3D Deformable Shape Reconstruction with Diffusion Maps

Lili Tao
lltao@uclan.ac.uk
Bogdan J. Matuszewski
bmatuszewski1@uclan.ac.uk

Applied Digital Signal and Image Processing Research Centre
University of Central Lancashire, UK

Abstract

This paper presents a method for recovering deformable shape and motion from uncalibrated 2D video sequence in the presence of missing data. Highly deformable shapes are hard to describe under previously used assumptions, such as global constraint enforcing shapes to lie within a linear subspace. Considering that the data dimensionality may not represent the true complexity of the problem, we suggest that the shapes can be well-modelled in a low dimensional manifold. However, building a dense representation of the manifold requires a large amount of training data which is not feasible in many real applications. The main contribution of this paper is to propose a novel approach for estimating accurate 3D reconstructions utilising manifold learned from a relatively small number of training samples. The problem is addressed by grouping shapes into evolving clusters, with the shapes in each cluster represented in the linear subspace, estimated based on the observations and the prior learned manifold. Results are presented using motion capture data and real video sequences, showing that the proposed method can better model shapes with complex deformations compare to several state-of-the-art techniques, and is robust against noise and missing data.

1 Introduction

Structure from motion (SfM) is defined as a problem of modelling 3D objects and estimating corresponding camera motion trajectories based on a set of observed images. While reconstructing 3D geometry has been well-studied under the assumption of object rigidity [11], in many real applications, such as the human face or body, objects often deform over time.

To extend the rigid SfM to the case of 3D deformable objects, a low rank shape model has been widely used in the non-rigid and articulated object reconstructions [15, 19]. In addition, since the high number of degrees of freedom and motion degeneracy may lead to the methods failing to obtain meaningful reconstructions, prior information can be used to improve the quality of recovered shapes and motion [8]. Another class of algorithms, so called trajectory approaches were inspired by the shape basis model but using predefined basis trajectories instead of estimated basis shapes, thus removing a large number of unknowns from the optimisation [1, 9, 10]. Template based reconstruction usually relies on a known reference frame and works well especially for reconstruction of inextensible surfaces, but a common drawback of this approach is that the initialisation has to be close to the solution [5].

The problem becomes more difficult when the observations are incomplete. The methods addressing this problem can be divided into three categories: imputation, alternation and non-linear optimisation. Imputation algorithms attempt to fill in the missing data entries using complete subset of the data [18, 22]. These methods are simple but cannot handle real data, which often tend to be very noisy. Alternation algorithms solve the problem based on closed-form solution using a rank constraint imposed on the measurement matrix without estimating the missing values in advance [13, 15]. Most existing methods for this problem followed this idea by iteratively updating motion and shape in terms of observed measurements [10]. Note that optimising the complete matrix using only rank constraint is often not sufficient, but for these methods it is difficult to incorporate additional constraints [9]. Therefore a careful initialisation is needed, otherwise the results can easily drift into a local minima. Non-linear optimisation is a direct solution for shape and motion recovery when measurement data are missing. Even though the inherently high number of degrees of freedom may lead to failure of obtaining reliable 3D reconstructions, additional constraints can naturally be included in the cost function.

Novelty. The main contribution of this paper is a novel approach for recovery of 3D non-rigid structures with large and/or complex deformations. The proposed method is shown to be flexible allowing a method extension to handle the case with missing measurements e.g. due to occlusion or feature track loss. The proposed method is based on a recently introduced manifold learning technique, Diffusion maps [6]. As claimed in [16], building a dense representation of the manifold enables to achieve better reconstruction performance when compared to other state-of-the-art approaches, but collecting sufficient number of training data may not be feasible in practice. The algorithm described in this paper is an improved version of the algorithm proposed in [16], with three main differences. First, the improved algorithm enables reconstruction with small number of training samples. Second, the proposed cost function includes additional term to relax the constraint on local basis shapes. Unlike in [16] these shapes do not have to match the local training samples. Third, the proposed algorithm has additional step solving the missing data problem. Despite the fact that manifold learning techniques are becoming increasingly popular in many different areas, such diffusion maps based approach has rarely been applied in the context of motion and non-rigid shape reconstruction, especially with missing data.

