Multiview Structure from Motion in Trajectory Space

Aamer Zaheer* Ijaz Akhter* Mohammad Haris Baig Shabbir Marzban Sohaib Khan
{az, akhter, mohammad.haris, shabbir.marzban, sohaib}@lums.edu.pk

LUMS School of Science and Engineering, Lahore, Pakistan
cvlab.lums.edu.pk

Abstract

Most nonrigid objects exhibit temporal regularities in their deformations. Recently it was proposed that these regularities can be parameterized by assuming that the nonrigid structure lies in a small dimensional trajectory space. In this paper, we propose a factorization approach for 3D reconstruction from multiple static cameras under the compact trajectory subspace representation. Proposed factorization is analogous to rank-3 factorization of rigid structure from motion problem, in transfromed space. The benefit of our approach is that the 3D trajectory basis can be directly learned from the image observations. This also allows us to impute missing observations and denoise tracking errors without explicit estimation of the 3D structure. In contrast to standard triangulation based methods which require points to be visible in at least two cameras, our approach can reconstruct points, which remain occluded even in all the cameras for quite a long time. This makes our solution especially suitable for occlusion handling in motion capture systems. We demonstrate robustness of our method on challenging real and synthetic scenarios.

1. Introduction

Nonrigid structures exhibit spatial and temporal regularities in their deformations. In nonrigid structure from motion, these regularities are usually exploited using subspace reduction techniques [5, 3]. Akhter *et al*. [3] proposed a trajectory based representation of the nonrigid structure to exploit the temporal smoothness in the trajectroies. In contrast to nonrigid structure from motion, multiview stereo methods and rigid structure from motion, work on the basic principle of triangulation i.e. a 3D point should lie at the intersection of the rays formed by its 2D projection on the image plane and camera center. In this paper, we show that the triangulation and compact trajectory subspace assumption can be combined, into a robust factorization approach for multiview structure reconstruction. Proposed method allows points to appear and disappear at any time. Points which are visible in just one camera or the points which remain occluded in all the cameras, even for a long duration of time can be reconstructed accurately. Proposed method is robust to noise in the image observations.

It was observed that the predefined basis like Discrete Cosine Transform (DCT) can approximate smooth trajectories very well [3]. Predefining the trajectory basis simplifies the estimation procedure on one hand, whereas compromise the representation quality on the other. Hence, it is hard to argue, which of these two approaches is better. We suggest a mixture of both of these approaches to estimate trajectory basis. Our key observation is that the trajectory basis can be directly learned from the image observations. Considering the 3D trajectory of a point and its image as a 2D trajectory in a static orthographic camera, we show that the X- and Y-components of the 2D trajectory are a linear combination of the X-, Y- and Z- components of the 3D trajectory. Therefore the 2D trajectory and the 3D trajectory share the same trajectory space and optimal trajectory basis can be learned from the image observations.

The observation that the 3D and 2D trajectory bases share the same trajectory space, gives us one more theoretical insight into the problem. We show that the 2D trajectory coefficients and 3D trajectory coefficients are related by a rank-3 factorization. This relation is analogous to the factorization of image observation matrix into the 3D structure and orthographic camera matrices in rigid structure from motion problem [18]. Due to the simplicity of estimation, proposed approach is equally stable as that of Tomasi and Kanade rigid structure factorization.

We tested the proposed technique on challenging synthetic and real sequences. We also evaluated our method on perspective cameras. Our experiments under noise, and various level of missing observations, highlight the significance of incorporating trajectory compactness assumption in multi-camera 3D reconstruction.

*Joint First Authors

2. Related work

The method of triangulation for 3D reconstruction is one of the most interesting and well studied shape cues in Computer Vision literature. Though it shows robust results for rigid 3D reconstruction but non-rigid structure is more difficult to track (due to non-rigid local deformations and scarcity of good visual tracking cues) and reconstruct (due to self occlusions, noise in tracking, trajectory labeling errors and correspondence problems). In most industry applications of motion capture, tracked infra-red markers are triangulated in a bundle adjustment framework. Infra-red markers reduce the tracking problems but self occlusions and broken trajectory labeling problems still remain. Therefore, good reconstruction still requires a dozen (or more) cameras and days of manual intervention for any reconstruction of significant size and duration [12, 20]. Hardware help has also been used in sensor fusion where constraints from multiple sensors are combined to produce robust results at significant additional cost [14].

