Vision Computing,* vol. 25, pp. 297-310, 2007.

[10] L. Torresani, D. Yang, E. Alexander, and C. Bregler, "Tracking and Modeling Non-Rigid Objects with Rank Constraints," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 493-500, 2001.

[11] M. Brand, "Morphable 3D Models from Video," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, 2001.

[12] H. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene from Two Projections," *Nature,* vol. 293, pp. 133-135, 1981.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision,* second ed. Cambridge Univ. Press, 2004.

[14] O. Faugeras and Q.T. Luong, *The Geometry of Multiple Images.* MIT Press, 2001.

[15] Y. Ma, S. Soatto, J. Kosecka, and S.S. Sastry, *An Invitation to 3-D Vision: From Images to Geomatric Models.* Springer, 2003.

[16] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorization Method," *Int'l J. Computer Vision,* vol. 9, pp. 137-154, 1992.

[17] L. Kontsevich, M. Kontsevich, and A. Shen, "Two Algorithms for Reconstructing Shapes," *Optoelectronics, Instrumentation and Data Processing,* vol. 5, pp. 76-81, 1987.

[18] S. Ullman, "Maximizing Rigidity: The Incremental Recovery of 3-D Structure from Rigid and Rubbery Motion," *Perception,* vol. 13, pp. 255-274, 1984.

[19] S. Chen, "Structure from Motion without the Rigidity Assumption," *Proc. IEEE CS Workshop Computer Vision,* pp. 105-112, 1985.

[20] S. Chen Penna, "Shape and Motion of Non-Rigid Bodies," *Computer Vision, Graphics, and Image Processing,* vol. 36, pp. 175-207, 1986.

[21] D. Shulman and J.Y. Aloimonos, "(Non-)Rigid Motion Interpretation: A Regularized Approach," *Proc. Royal Soc. London Series B,* vol. 233, pp. 217-234, 1988.

[22] J. Costeira and T. Kanade, "A Multibody Factorization Method for Independently Moving Objects," *Int'l J. Computer Vision,* vol. 49, pp. 159-179, 1998.

[23] L. Wolf and A. Shashua, "On Projection Matrices $\mathcal{P}^k \rightarrow \mathcal{P}^2$, $k = 3, \ldots, 6$, and Their Application in Computer Vision," *Int'l J. Computer Vision,* vol. 48, pp. 53-67, 2002.

[24] M. Han and T. Kanade, "Reconstruction of a Scene with Multiple Linearly Moving Objects," *Int'l J. Computer Vision,* vol. 59, pp. 285-300, 2004.

[25] A. Gruber and Y. Weiss, "Multibody Factorization with Uncertainty and Missing Data Using the EM Algorithm," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 707-714, 2004.

[26] R. Vidal and R. Hartley, "Motion Segementaion with Missing Data Using PowerFactorization and GPCA," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 310-316, 2004.

[27] A.D. Bue, X. Llado, and L. Agapito, "Non-Rigid Metric Shape and Motion Recovery from Uncalibrated Images Using Priors," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2006.

[28] A.D. Bue, "A Factorization Approach to Structure from Motion with Shape Priors," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2008.

[29] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd, "Coarse-to-Fine Low-Rank Structure-from-Motion," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2008.

[30] J. Yan and M. Pollefeys, "A Factorization-Based Approach to Articulated Motion Recovery," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 815-821, 2005.

[31] J. Yan and M. Pollefeys, "Automatic Kinematic Chain Building from Feature Trajectories of Articulated Objects," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, 2006.

[32] P. Tresadern and I. Reid, "Articulated Structure from Motion by Factorization," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2005.

[33] M. Paladini, A.D. Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito, "Factorization for Non-Rigid and Articulated Structure Using Metric Projections," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2009.

[34] V. Rabaud and S. Belongie, "Linear Embeddings in Non-Rigid Structure from Motion," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2009.

[35] I. Akhter, Y. Sheikh, and S. Khan, "In Defense of Orthonormality Constraints for Nonrigid Structure from Motion," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2009.

[36] J. Xiao and T. Kanade, "Uncalibrated Perspective Reconstruction of Deformable Structures," *Proc. 10th IEEE Int'l Conf. Computer Vision,* vol. 2, pp. 1075-1082, 2005.

[37] R. Vidal and D. Abretske, "Nonrigid Shape and Motion from Multiple Perspective Views," *Proc. European Conf. Computer Vision,* 2006.

[38] R. Hartley and R. Vidal, "Perspective Nonrigid Shape and Motion Recovery," *Proc. 10th European Conf. Computer Vision,* 2008.

[39] V. Rabaud and S. Belongie, "Rethinking Non-Rigid Structure from Motion," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* 2008.

[40] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Nonrigid Structure from Motion in Trajectory Space," *Proc. Neural Information Processing Systems,* 2008.

[41] S. Carlsson and D. Weinshall, "Dual Computation of Projective Shape and Camera Positions from Multiple Images," *Int'l J. Computer Vision,* vol. 27, no. 3, pp. 227-241, 1998.

[42] A. Shashua, "Trilinear Tensor: The Fundamental Construct of Multiple-View Geometry and Its Applications," *Proc. Int'l Workshop Algebraic Frames for the Perception-Action Cycle,* 1997.

[43] L. Zelnik-Manor and M. Irani, "Temporal Factorization vs. Spatial Factorization," *Proc. Eighth European Conf. Computer Vision,* 2004.