2 Problem statement

2.1 Formulation

In the case of non-rigid objects, the 3D shapes deforms throughout the time which makes the problem more difficult to solve. Assuming that a set of image feature points have been tracked in the 2D image sequence viewed by an orthographic camera, the problem consists of shapes $\mathcal{S} = \{\mathbf{S}_1, \dots \mathbf{S}_f\}$ and camera rotation $\mathcal{R} = \{\mathbf{R}_1, \dots, \mathbf{R}_f\}$, recovery from 2D observations $\mathcal{Y} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_f\}$, thus can be formulated as the following optimisation problem,

$$\underset{\mathbf{R}_t, \mathbf{S}_t}{\arg\min} \sum_{t=1}^{f} \|\mathbf{Y}_t - \mathbf{P} \cdot \mathbf{R}_t \cdot \mathbf{S}_t\|^2 \tag{1}$$

where $\mathbf{P}$ represents a known orthographic camera projection matrix, $\mathbf{Y}_t$ represents the 3D points projected onto t^{th} image. The camera translation can be eliminated, by expressing 2D

observations with respect to the data points centroid calculated for each observed image.

According to the shape basis assumption, shape $\mathbf{S}_t$ can be represented as a linear combination of n unknown but fixed basis shapes $\mathbf{B}_l$, $\mathbf{S}_t = \sum_{l=1}^{n} \theta_{tl}\mathbf{B}_l$, while the shape coefficients θ_{tl} are adjustable over time. Therefore the measurement can be rearranged as:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_f \end{bmatrix} = \begin{bmatrix} \theta_{11}\mathbf{P}\cdot\mathbf{R}_1 & \cdots & \theta_{1n}\mathbf{P}\cdot\mathbf{R}_1 \\ \vdots & \ddots & \vdots \\ \theta_{f1}\mathbf{P}\cdot\mathbf{R}_F & \cdots & \theta_{fn}\mathbf{P}\cdot\mathbf{R}_F \end{bmatrix} \begin{bmatrix} -\mathbf{B}_1- \\ \vdots \\ -\mathbf{B}_n- \end{bmatrix} = \mathbf{MB} \tag{2}$$

2.2 The diffusion framework

Diffusion maps have become a popular method in data dimensionality reduction given their capability to recover underlying structures of a complex manifold, robustness to noise, data outliers, and efficient implementation.

Let $\mathcal{X} = \{\mathbf{X}_1 \dots \mathbf{X}_M\}$ be a dataset with M training samples lying on an n dimensional manifold $\mathcal{M}$ embedded in a higher-dimensional space $\mathbb{R}^N$. The idea of dimensionality reduction is to learn a low dimensional representation $\{\mathbf{x}_1 \dots \mathbf{x}_M\}$ with $\mathbf{x}_i \in \mathbb{R}^n, n \ll N$ preserving implicit structure of the data. A mapping is defined by $\Psi : \mathcal{X} \mapsto \Psi(\mathcal{X})$, with the optimal embedding provided by eigenvalues λ_l and associated eigenvectors φ_l of the Laplace-Beltrami operator [7], such as,

$$\Psi(\mathbf{X}_i) \mapsto [\lambda_1\varphi_1(\mathbf{X}_i), \cdots, \lambda_n\varphi_n(\mathbf{X}_i)]^T \tag{3}$$

The details of building approximated Laplace-Beltrami operator can be found in [16].

3 Deformable shape reconstruction

The method presented in [16] introduced the non-linear manifold, learned based on 3D training samples, as shape prior for non-rigid shape reconstruction. Given the learned shape manifold and the observed 2D measurements, the algorithm iteratively refines the 3D reconstructed shapes for each frame by using its $n+1$ nearest shape neighbours on the manifold, as basis shapes. Although the method is able to achieve high quality shape reconstructions, the requirement of large number of training data to build a sufficiently dense representation of the manifold is not feasible for most real applications. To overcome this, the method proposed in this paper relaxes the constraint for basis shapes so as to make the algorithm more adaptable to the case when only a relatively small number of training samples have been used for the manifold learning.

3.1 Mapping and inverse mapping for previously unseen data

The diffusion maps can only provide embedding for the given training data without a clear strategy for embedding shapes which are not presented in the training set. Re-training of the whole manifold is impractical due to the computation cost. The out-of-sample issue was first demonstrated in [3] and applies to several spectral algorithms for manifold learning. The method proposed in [7] projects the previously unseen data $\mathbf{S}_t \in \mathbb{R}^N$ onto the lower dimensional feature space, such as $\mathbf{S}_t \mapsto \left(\hat{\Psi}_1(\mathbf{S}_t) \cdots \hat{\Psi}_n(\mathbf{S}_t)\right)$. For each new shape, an embedding $\hat{\Psi}$ is calculated by an approximation technique based on the Nyström extension.