The setup cost and robustness issues have led most recent work to reduce the problem complexity by imposing additional constraints on the shape or motion of the object(s) to be reconstructed. Shape constraints include articulated and parametric shape models [7, 9, 15, 17]. Such shape constraints struggle in characterizing most interesting non-rigid structures by imposing restrictive constraints. Constraints on motion are more widely applicable in complex scenarios since all realistic and interesting motion is limited by its speed and kinematic constraints to produce large temporal inconsistency. Some recent papers [16, 4, 6, 10, 11, 21] try to use temporal consistency in a local context and show promising results but lack a stable and global constraint formulation.

Monocular non-rigid reconstruction is usually posed as a non-rigid structure and motion problem where both the camera as well as the structure can move. This ill-posed problem may be solved using constraints on shape and motion of the non-rigid objects. Subspace reduction is the most popular approach in this regard which assumes that non-rigid structure lies in a linear subspace (based on shape or trajectory) of lower dimensionality. The shape constraints showed promising results but required the simultaneous computation of shape basis, shape coefficients and camera parameters, leading to a very difficult estimation problem [5, 19]. In contrast, the trajectory formulation [3] simplifies the reconstruction by using data-independent DCT basis but lacks the optimality of data dependent basis.

We combine the power of triangulation with temporal constraints in trajectory space [3, 13] in a principled way without making any restrictive assumption about the structure. Moreover, we show that data dependent trajectory bases can be pre-computed directly from 2D data thus increasing the compactness and applicability of trajectory subspace without requiring a simultaneous computation of basis, coefficients and cameras (like the shape basis methods). This compactness of representation and principled modeling of temporal consistency allows us to compute a motion capture comparable reconstruction with very few cameras while explicitly handling the noise and missing data in a provided tracking. Our factorization based approach builds on a linear relationship between 2D and 3D trajectory coefficients derived in the following section.

3. Trajectory space parametrization

Let us consider a nonrigid structure, consisting of P points, sampled at F time instances. We represent the trajectory of jth point as the following $3 \times F$ matrix

$$\mathbf{T}_j = \left(\mathbf{X}_{1j}^T \ \cdots \ \mathbf{X}_{Fj}^T\right),$$

where $\mathbf{X}_{fj} = [X_{fj} \ Y_{fj} \ Z_{fj}]$ denotes the 3D coordinates of the jth point at time instance f. We consider C static orthographic cameras viewing this trajectory. We denote $\mathbf{t}_{ij}$ as the image of the jth trajectory, viewed by the ith camera. The $3 \times F$ matrix $\mathbf{t}_{ij}$ is given by the following relation

$$\mathbf{t}_{ij} = \mathbf{R}_i \mathbf{T}_j - \mathbf{o}_i,$$

where $\mathbf{o}_i$ is a $2 \times F$ matrix containing the image of the world origin in ith camera as its columns. By combining the information of all the points in all cameras, we can write

$$\mathbf{W} = \mathbf{RS} - \mathbf{O}, \tag{1}$$

where

$$\mathbf{W} = \begin{pmatrix} \mathbf{t}_{11} & \cdots & \mathbf{t}_{1P} \\ \vdots & \ddots & \vdots \\ \mathbf{t}_{C1} & \cdots & \mathbf{t}_{CP} \end{pmatrix}, \mathbf{R} = \begin{pmatrix} \mathbf{R}_1 \\ \vdots \\ \mathbf{R}_C \end{pmatrix}$$

$$\mathbf{S} = (\mathbf{T}_1 \ \cdots \ \mathbf{T}_P) \text{ and } \mathbf{O} = \begin{pmatrix} \mathbf{o}_1 & \cdots & \mathbf{o}_1 \\ \vdots & \ddots & \vdots \\ \mathbf{o}_C & \cdots & \mathbf{o}_C \end{pmatrix}.$$

We can set $\mathbf{O} = \mathbf{0}$ in Equation 1 by adopting the convention (suggested in [18]) that world origin lies at the center of 3D structure (i.e. center of all points at *all* time instances), camera origin lies at the center of its 2D projection and the world origin gets imaged to the camera origin. Hence we can write

$$\mathbf{W}_{2C \times FP} = \mathbf{R}_{2C \times 3} \mathbf{S}_{3 \times FP}, \tag{2}$$

where the subscripts denote the size of each matrix. Equation 2 shows that nonrigid structure estimation from multiple static cameras can be reduced to the rigid structure from motion problem. We notice that in Equation 2, it is not required that all trajectories should be of the same length. In fact points can appear or disappear at any time. If the length of the jth trajectory is F_j, then the dimensions of $\mathbf{W}$ and $\mathbf{S}$ will become $2C \times \sum_j F_j$ and $3 \times \sum_j F_j$, respectively.

