

Non-rigid structure from motion using quadratic deformation models

João Fayad¹

jrkf@dcs.qmul.ac.uk

Alessio Del Bue²

adb@isr.ist.utl.pt

Lourdes Agapito¹

lourdes@dcs.qmul.ac.uk

Pedro M.Q. Aguiar²

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

Abstract

In this paper we present a new approach to the modelling of non-rigid 3D surfaces from the observation of 2D motion in images captured by an orthographic camera. Our aim is to characterize strong variations of the shape due, for instance, to bending motions. Such motions are hard to describe with previously used deformation models, such as the linear basis shapes model, which would tend to overestimate the dimensionality of the deformable data. Our approach uses a quadratic deformation model which is able to represent non-linear non-rigid motions such as bending, stretching, shearing and twisting. The model is bilinear and thus fits easily into previous schemes for Non-Rigid Structure from Motion (NRSfM). We formulate the NRSfM problem using a non-linear optimization scheme to minimize image reprojection error and recover the camera parameters, the 3D shape at rest and the quadratic deformation transformations. Our experiments with synthetic and real data show examples in which methods based on the linear basis shape model perform poorly or do not converge and instead the quadratic model is able to achieve accurate 3D reconstructions.

1 Introduction

In this paper we address the problem of reconstructing a 3D deformable surface from a set of 2D correspondences across an image sequence, a problem also known as Non-Rigid Structure from Motion (NRSfM). The most common assumption, made originally by Bregler *et al.* [5] and later adopted by most NRSfM methods, is that the deformable 3D shape can be represented as a linear combination of rigid basis shapes, or modes of deformation, with time varying coefficients. This linear shape model has allowed the development of a number of algorithms which can be seen as extensions of Tomasi and Kanade's classical rigid factorization algorithm [2, 3, 4, 6, 9, 13, 16, 18, 19].

Due to its simplicity, the linear basis shape model has proved a very successful representation, allowing solutions that can achieve good reconstructions under certain assumptions.

However, the problem of NRSfM is inherently under-constrained and remains open. In particular, although this low-rank shape model is well suited to representing the deformations of many common objects that exhibit small variations, such as faces, most methods have difficulties modelling strong deformations, since the underlying linear basis shapes model is restrictive. In this paper we are interested in modelling non-linear deformations (such as stretching and bending) that a linear basis shape model may have difficulty in explaining since the deformed surfaces may not lie on a low-dimensional linear shape manifold.

Inspired by previous work in the fields of computer graphics [10] and 3D human motion modelling [11], we depart from the linear setup by proposing a quadratic model for non-rigid deformations that uses geometric constraints. The 3D shape coordinates are augmented with quadratic and mixed terms allowing the model to represent non-linear local deformations such as stretching, bending and twisting motions. In this paper we show that the new model can deal with challenging sequences where algorithms based on the former linear bases model sometimes fail to converge. Moreover, since the formulation is bilinear, the quadratic model fits nicely into the NRSfM framework.

The main contribution of this paper is to introduce a physically grounded deformation model into the NRSfM formulation, which allows the 3D reconstruction of local non-linear deformations viewed by an orthographic camera. We then formulate the NRSfM problem using a non-linear optimization scheme to minimize image reprojection error and recover the camera parameters, the 3D shape at rest and the quadratic deformation transformations.

2 Previous Work

In their pioneering work to extend structure from motion to the case of non-rigid objects, Bregler *et al.*'s key insight [12] was to use a low-rank shape model to represent a deforming shape as a linear combination of k basis shapes which encode the main modes of deformation of the object. Based on this model, they proposed a non-rigid factorization method for an affine camera that exploited the rank constraint on the measurement matrix and enforced orthonormality constraints on camera rotations to recover the motion and the non-rigid 3D shape. Different iterative optimization schemes were later introduced by Torresani *et al.* [13] and Brand [14] to improve the computation of the metric upgrade.

Although the low-rank linear shape model has proved a successful representation, the NRSfM problem is inherently under-constrained. Recent approaches have focused on overcoming the problems caused by inherent ambiguities and degeneracies by proposing the use of generic priors or different optimization schemes. Aanaes *et al.* [15] impose the prior knowledge that the reconstructed shape does not vary much from frame to frame while Del Bue *et al.* [16] impose the constraint that some of the points on the object are rigid. Both approaches use bundle adjustment to refine all the parameters of the model simultaneously. Bartoli *et al.* [17] on the other hand, use a coarse to fine shape model where new deformation modes are added iteratively to capture as much of the variance left unexplained by previous modes as possible. Torresani *et al.* [13] propose to place a Gaussian prior distribution on the deformation weights which represents an explicit assumption that these will be similar to each other for each pose. They then generalise the model to represent linear dynamics in the deformations.