Initial shapes and camera motion are estimated by running a few iteration of the optimisation process using a linear method [17].

Once the initial shapes have been embedded into a lower dimensional space, finding their inverse mapping (the pre-image problem) can help to update shapes. However, the exact pre-image typically does not exist, and the problem can only be defined as an approximate solution [12]. Suppose we have an embedded point $\mathbf{b}_t \in \mathbb{R}^n$, a Delaunay triangulation [4] can be computed in n dimensional reduced space, enabling selection of $n+1$ nearest neighbours $\mathbf{x}_{tl}$ of $\mathbf{b}_t$. Each point $\mathbf{b}_t$ can be represented as $\mathbf{b}_t = \sum_{l=1}^{n+1} \theta_{tl}\mathbf{x}_{tl}$, where the coefficients $\theta = \{\theta_{t1}, \ldots, \theta_{t(n+1)}\}$ are the barycentric coordinates of $\mathbf{b}_t$ and the inverse mapping can be formulated as,

$$\hat{\Psi}^{-1}(\mathbf{b}_t) = \sum_{l=1}^{n+1} \theta_{tl}\mathbf{X}_{tl} \quad with \quad \sum_{l=1}^{n+1} \theta_{tl} = 1, 0 \leq \theta_{tl} \leq 1 \tag{4}$$

where training sample $\mathbf{X}_{tl}$ is the pre-image of $\mathbf{x}_{tl}$.

3.2 Shape clustering

Given a set of estimated shapes $\mathcal{S} = \{\mathbf{S}_1, \ldots \mathbf{S}_f\}$, the aim of the clustering is to partition f shapes into K clusters, in which the shapes have similar structure, with each shape cluster denoted by $\mathcal{T}_i, i \in 1 \ldots K$. The clusters are obtained by performing the Delaunay triangulation in the reduced space. As defined in [4], any "angle-optimal" triangulation of a set of points is a Delaunay triangulation of these points. This can help to avoid "skinny triangles", for which the corresponding shape of each vertex could be significantly different, thus may lead to meaningless reconstructions.

Diffusion maps are based on distance preserving mapping, meaning that the points relatively close in reduced space correspond to the similar shapes. As a consequence we stipulate that the points in the reduced space belong to the same Delaunay simplex (i.e. cluster), can be modelled by the same linear subspace embedded in $\mathbb{R}^N$, and therefore all corresponding reconstructed shapes (represented by that cluster) can be approximated by a linear combination of the same set of unknown but fixed basis shapes. Thus all the shapes in the cluster i can be represented as $\mathbf{S}_t = \sum_{l=1}^{n+1} \theta_{tl}\mathbf{B}_l^i, \forall t \in \mathcal{T}_i$, where a set of basis shapes $\mathcal{B}^i = \{\mathbf{B}_1^i \ldots \mathbf{B}_{n+1}^i\}$ is spanning the tangent linear subspace representing all the shapes from the cluster i.

The reconstructed shapes are often different from the training samples, therefore cannot be perfectly mapped into the manifold $\mathcal{M}$. As the result we relax the constraint for the basis shapes, only "encouraging" them to be close to the basis shapes spanning the tangent subspace, instead of being exactly the same. The additional constraint applied to the i^{th} set of basis shapes is,

$$\varepsilon_{bs}^i = \sum_{l=1}^{n+1} \left\| \mathbf{B}_l^i - \mathbf{X}_l^i \right\|^2, \mathbf{X}_l^i \in \mathcal{X} \tag{5}$$

Figure 1 illustrates an example of how the initial shapes are redistributed in the reduced space after algorithm has converged. As shown in (a) the initial shapes are embedded in a two dimensional space which fall into three clusters, $K = 3$. (b) shows the embedding of optimal shapes which produced by the non-linear optimisation (see Section 3.3) with $K = 11$.

This approach differs from the one presented in [16] as all the shapes belonging to the same cluster are being jointly optimised, whereas in [16] all the shapes would have been reconstructed independently if not for the temporal smoothness constraint(not used in the algorithm proposed in this paper). Additionally the proposed algorithm relaxes the constraint on the tangent subspace as it only encourages that the basis shapes to be "close" to this subspace.