Equation 2 is not imposing any spatial or temporal regularities on the nonrigid structure. Since the nonrigid structure mostly exhibits such regularities, therefore they can be exploited to increase the robustness of the structure estimation. According to the trajectory model of Akhter *et al*. [3] the X, Y and Z components of the jth trajectory can be approximated as a multiplication of the trajectory coefficients and the trajectory basis as follows

$$\mathbf{T}_j = \mathbf{A}_j \mathbf{\Theta}_j, \tag{3}$$

where $\mathbf{\Theta}_j$, a $K_j \times F_j$ matrix, consists of the K_j trajectory basis along its rows and $K_j \ll F_j$. F_j is the length of the jth trajectory. $\mathbf{A}_j$ is a $3 \times K_j$ matrix and consists of the trajectory coefficients of the X, Y and Z components of the trajectory along its rows. Considering all the trajectories, we can write

$$\mathbf{S} = \mathbf{A}\mathbf{\Theta}, \tag{4}$$

where $\mathbf{A} = (\mathbf{A}_1, \cdots, \mathbf{A}_P)$ and $\mathbf{\Theta}$ is given by

$$\mathbf{\Theta} = \begin{pmatrix} \mathbf{\Theta}_1 & & \\ & \ddots & \\ & & \mathbf{\Theta}_P \end{pmatrix}.$$

Combining equations 2 and 4, we can write

$$\mathbf{W} = \mathbf{R}\mathbf{A}\mathbf{\Theta}.$$

Since $\mathbf{\Theta}\mathbf{\Theta}^T$ is an identity matrix, therefore we can write

$$\begin{aligned} \mathbf{W}\mathbf{\Theta}^T &= \mathbf{R}\mathbf{A}. \\ \mathbf{D} &= \mathbf{R}\mathbf{A}, \end{aligned} \tag{5}$$

where we denote $\mathbf{D} = \mathbf{W}\mathbf{\Theta}^T$. We notice that $\mathbf{D}$ consists of the coefficients of 2D trajectories of $\mathbf{W}$, just like $\mathbf{A}$ consists of the coefficients of 3D trajectories. This means that the image observation and the 3D structure share the same trajectory space. A key theoretical insight of Equation 5 is that the 2D trajectory coefficients are a linear combination of the 3D trajectory coefficients and leads to all the interesting benefits of our approach. Equation 5 is analogous to Equation 2 in ceoefficient domain and can be seen as structure from motion in transformed domain. Conventional Tomasi-Kanade factorization [18] can be used to recover the 3D coefficients from the 2D coefficients. This factorization can be done using linear least square optimization, hence provides a robust estimation of 3D structure. Once $\mathbf{A}$ is known, the structure can be recovered using Equation 4.

Since both 2D and 3D trajectories share the same low-dimensional trajectory space, the trajectory basis can be learnt directly from the image observations. Next, we explore the link between trajectory space of $\mathbf{W}$ and $\mathbf{S}$ in more detail. We also show how trajectory smoothness assumption may be used to impute missing data.

4. Computing 2D trajectory coefficients

In the previous section, we formulated the trajectory space parametrization for the estimation of nonrigid structure using multiple static cameras. Now we will show how data-independent and data-dependent bases may be used for structure estimation and imputation of missing data, followed by a detailed comparison between the two approaches.

4.1. Using data-independent trajectory basis

The 2D trajectory coefficients $\mathbf{D}$ may be computed from image observation matrix $\mathbf{W}$ using $\mathbf{D} = \mathbf{W}\mathbf{\Theta}^T$ if the trajectory basis $\mathbf{\Theta}$ is known (Eq 5). Akhter *et al*. proposed using DCT basis and showed that DCT basis can compactly represent smooth trajectories and approach to the optimal linear 3D basis for a large training data [3]. If we use DCT trajectory basis then $\mathbf{D}$ may be directly computed if all the entries in $\mathbf{W}$ are known. When $\mathbf{W}$ is incomplete (due to occlusions) trajectory coefficients can still be estimated using a least squares fit as described below.

