Non-rigid structure from motion using quadratic deformation models

João Fayad[1]
jrkf@dcs.qmul.ac.uk

Alessio Del Bue[2]
adb@isr.ist.utl.pt

Lourdes Agapito[1]
lourdes@dcs.qmul.ac.uk

Pedro M.Q. Aguiar[2]
aguiar@isr.ist.utl.pt

[1] School of Electronic Engineering and Computer Science
Queen Mary, University of London
London, UK

[2] Institute of Systems and Robotics
Instituto Superior Técnico
Lisbon, Portugal

Reconstructing a 3D deformable surface from a set of 2D correspondences across an image sequence is the problem known as Non-Rigid Structure from Motion (NRSfM). In this scenario, it is usually assumed that the 3D deformable shape can be represented by a combination of rigid basis shapes with time varying coefficients. Due to its simplicity, linear basis shape models have provided good reconstructions under certain assumptions. However, NRSfM is inherently underconstrained and it remains an open problem. In particular these models successfully reconstruct objects which exhibit small deformations, such as faces, but have difficulties in modelling objects with strong ones.

Inspired by previous work in the fields of computer graphics [1, 2], we propose a quadratic model for non-rigid deformations based on geometric constraints. The shape coordinates are augmented with quadratic and cross terms allowing the model to represent non-linear local deformations.We show that our new model can deal with challenging sequences where algorithms based on the former linear bases model sometimes fail to converge.

In the case of a rigid object, the 3D coordinates of a world point $\mathbf{S}_j = [X_j Y_j Z_j]^T$ are projected on the image following the orthographic projection equation:

$$\mathbf{w}_{ij} = \begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} = \begin{bmatrix} r_{i1} & r_{i2} & r_{i3} \\ r_{i4} & r_{i5} & r_{i6} \end{bmatrix} \begin{bmatrix} X_j \\ Y_j \\ Z_j \end{bmatrix} + \mathbf{T}_i, \tag{1}$$

where $\mathbf{w}_{ij} = (u_{ij}\ v_{ij})^T$ are the non-homogeneous coordinates of point $\mathbf{S}_j$ in frame i; $\mathtt{R}_i$ is a 2×3 orthographic camera matrix that contains the first two rows of a rotation matrix and $\mathbf{T}_i$ is the 2×1 translation vector. Our quadratic deformation model for non-rigid bodies augments the rigid shape matrix with quadratic and cross-term components to account for the deformations of the object. We define the new shape matrix as

$$\mathtt{S} = \left[\begin{array}{cccc} X_1 & X_2 & \dots & X_p \\ Y_1 & Y_2 & \dots & Y_p \\ Z_1 & Z_2 & \dots & Z_p \\ \hline X_1^2 & X_2^2 & \dots & X_p^2 \\ Y_1^2 & Y_2^2 & \dots & Y_p^2 \\ Z_1^2 & Z_2^2 & \dots & Z_p^2 \\ \hline X_1Y_1 & X_2Y_2 & \dots & X_pY_p \\ Y_1Z_1 & Y_2Z_2 & \dots & Y_pZ_p \\ Z_1X_1 & Z_2X_2 & \dots & Z_pX_p \end{array}\right] = \left[\begin{array}{c} \mathtt{S}^{(\Gamma)} \\ \hline \mathtt{S}^{(\Omega)} \\ \hline \mathtt{S}^{(\Lambda)} \end{array}\right], \tag{2}$$

where $\mathtt{S}^{(\Gamma)}$ is the $3 \times p$ linear shape matrix which contains the 3D coordinates of the shape at rest and $\mathtt{S}^{(\Omega)}$ and $\mathtt{S}^{(\Lambda)}$ are simply the $3 \times p$ matrices that contain the quadratic and cross-terms respectively. To each of these three components of the new shape matrix is associated a deformation matrix such that the 3D coordinates of the deforming body at each frame i are defined as

$$\mathtt{S}_i = \mathtt{A}_i \mathtt{S} = \begin{bmatrix} \Gamma_i & \Omega_i & \Lambda_i \end{bmatrix} \mathtt{S} \tag{3}$$

where Γ_i, Ω_i and Λ_i are the 3×3 transformation matrices associated respectively with the linear, quadratic and cross-term deformations at frame i. Notice that $\mathtt{S}$, is fixed for all the frames while $\mathtt{A}_i$ varies frame-wise. By analogy with equation 1, the 2D image coordinates of the j^{th} world point in the i^{th} image will be given by $\mathbf{w}_{ij} = \mathtt{R}_i \mathtt{A}_i \mathbf{S}_j + \mathbf{T}_i$. Registering the 2D coordinates to the centroid we can form the bi-linear matrix model by stacking all the sub-block matrices for each frame obtaining:

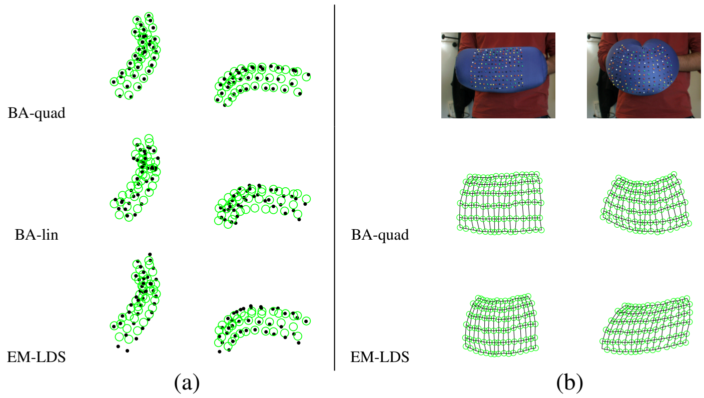

Figure 1: (a) 3D reconstruction of a cilindric object performing a bending motion. This sequence was obtained with a VICON Motion Capture system and artificially projected to 2D. Ground truth data is represented as green circles, while the reconstructions are presented as black dots; (b) 3D reconstruction of a cushion performing bending a stretching motions.

$$\mathtt{W} = \begin{bmatrix} \mathtt{R}_1 & & & \\ & \mathtt{R}_2 & & \\ & & \ddots & \\ & & & \mathtt{R}_f \end{bmatrix} \begin{bmatrix} \Gamma_1 & \Omega_1 & \Lambda_1 \\ \Gamma_2 & \Omega_2 & \Lambda_2 \\ \vdots & \vdots & \vdots \\ \Gamma_f & \Omega_f & \Lambda_f \end{bmatrix} \begin{bmatrix} \mathtt{S}^{(\Gamma)} \\ \mathtt{S}^{(\Omega)} \\ \mathtt{S}^{(\Lambda)} \end{bmatrix} = \mathtt{MS}, \tag{4}$$

where $\mathtt{W}$ is the $2f \times p$ measurement matrix.

Due to the non-linear nature of the model we have used a non-linear optimization to minimize image reprojection error, generically known as bundle adjustment [4].

Defining the 2D coordinates of point j at frame i given by our quadratic model as $\hat{\mathbf{w}}_{ij}$, the parameters of the model are then estimated by minimizing the following cost function:

$$\underset{\mathtt{R}_i,\mathbf{T}_i,\Gamma_i,\Omega_i,\Lambda_i,\mathbf{S}_j}{\arg\min} \sum_{i,j}^{f,p} \left|\left|\mathbf{w}_{ij} - \hat{\mathbf{w}}_{ij}\right|\right|^2 = \underset{\mathtt{R}_i,\mathbf{T}_i,\Gamma_i,\Omega_i,\Lambda_i,\mathbf{S}_j}{\arg\min} \sum_{i,j}^{f,p} \left|\left|\mathbf{w}_{ij} - \mathtt{R}_i \left[\Gamma_i \,\Omega_i \,\Lambda_i\right] \mathbf{S}_j - \mathbf{T}_i\right|\right|^2. \tag{5}$$

Synthetic and real experiments show that our quadratic model provides more accurate reconstructions as well as a higher convergence rate, on motions with strong deformations, when compared to a linear basis shapes model using the same image reprojection error approach (BA-lin), as well as the current state of the art method proposed by Torresani *et al.* [3].

[1] M. Müller, B. Heidelberger, M. Teschner, and M. Gross. Meshless deformations based on shape matching. In *SIGGRAPH 2006*, volume 24, pages 471–478, New York, NY, USA, 2005. ACM.

[2] S. I. Park and J. K. Hodgins. Capturing and animating skin deformation in human motion. In *SIGGRAPH 2006*, pages 881–889, New York, NY, USA, 2006. ACM.

[3] L. Torresani, A. Hertzmann, and C. Bregler. Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. *T-PAMI*, pages 878–892, 2008.

[4] B. Triggs, P. McLauchlan, R. I. Hartley, and A. Fitzgibbon. Bundle adjustment – A modern synthesis. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000.