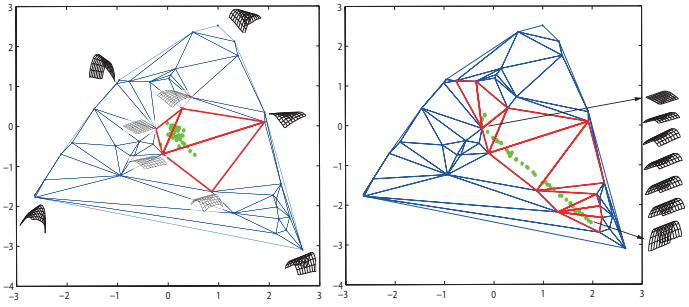


Figure 1: Delaunay triangulations (blue line) in the reduced space; Left: Embedded initial shapes (green dots) in a reduced space and the actual used triangles (red line), together with representative corresponding shapes from the total of 40 training samples; Right: Embedded reconstructed shapes (green dots) in a 2D reduced space and the actual used triangles (red line), with some reconstructed shapes

3.3 Non-linear refinement

The parameters $\theta_{tl}, \mathbf{B}_l^i$ and $\mathbf{R}_t$ are optimised simultaneously by minimising the 2D re-projection error with additional constraints on basis shapes and rotation matrices. The cost function can be written as,

$$E(\mathbf{R}_t, \mathbf{B}_l^i, \theta_{tl}) = \sum_{t \in \mathcal{T}_i} \left\| \mathbf{Y}_t - \mathbf{P} \cdot \mathbf{R}_t \sum_{l=1}^{n+1} \theta_{tl} \mathbf{B}_l^i \right\|^2 + \lambda_B \varepsilon_{bs}^i + \lambda_R \sum_{t \in \mathcal{T}_i} \varepsilon_{rot} \tag{6}$$

where $\varepsilon_{rot} = \left\| \mathbf{R}_t \mathbf{R}_t^T - \mathbb{I} \right\|$ enforces orthonomality of all $\mathbf{R}_t$. The parameters λ_B and λ_R are regularisation constants selected experimentally. A non-linear optimisation based on bundle adjustment using Levenberg-Marquardt algorithm was applied to minimize this cost function.

Because the quality of the provided initial shapes may seriously affect the results of the optimisation, we try to avoid this by updating the basis shapes $\mathcal{B}^i$(re-cluster the data) and the corresponding shape coefficients in each iteration until 2D measurement error is less than the defined threshold (10^{-3} in this case) and the error between two adjacent frames is relatively small. The pre-image of the vertices of Delaunay triangles are used to constraint the basis shapes, Figure 1 shows which Delaunay simplexes are being used along the iterations. The algorithm for iteratively 3D shape estimation is summarised in Algorithm 1.

4 Reconstruction with Missing Data

The algorithm described above assumes the measurements $\mathcal{Y}$ are complete, all the feature points are identified in all the images in the sequence. In practice, some of the points cannot be detected in all the images due to the occlusions, feature detection problems, or tracking failures and therefore acquiring complete set of measurements is unlikely. We present two methods which efficiently handle the case of missing data in the shape estimation problem.

4.1 Linear approach

If the input data is incomplete, instead of considering more complex and time-consuming optimisation algorithms, we briefly summarise a recently proposed linear method based on

Algorithm 1 Iteratively 3D shape estimation

Input: 2D points with known correspondence, diffusion map calculated from the training dataset $\mathcal{X}$.

1: Initialisation: Obtain initial shapes $\mathbf{S}'$ and camera motion $\mathbf{R}'$. for each frame t.
2: **repeat**
3: Compute the embedding $\hat{\Psi}$ of new shapes $\mathbf{S}_t \mapsto \hat{\Psi}(\mathbf{S}_t)$
4: Find $n+1$ nearest neighbours $\mathbf{x}_{tl}$ and its corresponding training samples $\mathbf{X}_{tl}$ of the embedded point $\mathbf{b}_t$
5: Calculate the barycentric coordinates θ_{tl} of $\mathbf{b}_t$
6: Perform clustering $\mathcal{T}_i$ of the estimated shapes $\mathcal{S}$
7: Refine $\theta_{tl}, \mathbf{B}_l^i, \mathbf{R}_t$ as to the cost function Eq. 6
8: Update the reconstructed 3D shapes $\mathbf{S}'_t = \sum_{l=1}^{n+1} \theta_{tl}\mathbf{B}_l^i$
9: Set $\mathbf{S}_t = \mathbf{S}'_t$
10: **until** $(\|r\| > r_T)$ *and* $\left(\|r_t\| - \|r_{t-1}\| > 10^{-3}\right)$

Output: 3D reconstructed shapes $\mathcal{S}$ and camera motion $\mathcal{R}$.