Let us denote $\mathbf{d}_{ij}$ as the 2D trajectory coefficients of trajectory $\mathbf{t}_{ij}$ i.e. $\mathbf{t}_{ij} = \mathbf{d}_{ij}\mathbf{\Theta}_j$. The estimation of $\mathbf{d}_{ij}$ is over-constrained and can be done even if there are missing points in $\mathbf{t}_{ij}$ (as long as the number of points available in the trajectory is at least K_j). Let $\hat{\mathbf{t}}_{ij}$ be the matrix, obtained from $\mathbf{t}_{ij}$, by deleting the columns corresponding to the missing points. Similarly, $\hat{\mathbf{\Theta}}_j$ denotes the matrix obtained from $\mathbf{\Theta}_j$ by deleting the columns corresponding to the missing points in $\mathbf{t}_{ij}$. $\mathbf{d}_{ij}$ can be estimated as

$$\mathbf{d}_{ij} = \hat{\mathbf{t}}_{ij}\hat{\mathbf{\Theta}}_j^+, \tag{6}$$

where $\hat{\mathbf{\Theta}}_j^+$ denotes the pseudo-inverse of $\hat{\mathbf{\Theta}}_j$. Hence missing data is imputed directly from the image observations. Then $\mathbf{d}_{ij}$s are stacked to form $\mathbf{D}$ and factorized to compute 3D trajectory coefficients $\mathbf{A}$. Finally $\mathbf{S}$ is estimated using Equation 4.

4.2. Learning data-dependent trajectory basis

DCT basis are simpler to use but data-dependent basis represent smooth trajectories more compactly. In non-rigid structure from motion, computing the optimal trajectory basis while simultaneously estimating the coefficients and cameras, results in a difficult trilinear factorization problem. Since our proposed factorization works in the transformed domain, we do not face this problem. We compute the trajectory basis $\mathbf{\Theta}$ directly from the image observations and the factorization is done afterwards. Here, all the trajectories are assumed to be of the same length $F_j = F$ and lie in the same $K_j = K$ dimensional subspace.

The image of the jth trajectory in ith camera is given by

$$\mathbf{t}_{ij} = \mathbf{R}_i \mathbf{T}_j. \tag{7}$$

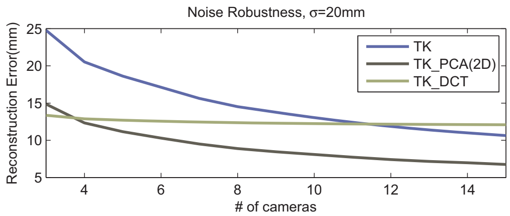


Figure 1. Robustness of trajectory parametrization in structure estimation: The experiment evaluates the effect on accuracy of reconstruction by increasing the number of cameras when using different types of basis. Here comparison is done on accuracy yielded by PCA(2D) , DCT and simple Tomasi Kanade when noise with a normal distribution having standard deviation of 20mm is added to a synthetic Dataset.

Equation 7 shows that the x and y components of $\mathbf{t}_{ij}$ are the linear combination of the X, Y and Z components of the 3D trajectory. Hence if the 3D trajectory lies in a low dimensional space, so does its 2D projection. Optimal 3D trajectory basis can be learned through Principal Component Analysis (PCA), if structure $\mathbf{S}$ is known. Singular Value Decomposition (SVD) of the following matrix gives the PCA trajectory basis

$$\mathcal{S}_{F\times 3P} = \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1P} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{X}_{F1} & \mathbf{X}_{F2} & \cdots & \mathbf{X}_{FP} \end{pmatrix}. \tag{8}$$

Having rearranged the structure into matrix $\mathcal{S}$, we denote the image observation, viewed by the ith orthographic camera as a $F \times 2P$ matrix, $\mathcal{W}_i$. The relation between $\mathcal{S}$ and $\mathcal{W}_i$ is as following

$$\mathcal{W}_i = \mathcal{S}\mathcal{R}_i, \tag{9}$$

where $\mathcal{R}_i = \mathbf{I} \otimes \mathbf{R}_i^T$ and $\mathbf{I}$ is a $P \times P$ identity matrix. Hence $\mathcal{R}_i$ contains the truncated rotation matrix, $\mathbf{R}_i^T$ along its 3×2 diagonal blocks. In Equation 9, we adopt the same centering convention as Section 3. Since the column space of $\mathcal{W}_i$ is a linear combination of the column space of $\mathcal{S}$, therefore PCA basis learned over the horizontal concatenation of $\mathcal{W}_i$s can also serve as the trajectory basis for the structure $\mathcal{S}$. Hence optimal trajectory basis can be learned for a given K. Once trajectory basis are known, trajectory coefficients can be estimated using Equation 5. To the best of our knowledge, we are the first to demonstrate that optimal 3D trajectory basis can be directly learned from the image observations and use these basis for 3D reconstruction. Optimality of the basis improves the quality of 3D reconstruction. Moreover, decoupling the basis estimation from 3D reconstruction results in a numerically stable and efficient factorization algorithm.

