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

Motivation This paper presents a method for recovering deformable shape and motion from uncalibrated 2D video sequence in the presence of missing data. 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 [3]. The main contribution of this paper is to propose a novel approach for estimating accurate 3D reconstructions using manifold learning technique, namely Diffusion maps, 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.

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 $\mathcal{Y} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_f\}$. 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 arranged as:

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

$\mathbf{P}$ represents a known orthographic camera projection matrix.

The initial shapes and motion are calculated based on linear approach.

Shape embedding Having a shape $\mathbf{S}_t$ not present in the training set $\mathbf{X}$, an embedding $\mathbf{S}_t \mapsto \left(\hat{\Psi}_1(\mathbf{S}_t), \cdots, \hat{\Psi}_K(\mathbf{S}_t)\right)$ of this new shape is calculated from the Nyström extension [1]:

$$\hat{\Psi}_k(\mathbf{S}_t) = \sum\nolimits_{\mathrm{X}_j \in \mathbf{X}} p(\mathbf{S}_t, \mathrm{X}_j)\varphi_k(\mathrm{X}_j) \tag{2}$$

where $p(\mathbf{S}_t, \mathrm{X}_j)$ is from diffusion operator [2].

Shape update 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. Suppose we have an embedded point $\mathbf{b}_t \in \mathbb{R}^n$, a Delaunay triangulation 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 $\theta = \left\{\theta_{t1}, \dots, \theta_{t(n+1)}\right\}$ are the barycentric coordinates of $\mathbf{b}_t$ and the inverse mapping can be formulated as,

$$\hat{\Psi}^{-1}(\mathbf{b}_t) = \sum\nolimits_{l=1}^{n+1} \theta_{tl}\mathbf{X}_{tl} \quad with \quad \sum\nolimits_{l=1}^{n+1} \theta_l = 1, 0 \le \theta_l \le 1 \tag{3}$$

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

Shape clustering 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 $\mathcal{B}^i = \left\{\mathbf{B}_1^i \dots \mathbf{B}_{n+1}^i\right\}$ 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,

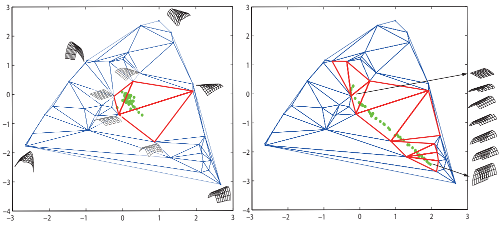

Figure 1: Delaunay triangulations (blue line) in the reduced space; Embedded initial shapes (left) and reconstructed shapes(right) (green dots) and the actual used triangles (red line)

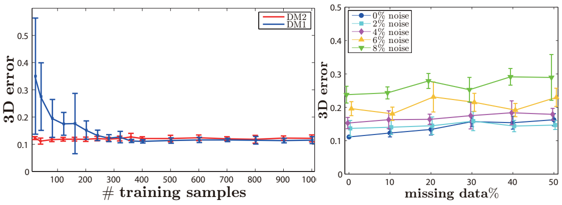

Figure 2: Results for the *cardboard* data. Left: 3D error as function of the number of training samples. Right: Varying levels of missing data and 5 levels of noise.

$$\varepsilon_{bs}^i = \sum_{l=1}^{n+1} \left\|\mathbf{B}_l^i - \mathbf{X}_l^i\right\|^2, \mathbf{X}_l^i \in \mathcal{X} \tag{4}$$

Fig. 1 illustrates an example of how the initial shapes are redistributed in the reduced space after algorithm has converged.

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,

$$E(\mathbf{R}_t, \mathbf{B}_l^i, \theta_{tl}) = \sum_{t \in \mathcal{T}_i} \left\|\mathbf{Y}_t - \mathbf{P}\cdot\mathbf{R}_t \sum_{l=1}^{n+1} \theta_{tl}\mathbf{B}_l^i\right\|^2 + \lambda_B \varepsilon_{bs}^i + \lambda_R \sum_{t \in \mathcal{T}_i} \varepsilon_{rot} \tag{5}$$

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

Evaluation Fig.2(Left) shows the effect of the number of training shapes on the reconstruction accuracy for [3] (DM1) and the proposed one (DM2). The average reconstruction errors with the standard deviation calculated over 10 trials, each using different data subset for training. In real cases, missing data and measurement noise are distorting the observations in the same time. Fig.2(Right) shows the relevant results.

Real data result shows in Fig.3.

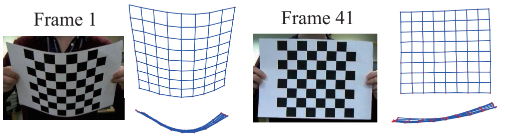

Figure 3: Selected 2D frames from paper bending video sequence. Front and top views of the corresponding 3D reconstructed results.

[1] P. Arias, G. Randall, and G. Sapiro. Connecting the out-of-sample and pre-image problems in kernel methods. In *CVPR*, 2007.

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

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