One advantage of the linear subspace model is that it has allowed closed form solutions to be proposed, for the cases of both affine [18] and perspective [9, 13] viewing conditions. However, closed form solutions have been reported to be very sensitive to noise [9, 17] and to the selection of the number of bases. Moreover, none of the closed form solutions

proposed so far can deal with missing data which becomes crucial when dealing with real video sequences.

Previous attempts to move away from the low-rank shape model within the NRSfM framework include work by Rabaud and Belongie [13] who define the set of possible 3D shapes as a smooth and low-dimensional manifold. Their approach assumes a non-linear optimization scheme imposing smoothness constraints on the deformations and enforcing a manifold dimensionality constraint on the 3D shapes.

Quadratic deformation models have been used before in the context of computer graphics [14] and human motion modelling using motion capture data [8, 15]. In the first case [14], the quadratic model was used to produce realistic simulations of deformable objects for computer games while in the second [8, 15], it was used to model skin deformations and demonstrated by capturing flexing muscles. However, both these approaches deal directly with 3D data. This paper proposes a new algorithm that extends the use of the quadratic deformation models to the problem of inferring 3D deformable models from 2D data using a NRSfM approach.

The paper is organised as follows. In Section 3 we review the problem formulation for NRSfM and describe the low-rank shape model. Section 4 describes the quadratic deformation model in detail including the physical meaning of its components. Section 5 describes our new NRSfM algorithm based on the quadratic shape model and finally Section 6 shows results on synthetic, motion capture and real sequences.

3 Non-rigid Structure from Motion

Consider the set of 2D image trajectories obtained when p points lying on the surface of a 3D object are viewed by a moving orthographic camera. 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:

$$\mathbf{w}_{ij} = \begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{i4} & r_{i5} & r_{i6} \end{bmatrix} \begin{bmatrix} X_j \\ Y_j \\ Z_j \end{bmatrix} + \mathbf{T}_i \quad (1)$$

where $\mathbf{w}_{ij} = (u_{ij} \ v_{ij})^T$ are the non-homogeneous coordinates of point \mathbf{S}_j in frame i ; \mathbf{R}_i is a 2×3 orthographic camera matrix that contains the first two rows of a rotation matrix (i.e. $\mathbf{R}_i \mathbf{R}_i^T = \mathbf{I}_{2 \times 2}$) and \mathbf{T}_i is the 2×1 translation vector.

Structure from Motion (SfM) can be defined as the problem of combined inference of the camera matrices (\mathbf{R}_i and \mathbf{T}_i) and the 3D shape S_j when only the 2D positions of the tracks \mathbf{w}_{ij} are known.

3.1 Low-Rank Linear Shape Model

In the case of deformable objects the observed 3D points change as a function of time. The low-rank shape model was first used in the context of NRSfM by Bregler *et al.* [5]. According to this model the 3D points deform as a linear combination of a fixed set of k rigid shape bases according to time varying coefficients. In this way, $\mathbf{S}_i = \sum_{d=1}^k l_{id} \mathbf{B}_d$ where $\mathbf{S}_i = [\mathbf{S}_{i1}, \dots, \mathbf{S}_{ip}]$ is the $3 \times p$ matrix that contains the 3D coordinates of the p points of the object at frame i , the $3 \times p$ matrices \mathbf{B}_d are the shape bases and l_{id} are the coefficient weights. Assuming an orthographic projection model, the coordinates of the 2D image points

observed at each frame i are then given by:

$$\mathbf{W}_i = \begin{bmatrix} u_{11} & \cdots & u_{1p} \\ v_{11} & \cdots & v_{1p} \end{bmatrix} = \mathbf{R}_i \left(\sum_{d=1}^k l_{id} \mathbf{B}_d \right) + \mathbf{T}_i, \quad (2)$$

where \mathbf{R}_i is a 2×3 orthographic camera matrix and the $2 \times p$ matrix \mathbf{T}_i aligns the image coordinates to the image centroid. When the image coordinates are registered to the centroid of the object and we consider all the frames in the sequence, we may write the measurement matrix \mathbf{W} as