Estimation of trajectory basis can also be done when some of image observations are missing. The assumption that $\mathcal{W}_i$ is low rank can be used to impute its missing values

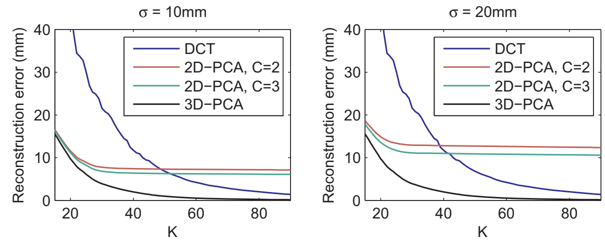


Figure 2. Comparison of PCA basis learned over a noisy $\mathcal{W}$ (2D-PCA) with DCT and PCA over ground truth structure $\mathcal{S}$ (3D-PCA). Comparison is done on a Mocap dance sequence with F=1000. The height of the actor is approximatley=1600mm. We generate synthetic image observation matrices $\mathcal{W}_i$ with C number of cameras, where $C = \{2, 3\}$. We add Gaussian noise in $\mathcal{W}_i$s with standard deviation = 10mm in left subplot and 20mm in right subplot. We learn PCA basis over noisy $\mathcal{W}_i$s (i.e. 2D-PCA) and compare their reconstructions with DCT and 3D-PCA. Plots show that for smaller values of K, 2D-PCA performs better, whereas for larger values of K, DCT is a good choice of basis.

using missing data factorization methods like [2, 8]. However, we adopt a very naive approach to estimate missing values in $\mathcal{W}_i$. In our synthetic experiments, we learn PCA basis over the 2D trajectories which were completely visible in $\mathcal{W}_i$s. For real data, we first impute the missing values using DCT-based interpolation, then learn the trajectory basis using PCA. Once $\mathcal{W}_i$ is complete, trajectory basis can be learned using PCA. In the following section, we present a comparison of data-dependent and data-independent bases.

4.3. Data-dependent vs. data-independent bases

We discussed in previous sections that the proposed factorization approach and the compact trajectory parametrization explicitly handle missing observations and noise in the data. In this section, we compare the reconstruction accuracy of proposed approach using both data-dependent (PCA) and data-independent bases (DCT) under various levels of noise and missing observations. This comparison was done on a Motion Capture dance sequence consisting of 1000 frames and 41 points. We generated synthetic images and tested proposed factorization approach given in Equation 5. As a matter of notation, we call proposed structure from motion method SFM_DCT, when DCT basis is used and SFM_PCA when PCA basis is used. Our experiments show that in an ideal case i.e. when $\mathcal{S}$ lies in a K dimensional subspace, SFM_PCA gives numerically zero reconstruction error. In Figure 1, we compare SFM_DCT, SFM_PCA and factorization method of Tomasi and Kanade [18] in the presence of noise. Plots show that more cameras are needed in Tomasi and Kanade factorization to get the same reconstruction accuracy, as compared to the proposed method. Specifically 14 cameras are needed in Tomasi and Kanade method to get the same accuracy as that of proposed method with just four cameras. Plots also show that the PCA basis performs significantly better as compared to the DCT basis, as number of cameras increases. Next we evaluate proposed factorization approach under different values of K in Figure 2. We tried proposed factorization approach using trajectory basis learned over ground truth 3D structure (3D-PCA), trajectory basis learned from image observations (2D-PCA) and DCT. Plots show that under small standard deviation of noise and for small values of K, 2D-PCA out-performs DCT. Plots also show that accuracy of 2D-PCA increases with the increase of cameras.