Principal Component Analysis (PCA) [17], with the missing data recovered before estimating the shapes and motion.

Assuming p feature points lie on the surface of an object, we set $\mathbb{I} = \bar{\Pi}_t + \bar{\Pi}_t^*$, where $\mathbb{I}$ is the identity matrix and $\bar{\Pi}_t$ is a $p \times p$ diagonal matrix such that $\bar{\Pi}_t(k,k) = 0$ indicates that the point k is missing in image t, otherwise $\bar{\Pi}_t(k,k) = 1$. The observations of time t can be represented as $\hat{\mathbf{Y}}_t = \mathbf{Y}_t\Pi_t$ and the missing measurements as $\hat{\mathbf{Y}}_t^* = \mathbf{Y}_t\Pi_t^*$, where matrix Π_t and Π_t^* are obtained from $\bar{\Pi}_t$ and $\bar{\Pi}_t^*$ by removing all columns for which entries are all zeros. According to Eq. 2, measurements can be factorised using motion $\mathbf{M}$ and shape bases $\mathbf{B}$ matrices, the incomplete measurement can be written as: $\hat{\mathbf{Y}}_t = \mathbf{M}_t\mathbf{B}\Pi_t$.

We firstly compute the motion matrix $\mathbf{M}_t$ using the available 2D measurements and the eigenshapes $\mathbf{E}$, approximating the unknown bases $\mathbf{B}$, obtained from the training dataset $\mathcal{X}$, $\mathbf{M}_t = \hat{\mathbf{Y}}_t(\mathbf{E}\Pi_t)^\dagger$, where $(\cdot)^\dagger$ indicates Moore-Penrose pseudo-inverse. The missing entries can be calculated as $\hat{\mathbf{Y}}_t^* = \mathbf{M}_t\mathbf{E}\Pi_t^*$. Thus the completed measurement matrix is,

$$\mathbf{Y}_t = \hat{\mathbf{Y}}_t\Pi_t^T + \hat{\mathbf{Y}}_t^*\Pi_t^{*T} \tag{7}$$

4.2 Non-linear approach

Since PCA is a linear manifold, the linear method is only able to cope well with small deformations. Although the method is not suitable when the deformations are relatively large or complex, it still can be used for providing a good starting point for the optimisation using the non-linear approach. The diffusion maps based method can be easily extended to handle the case with missing data. To facilitate this, modification of the Eq. 6 is introduced where the cost function can be rewritten as $E(\mathbf{R}_t, \mathbf{B}_l^i, \theta_{tl}, \mathbf{Y}_t\Pi_t^*)$. And therefore depends explicitly on the missing observations $\mathbf{Y}_t\Pi_t^*$. As results the cost function in Eq. 6 is simultaneously minimised with respect to rotation, shape basis, shape coefficients and the missing observations. It should be pointed out that we only optimise the missing entries in the observation not the whole 2D measurements $\mathbf{Y}_t$.

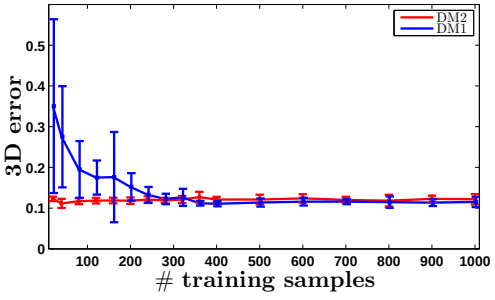


Figure 2: 3D error as function of the number of training samples for the *cardboard* data.

5 Experiments

We evaluate the performance of the proposed diffusion maps algorithm on both motion capture and real data. Several state-of-the-art algorithms are compared in our experiments: **CSF**: The column space fitting method [9]. **KSFM**: The kernel non-rigid structure from motion approach [10]. **IPCA**: The incremental principal components analysis based method [17]. **DM1**: The diffusion maps based method without basis shape optimisation, requiring large amount of training data [16] **DM2**: The proposed method.