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_{11} & \cdots & \mathbf{w}_{1p} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{f1} & \cdots & \mathbf{w}_{fp} \end{bmatrix} = \begin{bmatrix} l_{11}\mathbf{R}_1 & \cdots & l_{1k}\mathbf{R}_1 \\ \vdots & \ddots & \vdots \\ l_{f1}\mathbf{R}_f & \cdots & l_{fk}\mathbf{R}_f \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_k \end{bmatrix} = \mathbf{M}\mathbf{S}. \quad (3)$$

It is clear that the rank of \mathbf{W} is constrained to be at most $3k$. This rank constraint has allowed the extension of the classical rigid factorization framework [14] to the case of non-rigid objects. The factorization of the measurement matrix is not unique since the alternative factorization $\mathbf{W} = (\tilde{\mathbf{M}}\mathbf{Q})(\mathbf{Q}^{-1}\hat{\mathbf{S}})$ is also possible. The NRSfM problem is that of finding the matrix \mathbf{Q} that removes the affine ambiguity, upgrading the reconstruction to a metric one and rendering the correct repetitive structure to the motion matrix.

4 Quadratic Deformation Model for Non-Rigid Bodies

4.1 Quadratic model formulation

The 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. Let us first define the shape matrix for the quadratic model as

$$\mathbf{S} = \begin{bmatrix} X_1 & X_2 & \cdots & X_p \\ Y_1 & Y_2 & \cdots & Y_p \\ Z_1 & Z_2 & \cdots & Z_p \\ \hline X_1^2 & X_2^2 & \cdots & X_p^2 \\ Y_1^2 & Y_2^2 & \cdots & Y_p^2 \\ Z_1^2 & Z_2^2 & \cdots & Z_p^2 \\ \hline X_1Y_1 & X_2Y_2 & \cdots & X_pY_p \\ Y_1Z_1 & Y_2Z_2 & \cdots & Y_pZ_p \\ Z_1X_1 & Z_2X_2 & \cdots & Z_pX_p \end{bmatrix} = \begin{bmatrix} \mathbf{S}^{(\Gamma)} \\ \mathbf{S}^{(\Omega)} \\ \mathbf{S}^{(\Lambda)} \end{bmatrix}, \quad (4)$$

where $\mathbf{S}^{(\Gamma)}$ is the $3 \times p$ linear shape matrix which contains the 3D coordinates of the shape at rest and $\mathbf{S}^{(\Omega)}$ and $\mathbf{S}^{(\Lambda)}$ are simply the $3 \times p$ matrices that contain the quadratic and cross terms respectively. We now define \mathbf{A}_i , the quadratic deformation transformation matrix at frame i , as follows

$$\mathbf{A}_i = \begin{bmatrix} \Gamma_i & \Omega_i & \Lambda_i \end{bmatrix}, \quad (5)$$

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 . Applying the quadratic transformation to the shape matrix \mathbf{S} we obtain the 3D coordinates of the deforming body at each frame i as

$$\mathbf{S}_i = \mathbf{A}_i\mathbf{S} = \begin{bmatrix} \Gamma_i & \Omega_i & \Lambda_i \end{bmatrix} \mathbf{S}. \quad (6)$$

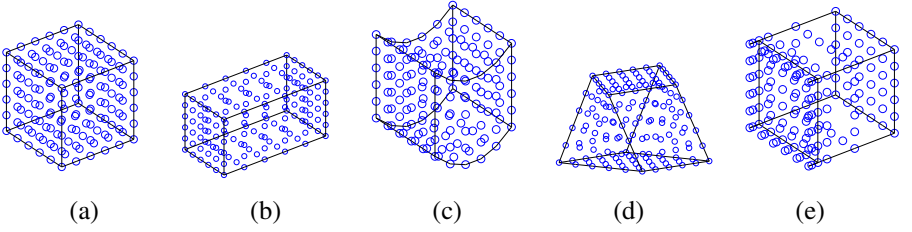


Figure 1: (a) Synthetic cube at rest. (b) Linear extension mode from the diagonal entries of Γ . (c) Bending deformation caused by the off-diagonal entries of Ω . (d) Twisting deformations caused by some entries of Λ . (e) Planar interpenetration caused by the diagonal entries of Ω .