To test the accuracy of structure estimation under missing observations, we create synthetic occlusion scenarios in the dance sequence at random locations and reconstruct the structure. We choose 20 (about half of the points) random trajectories and delete a set of contiguous points of a certain length. We test the proposed method of structure estimation using Equation 5 on both DCT and PCA basis. PCA basis were learned over image observations. In Figure 3 we plot reconstruction error of SFM_DCT and SFM_PCA by varying the gap length for different values of K. We create three synthetic cameras and project the 3D structure to get $\mathcal{W}_i$s. In Figure 3(a) trajectories were invisible in all three cameras for a particular duration of time, whereas in Figure 3(b), trajectories were visible in one camera only. Plots show that SFM_PCA reconstructs 3D trajectories accurately, even when 2D trajectories were invisible in all three cameras for about 300 consecutive frames in the 1000 frame long sequence. Figure 3(b) shows that the imputation accuracy improves when complete trajectory is visible in one camera. Plots also show that SFM_DCT is relatively unstable in finding imputation of the missing data, when gap length is large. Though PCA based imputation is more stable than the DCT, its accuracy depends upon the learning of the PCA basis from the incomplete image observation matrix. Incomplete matrix factorization can be used to estimate PCA basis. However, this itself is an open research area. According to our knowledge, global solution of incomplete matrix factorization is still intractable.

5. Results

We did an extensive quantitative and qualitative evaluation of the proposed method. For quantitative evaluation, we chose 20 different Motion Captured sequences, like running, jumping, boxing, crawling and dancing. We generated synthetic perspective cameras and created images. Synthetic images were used to form image observation matrix $\mathbf{W}$. We run SFM_DCT and SFM_PCA on $\mathbf{W}$ and computed the per point reconstruction error. In Figure 4 we plot the average reconstruction error of 20 sequences by varying the distance of the camera from the structure. We did this experiment with 3, 5 and 7 cameras. We see that the reconstruction error decreases with the increase of distance

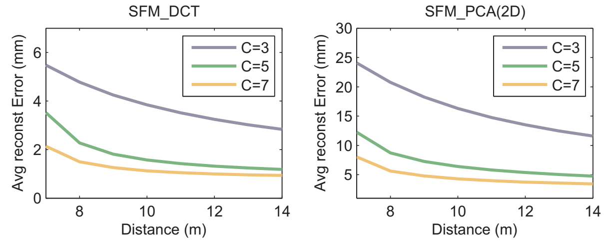
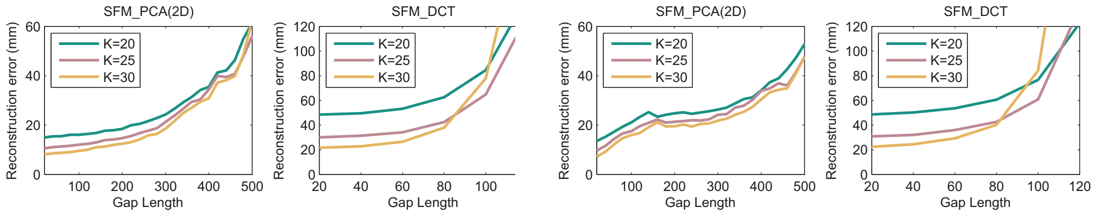


Figure 4. Effect of distance on reconstruction error with a Perspective Camera Model: The experiment evaluates the performance of DCT and PCA(2D) basis in reducing noise introduced due to perspective camera model. Here a circular camera setup is used and the experiment is repeated with varying number of cameras, in each sucessive experiment increasing radial distance.

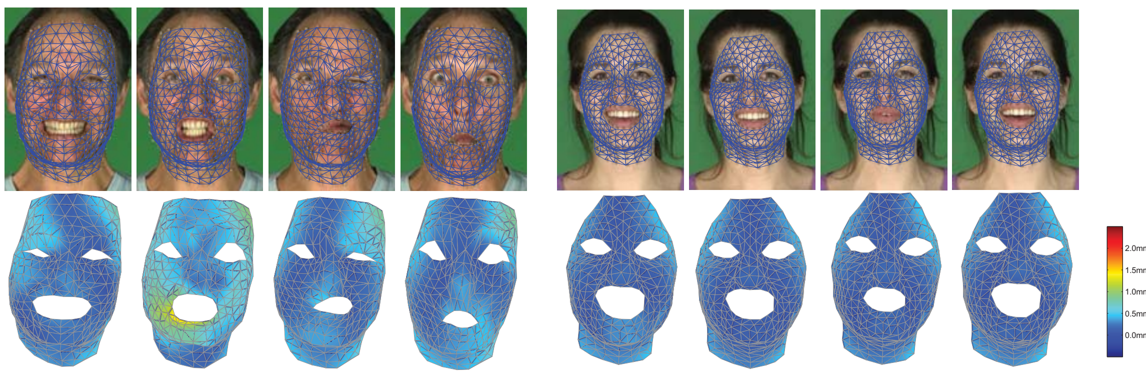
and increase of cameras. This experiment suggests the validity of the method on a variety of actions. We also tested our method on two Motion Captured sequences of face and two dense full body sequences from Performance Capture database [7]. For face datasets, we created three synthetic perspective cameras at a distance of 3m from the object, with one camera fronto parallel to the object and the other two with an angle of $\pm 45 \deg$ with the center one. For full body sequences, we generated six perspective cameras. We used actual camera calibration matrices to get synthetic images. In Figure 5 we plot the 3D reconstruction on face and body sequences on a few frames. The colorbar shows the error in reconstruction in mm. Figure 5(a) and 5(b) show that the maximum error on face sequences is 1.5mm. Figure 5(c) and 5(d) demonstrate qualitatively appealing results on full body sequences. This experiment demonstrates the performance of the algorithm, under a breakdown of orthographic camera assumption i.e. reconstruction looks good qualitatively though reconstruction error is large in comparison with the ground truth.