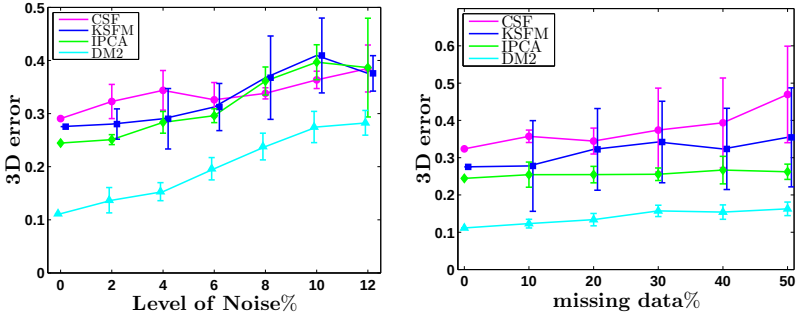
Data used for evaluation include: two articulated face sequences, *surprise* and *talking*, both captured using passive 3-D scanner with 3D tracking of 83 facial landmarks [14]; two surface models, *cardboard* and *cloth* [20]; and human action sequence *yoga* from CMU motion capture database [1]. Diffusion maps require training, for the face and surface sequences training datasets are taken from the BU-3DFE facial database [21] and from [20] respectively. Since no separate training data are provided for human action, then we use part of the frames for manifold learning and the rest for testing. All the training data has been rigidly co-registered. The same testing data has been applied for other methods which do not need training. We projected the 3D data using simulated orthographic camera.

In the following experiments, the reconstructed shapes are aligned using a single global rotation based on Procrustes alignment [1], and the errors are compared using normalised means of the 3D error [10] over all frames and all points.

5.1 Quantitative evaluation

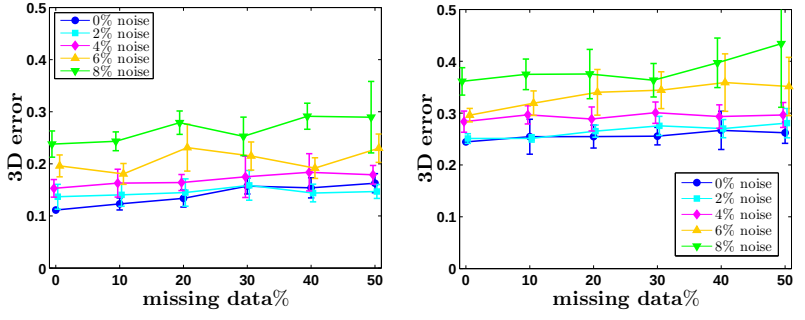
As it was stipulated in the previous sections, only a small number of training samples are required by the proposed method. We firstly investigate the effect of the number of training shapes on the reconstruction accuracy. The average reconstruction errors with the standard deviation calculated over 10 trials (each using different data subset for training) are shown in Figure 2. It can be seen that although the two methods are comparable when over 400 training samples are used, DM2 is more stable and outperforms DM1 when relatively small shape sample is used for training. For the comparative evaluation, performance of the proposed method is tested against three previous approaches. The experiment is design to test the robustness of our approach when data is corrupted by noise. The measurements $\mathcal{Y}$ were perturbed by Gaussian noise with varied level of noise. For each selected level of noise, the experiments were repeated 10 times. The results in Figure 3(a) show our method provides smaller reconstruction errors.

[1] The data was obtained from http://mocap.cs.cmu.edu.



(a) (b)

Figure 3: (a) Reconstruction error as function of the measurement noise for the *cardboard* data. (b) The influence of the observations missing data on the reconstruction error.



(a) (b)

Figure 4: Reconstruction results for varying levels of missing data and 5 levels of noise for the *cardboard* data. (a) Results using non-linear method with DM2; (b) Results using linear method.

To simulate the missing observations, we randomly discard 10%, 20%, 30%, 40% and 50% of the 2D entries in $\mathcal{Y}$. The results in Figure 3(b) are calculated by averaging over 10 trials. With the missing data ratio of up to 50% , the average (maximum) 3D and 2D reconstruction errors were 0.1629 (0.1881) and 0.0032 (0.0053) respectively, where errors were calculated as $\|\mathbf{Y}-\mathbf{Y}'\|/\|\mathbf{Y}\|$, where $\mathbf{Y}'$ is the reconstructed measurement matrix.