Notice that the shape matrix S , which encodes the augmented coordinates of the shape at rest, is fixed for all the frames while the deformation matrix A_i varies frame-wise. This formulation assumes that the reference frame of the 3D shape $S^{(\Gamma)}$ is aligned with the principal axes of deformation of the object and centered at its centre of deformation. Figure 1 shows some examples of possible transformations applied to a synthetic cube. It is easy to observe that the linear transformation Γ_i accounts for stretching and shearing while the quadratic transformation Ω_i accounts for bending and the cross-term transformation Λ_i for twists.

4.2 Physical constraints on the quadratic model

One of the interesting features of the quadratic deformation model is that the entries of the transformation matrices have a physical meaning. Therefore, if prior knowledge exists about the physical properties of an object which could affect the way in which it deforms, this information could be used to pre-define some of the entries in the transformation matrices. For instance, when dealing with sequences of objects that cannot stretch, one should not consider the entries that account for it in Γ . On the other hand, if the object were not able to twist then entries Λ_{12} , Λ_{23} and Λ_{13} should be set to zero.

Moreover, if numerical bounds were not applied to the coefficients of the deformation matrices, the model could represent unrealistic deformations such as infinite extensions, or compressions to the point where the object would collapse onto a plane. Exercising all the degrees of freedom of the deformation matrices Γ_i , Ω_i and Λ_i reveals that some values of the coefficients of these matrices are not consistent with what is expected when modelling natural objects. For instance, deformations that involve interpenetration of the shapes are not physically plausible. An example of this kind of deformation is shown in Figure 1(e). These configurations can be avoided by setting the diagonal values in the quadratic transformation matrix Ω to zero.

An important observation concerning the linear deformation matrix Γ is that it can be decomposed using QR decomposition into the product of a rotation times an upper triangular matrix. Since the Γ must be expressed in the local referential, to avoid ambiguities the rotation matrix should be the identity and therefore Γ can be parameterized as an upper triangular matrix.

An alternative parameterization could be achieved by decomposing Γ via the polar decomposition, following the approach taken in the theory of elasticity in physics. In this case, to avoid ambiguities, Γ would be parametrized as a symmetric matrix. We carried out experiments to compare the performance of both parameterizations and found no difference

between them.

5 Non-Rigid SfM with a Quadratic Deformation Model

If we assume the deforming shape is viewed by an orthographic camera, the 2D image coordinates of the j^{th} world point in the i^{th} image will be given by

$$\mathbf{w}_{ij} = \Pi \mathbf{Q}_i [\Gamma_i \Omega_i \Lambda_i] \mathbf{S}_j + \mathbf{T}_i, \quad (7)$$

where the orthographic camera matrix Π is defined as

$$\Pi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad (8)$$

\mathbf{Q}_i is the 3×3 rotation matrix for frame i , \mathbf{T}_i is the 2D translation vector and \mathbf{S}_j is the vector that holds the 3D coordinates of point j when the object is at rest. Equation (7) can be rewritten as $\mathbf{w}_{ij} = \mathbf{R}_i \Lambda_i \mathbf{S}_j + \mathbf{T}_i$ where \mathbf{R}_i is the 2×3 truncated rotation matrix for frame i and Λ_i is the quadratic deformation matrix. Assuming that the 2D coordinates are registered to the centroid we can stack all the sub-block matrices for each frame obtaining:

$$\mathbf{W} = \begin{bmatrix} \mathbf{R}_1 & & & \\ & \mathbf{R}_2 & & \\ & & \ddots & \\ & & & \mathbf{R}_f \end{bmatrix} \begin{bmatrix} \Gamma_1 & \Omega_1 & \Lambda_1 \\ \Gamma_2 & \Omega_2 & \Lambda_2 \\ \vdots & \vdots & \vdots \\ \Gamma_f & \Omega_f & \Lambda_f \end{bmatrix} \begin{bmatrix} \mathbf{S}^{(\Gamma)} \\ \mathbf{S}^{(\Omega)} \\ \mathbf{S}^{(\Lambda)} \end{bmatrix} = \mathbf{M}\mathbf{S}, \quad (9)$$

where \mathbf{W} is the $2f \times p$ measurement matrix. Note that if $\Gamma_i = I_{3 \times 3}$, $\Omega_i = O_{3 \times 3}$ and $\Lambda_i = O_{3 \times 3}$ for every $i = 1 \dots f$, the formulation will reduce to the classical rigid factorization model [14] with $\text{rank}(\mathbf{W}) \leq 3$. Also note that while the coefficients in \mathbf{R}_i , Λ_i , Ω_i and Γ_i vary from frame to frame to encode the rotations and the deformations, the shape matrix is fixed for all frames.