We also tested proposed method on two real sequences named dance and jumping-dog from 4D repository [1] as shown in Figure 6. We did manual tracking in both of these sequences. We tracked 200 frames in 4 cameras in dance sequence, whereas in jumping-dog sequence, we tracked 3 cameras in 150 frames. These sequences provide challenging test cases for the proposed method because of the extreme tracking noise and difficult occlusion scenario. We observe that out of 34 points in jumping-dog sequence one points was invisible in all three cameras for eight consecutive frames. About 9 points were invisible in two cameras for 15 or more consecutive frames, one point was even invisible in all 34 consecutive frames. Figure 6 shows the reconstruction using SFM_PCA. Results demonstrate the robustness of the proposed method under tracking noise and occlusion of the points.



(a) Trajectories invisible in all three cameras (b) Trajectories visible in one camera only

Figure 3. Handling Missing data in image observation matrix: we create synthetic occlusions in cameras for a certain number of frames and reconstruct the trajectory with SFM_PCA and SFM_DCT for different values of K. Plots show that the occluded points can be reliably reconstructed. SFM_PCA demonstrate better reconstruction accuracy for larger gaps than SFM_DCT.



(a) (b)

(c) (d)

Figure 5. Perspective camera reconstruction results on Mocap face sequences. Sub-figures (a) and (b) show our results projected on images (first row) TK_DCT reconstruction (second row). Sub-figures (c) and (d) show perspective camera reconstruction results on Performance Capture data [7]. Colors indicate error in reconstruction with respect to the ground truth.

Acknowledgments

We thank Arif Zaman and Murtaza Taj for valuable feedback. We also thank Muhammad Ahmad Bashir, Hassan Bukhari, Arslan Akram Sandhu, Mehr-un-Nisa Shahid, Ali Rehan, Maria Zubair, Kashif Murtaza and Abdul Rehman Naeem for help with tracking of some input data.

References

[1] http://4drepository.inrialpes.fr/.

[2] M. Q. Aguiar and M. Stosic. Spectrally optimal factorization of incomplete matrices. *CVPR*, 2008.

[3] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from motion. In *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2011.

[4] A. Baak, B. Rosenhahn, M. Muller, and H.-P. Seidel. Stabilizing motion tracking using retrieved motion priors. In *Computer Vision, 2009 IEEE 12th International Conference on*, 29 2009.