In real cases, missing data and measurement noise are distorting the observations in the same time. The aim of the following experiment is to evaluate the methods' performance in such situations. We compare results of the 3D error obtained using the PCA based method to fill the missing entries in the measurement and then apply DM2, with the results obtained using the non-linear approach. Results plotted in Figure 4 show the reconstruction error as function of the amount of the missing data for different level of noise in the observations. As it can be seen that both methods are robust with respect to missing data, however, the non-linear method provides smaller errors both in terms of means and standard deviations.

5.2 Qualitative evaluation

Motion capture data: Table 1 shows the 3D reconstruction error for different methods on different sequences. For DM we present both initial error and final result produced by DM1 and DM2. The errors are chosen with the optimal number of basis n, with the optimal n selected based on running the trials with n varying from 2 to 10. As shown in the table, DM1

	CSF	KSFM	IPCA	DM		
				Initial	DM1	DM2
Surprise	0.0396(3)	0.0381(4)	0.0829	0.3154	0.0352(10)	**0.0208**(10)
Talking	0.0573(3)	0.0498(4)	0.0986	0.9657	0.0350(10)	**0.0280**(10)
Cardboard	0.3237(3)	0.2753(2)	0.2445	0.2674	**0.1064**(10)	0.1114(10)
Cloth	0.2609(6)	0.1806(2)	0.1909	0.2967	**0.0287**(7)	0.0556(5)
Yoga	0.1467(7)	0.1474(7)	0.2626	0.2628	**0.0768**(10)	0.1197(10)

Table 1: Normalised mean 3D error (number of bases n) of reconstruction results using different methods.

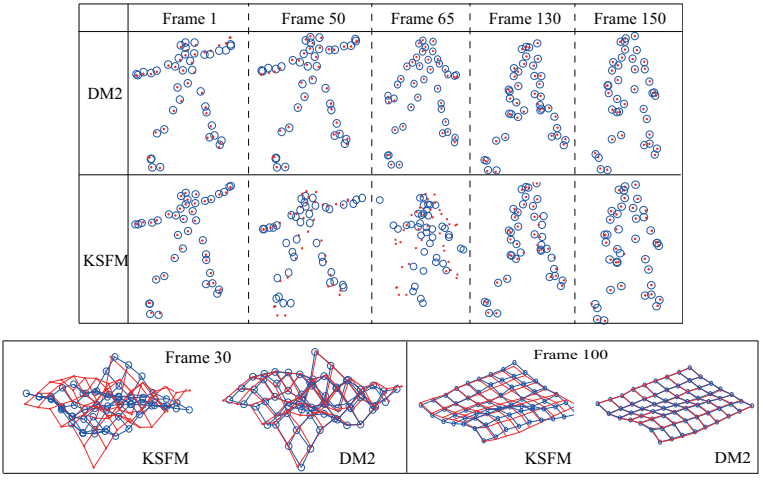


Figure 5: Reconstruction results on the *yoga* (upper) and *cloth* (bottom) sequences. Reconstructed 3D shapes (blue circles), together with ground truth (red dots) are displayed.

and DM2 consistently outperform other methods, especially for the sequences with large deformations. Even though the initial error is big, the proposed method is still able to provide accurate reconstruction results. DM1 and DM2 are comparable, but DM2 uses much less training data than DM1, e.g. for cardboard sequence, DM1 required a dense representation of the manifold, for which 1000 shapes have been used for training, while DM2 only used 40 shapes for training. More results comparing DM1 against other approaches can be found in [16].

In Figure 5, we visually compare the results of KSFM and DM2 against ground truth shapes. We can observe that DM2 generally gives better results, especially for the cloth sequence. This was to be expected since shapes can be better modelled in a non-linear manifold.

Real data: The algorithms used in the motion capture experiments above were applied to real data as shown in Figure 6. In the video, 81 features were tracked along 61 frames showing approximately two periods of paper bending movement.