[44] S. Olsen and A. Bartoli, "Implicit Non-Rigid Structure-from-Motion with Priors," *J. Math. Imaging and Vision,* vol. 31, nos. 2/3, pp. 233-244, 2008.

[45] G. Golub and W. Kahan, "Calculating the Singular Values and Pseudo-Inverse of a Matrix," *J. SIAM: Series B, Numerical Analysis,* vol. 2, pp. 205-224, 1965.

[46] R. Zelinski and P. Noll, "Adaptive Transfrom Coding of Speech Signals," *IEEE Trans. Acoustics, Speech, and Signal Processing,* vol. 25, no. 4, pp. 299-309, Aug. 1977.

[47] J. Huang and Y. Zhao, "A DCT-Based Fast Signal Subspace Technique for Robust Speech Recognition," *IEEE Trans. Speech and Audio Processing,* vol. 8, no. 6, pp. 747-751, Nov. 2000.

[48] L.A. Park, M. Palaniswami, and K. Ramamohanarao, "A Novel Document Ranking Method Using the Discrete Cosine Transform," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 1, pp. 130-135, Jan. 2005.

[49] Z.M. Hafed and N.D. Levine, "Face Recognition Using the Discrete Cosine Transform," *Int'l J. Computer Vision,* vol. 43, no. 3, pp. 167-188, 2001.

[50] H. Yan Li and T. Wang, "Motion Texture: A Two-Level Statistical Model for Character Motion Synthesis," *ACM Trans. Graphics,* vol. 21, no. 3, pp. 465-472, 2002.

[51] Carnegie-Mellon Mocap Database, http://mocap.cs.cmu.edu/, 2003.

[52] J. Xiao and T. Kanade, "Non-Rigid Shape and Motion Recovery: Degenerate Deformations," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 668-675, 2004.

[53] R. White, K. Crane, and D. Forsyth, "Capturing and Animating Occluded Cloth," *Proc. ACM SIGGRAPH,* 2007.

[54] A. Datta, Y. Sheikh, and T. Kanade, "Linear Motion Estimation for Systems of Articulated Planes," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, 2008.

Ijaz Akhter received the BSc and MSc degrees in physics from the University of the Punjab, Pakistan, in 1999 and 2001, respectively. He received the MS degree in computer science from Lahore University of Management Sciences, Pakistan, in 2006, where he is currently working toward the PhD degree in computer science. He was a lab associate at Disney Research, Pittsburgh, Pennsylvania, in 2009-2010, where he worked on automated techniques for labeling and reconstruction of raw motion capture data. His research interests lie in the representation and modeling of dynamic objects, particularly in developing robust techniques for inferring 3D structure of nonrigid objects. He is a student member of the IEEE.

Yaser Sheikh received the doctoral degree from the University of Central Florida in 2006. He is an assistant research professor at the Robotics Institute, at Carnegie Mellon University. His research focus is computer vision, primarily in analyzing dynamic scenes including human activity analysis, dynamic 3D reconstruction, mobile camera networks, and nonrigid motion estimation. He won the Honda Initiation Award in 2010, the best paper award at SCA 2010, ICCV THEMIS in 2009, and the Hillman Fellowship for Excellence in Computer Science Research in 2004. He is a member of the IEEE and the IEEE Computer Society.

Sohaib Khan received the BE degree in electronics engineering from the GIK Institute of Engineering Sciences and Technology, Topi, Pakistan, in 1997, and the PhD degree in computer science in 2002 from the University of Central Florida, specializing in computer vision. He is an associate professor of computer science at the LUMS School of Science and Engineering, Lahore, Pakistan, and the founding director of the Computer Vision Lab (http://cvlab.lums.edu.pk) at LUMS. His research interests broadly span the areas of image and video analysis, including image registration, multiple camera surveillance systems, structure from motion and satellite, and aerial image processing. He is a member of the IEEE.

Takeo Kanade received the doctoral degree in electrical engineering from Kyoto University, Japan, in 1974. He is the U.A. and Helen Whitaker University Professor of Computer Science and Robotics and the director of the Quality of Life Technology Engineering Research Center at Carnegie Mellon University. After holding a faculty position in the Department of Information Science, Kyoto University, he joined Carnegie Mellon University in 1980. He was the director of the Robotics Institute from 1992 to 2001. He also founded the Digital Human Research Center in Tokyo and served as the founding director from 2001 to 2010. Dr. Kanade works in multiple areas of robotics: computer vision, multimedia, manipulators, autonomous mobile robots, medical robotics, and sensors. He has written more than 350 technical papers and reports in these areas, and holds more than 20 patents. He has been the principal investigator of more than a dozen major vision and robotics projects at Carnegie Mellon. Dr. Kanade has been elected to the National Academy of Engineering and the American Academy of Arts and Sciences. He is a fellow of the IEEE, a fellow of the ACM, a founding fellow of the American Association of Artificial Intelligence (AAAI), and the former and founding editor of the *International Journal of Computer Vision*. He has received several awards, including the Franklin Institute Bower Prize, Okawa Award, C&C Award, Tateishi Grand Prize, Joseph Engelberger Award, IEEE Robotics and Automation Society Pioneer Award, FIT Accomplishment Award, and IEEE PAMI-TC Azriel Rosenfeld Lifetime Accomplishment Award.