[5] C. Bregler, A. Hertzmann, and H. Biermann. Recovering

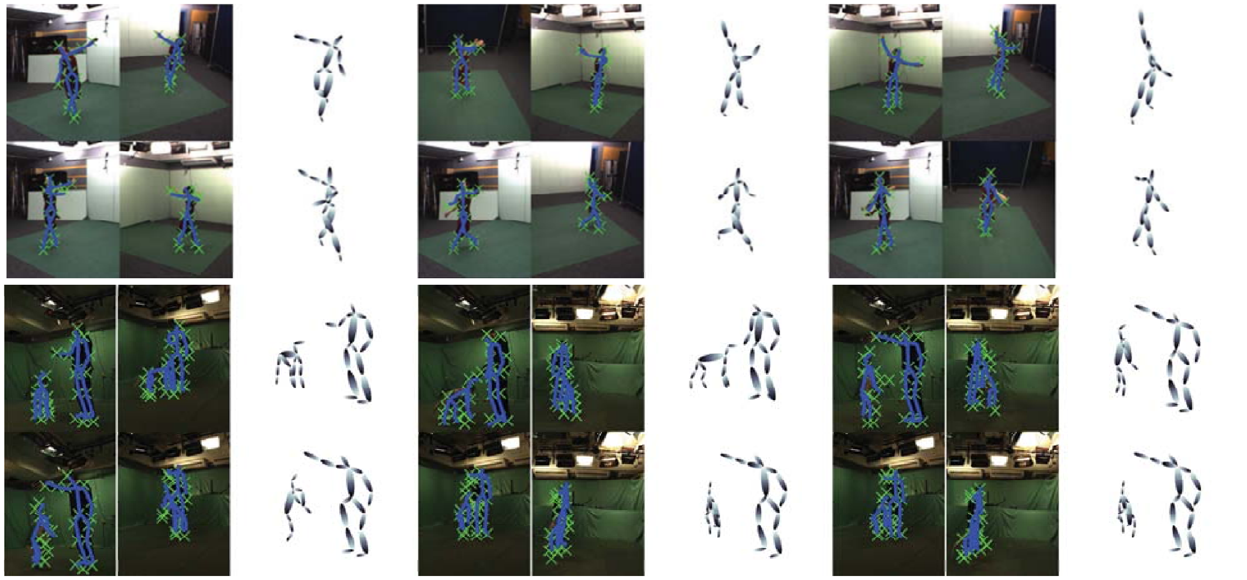


Figure 6. Results on real datasets: We took two datasets from 4D repository [1], named dance and jumping-dog. We manually track the frames and reconstruct structure using SFM_PCA. We tracked 4 cameras in dance sequence and 3 in jumping-dog. Despite the noise in manual tracking and occlusion of points, proposed method qualitatively accurate results.

non-rigid 3D shape from image streams. *CVPR*, 2:690–696, 2000.

[6] J. Davis, D. Nehab, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(2):296–302, Feb. 2005.

[7] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun. Performance capture from sparse multi-view video. *ACM Trans. Graph.*, 27(3), 2008.

[8] A. Eriksson and A. van den Hengel. Efficient computation of robust low-rank matrix approximations in the presence of missing data using the l_1 norm. *CVPR*, 2010.

[9] J. Fayad, A. Del Bue, L. Agapito, and P. Aguiar. Non-rigid structure from motion using quadratic deformation models. In *British Machine Vision Conference (BMVC2009), London, UK*, 2009.

[10] M. Gong. Enforcing temporal consistency in real-time stereo estimation. In *Computer Vision ECCV 2006*, volume 3953 of *Lecture Notes in Computer Science*, pages 564–577. Springer Berlin / Heidelberg, 2006.

[11] C. Lei, X. D. Chen, and Y.-H. Yang. A new multi-view spacetime-consistent depth recovery framework for free viewpoint video rendering. In *ICCV*, pages 1570–1577, 2009.

[12] T. B. Moeslund, A. Hilton, and V. Krger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3):90 – 126, 2006.

[13] H. Park, T. Shiratori, I. Matthews, and Y. Sheikh. 3D reconstructiopn of a moving point from a series of 2D projections. *ECCV*, 2010.

[14] G. Pons-Moll, A. Baak, T. Helten, M. Mu andller, H.-P. Seidel, and B. Rosenhahn. Multisensor-fusion for 3d full-body human motion capture. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 663 –670, 2010.

[15] M. Salzmann, R. Hartley, and P. Fua. Convex optimization for deformable surface 3-d tracking. In *IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007.

[16] M. Sizintsev and R. P. Wildes. Spatiotemporal stereo via spatiotemporal quadric element (stequel) matching. In *Computer Vision and Pattern Recognition*, 2009.

[17] C. Taylor. Reconstruction of articulated objects from point correspondences in a single uncalibrated image. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 677 –684 vol.1, 2000.

[18] C. Tomasi and T. Kanade. Shape and Motion from Image Streams Under Orthography: A Factorization Method. *IJCV*, 9(2), 1992.

[19] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from motion: Estimating shape and motion with hierarchical priors. *PAMI*, 30(5):878–892, May 2008.

[20] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment a modern synthesis. In *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 153–177. Springer Berlin / Heidelberg, 2000.

[21] L. Zhang, N. Snavely, B. Curless, and S. M. Seitz. Spacetime faces: High-resolution capture for modeling and animation. In *ACM Annual Conference on Computer Graphics*, pages 548–558, August 2004.