6 Conclusion

In this paper, a manifold based approach has been demonstrated to recover the shape and motion of non-rigid objects from monocular image sequences. The advantage of the proposed method is that the non-linear manifold is only learned from small number of samples and the reconstructed shapes are clustered into several local linear subspaces. By combin-

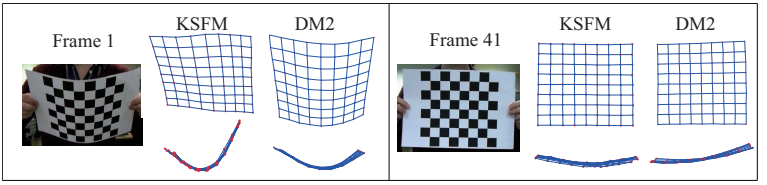


Figure 6: Selected 2D frames from the paper bending video sequence . Front and top views of the corresponding 3D reconstructed results using the proposed method (DM2) and KSFM.

ing non-linear manifold technique and low-rank shape model, the method produces accurate solutions to the shape recovery problem, and achieves better performance when compared with linear based methods, especially for the shapes with large and complex deformations. However the comparison of the proposed method with respect to the other methods may be seen as unfair, as better reconstruction accuracy of of the proposed method comes at the cost of required availability of a representative training dataset.

It should be noticed that selection of the training shapes has not been optimised leading to some badly shaped triangles in the clustered reduced space. The reconstruction results are affected if corresponding shapes are being clustered in such triangles. Future work will attempt to address the problem by either refining the Delaunay mesh or introducing a criterion for selection of the optimal training shapes. We are also investigating several extensions of this work to more challenging cases, such as to deal with the outliers and real time implementation.

References

[1] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1442–1456, 2011.

[2] P. Arias, G. Randall, and G. Sapiro. Connecting the out-of sample and pre-image problems in kernel methods. In *IEEE Conference on CVPR*, pages 1–8. IEEE, 2007.

[3] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. *Advances in Neural Information Processing Systems.*, 16, 2004.

[4] M. Berg, O. Cheong, M. Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag, 2008.

[5] F. Brunet, R. Hartley, A. Bartoli, N. Navab, and R. Malgouyres. Monocular template-based reconstruction of smooth and inextensible surfaces. In *ACCV*, pages 52–66. Springer, 2011.

[6] R. Coifman and S. Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[7] R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *National Academy of Sciences*, 102(21), 2005.

[8] A. Del Bue. A factorization approach to structure from motion with shape priors. In *IEEE Conference on CVPR*, pages 1–8. IEEE, 2008.

[9] P. F. U. Gotardo and A.M. Martinez. Computing smooth time-trajectories for camera and deformable shape in structure from motion with occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):878–892, 2008.

[10] P. F. U. Gotardo and A.M. Martinez. Kernel non-rigid structure from motion. In *IEEE Conference on ICCV*, pages 802–809. IEEE, 2011.

[11] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*, volume 2. Cambridge Univ Press, 2000.

[12] J. Kwok and I. Tsang. The pre-image problem in kernel methods. *IEEE Transactions on Neural Networks*, 15(6):1517–1525, 2004.

[13] M. Marques and J. Costeira. Estimating 3d shape from degenerate sequences with missing data. *Computer Vision and Image Understanding*, 113(2):261–272, 2009.

[14] B.J. Matuszewski, W. Quan, L-K. Shark, A. McLoughlin, C. Lightbody, H. Emsley, and C Watkins. Hi4d–adsip 3d dynamic facial articulation database. *Image and Vision Computing*, 10, 2012.

[15] M. Paladini, A. Bue, J. Xavier, M. Stosic, M. Dodig, and L. Agapito. Factorization for non-rigid and articulated structure using metric projections. In *IEEE Conference on CVPR*, pages 2898–2905. IEEE, 2009.

[16] L. Tao and B.J. Matuszewski. Non-rigid structure from motion with diffusion maps prior. In *IEEE Conference on CVPR*, pages 1530–1537, 2013.

[17] L. Tao, S.J. Mein, W. Quan, and B.J. Matuszewski. Recursive non-rigid structure from motion with online learned shape prior. *Computer Vision and Image Understanding*, 2013. doi: http://dx.doi.org/10.1016/j.cviu.2013.03.005.

[18] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[19] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5), 2008.

[20] A. Varol, M. Salzmann, P. Fua, and R. Urtasun. A constrained latent variable model. In *IEEE Conference on CVPR*, pages 2248–2255. IEEE, 2012.

[21] L. Yin, X. Wei, Y. Sun, J. Wang, and M.J. Rosato. A 3d face expression database for facial behavior research. In *Automatic face and gesture recognition*, pages 211–216. IEEE, 2006.

[22] A. Zaheer, I. Akhter, M.H. Baig, S. Marzban, and S. Khan. Multiview structure from motion in trajectory space. In *IEEE Conference on ICCV*, pages 2447–2453. IEEE, 2011.