The rank of the measurement matrix \mathbf{W} is at most 9. This property could be used to factorize \mathbf{W} into $\mathbf{W} = \mathbf{M}\mathbf{S}$ where \mathbf{M} would encapsulate the rotation and deformation matrices and \mathbf{S} the augmented shape matrix. However, this factorization would be ambiguous and problem would therefore consist of computing the transformation that renders the correct rotation and deformation matrices which would result in a non-linear problem.

5.1 Non-linear optimization

In order to fit the quadratic model to our data, we need to deal with the non-linearities of the model. An alternating approach such as *power-factorization* [15] would not be suitable in this case since the motion and shape matrices are intrinsically non-linear. Instead we have used non-linear optimization to minimize image reprojection error, or bundle adjustment [16]. This approach combines efficient minimization using a Levenberg-Marquardt optimization algorithm together with faster computation exploiting the sparse properties of the Jacobian and Hessian matrices computed at each iteration of the minimization.

The 2D coordinates of point j at frame i given by our quadratic model can be written as

$$\hat{\mathbf{w}}_{ij} = \mathbf{R}_i [\Gamma_i \Omega_i \Lambda_i] \mathbf{S}_j + \mathbf{T}_i. \quad (10)$$

The parameters of the camera matrices, the quadratic deformations and the 3D shape are then estimated by minimizing image reprojection error, defined as the difference between

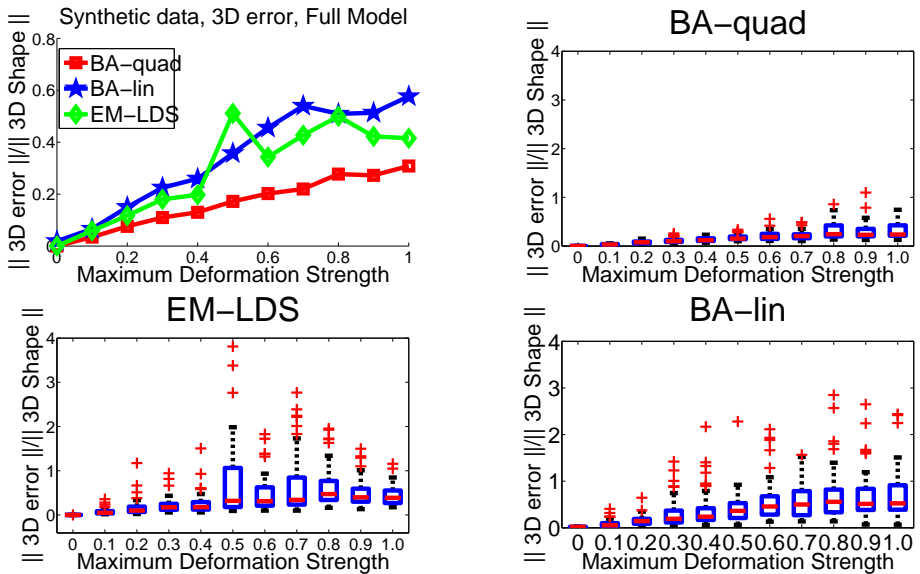


Figure 2: Results on synthetic data for varying deformation strengths. Top-left: Average 3D error plot for experiments that converged to a valid solution. Box-plots are provided to illustrate the rate of convergence of the three different algorithms. Notice the high rate of convergence failure of the EM-LDS and BA-Lin algorithms.

the measured and estimated image points:

$$\arg \min_{\mathbf{R}_i, \mathbf{T}_i, \Gamma_i, \Omega_i, \Lambda_i, \mathbf{S}_j} \sum_{i,j} \|\mathbf{w}_{ij} - \hat{\mathbf{w}}_{ij}\|^2 = \arg \min_{\mathbf{R}_i, \mathbf{T}_i, \Gamma_i, \Omega_i, \Lambda_i, \mathbf{S}_j} \sum_{i,j} \|\mathbf{w}_{ij} - \mathbf{R}_i [\Gamma_i \Omega_i \Lambda_i] \mathbf{S}_j - \mathbf{T}_i\|^2. \quad (11)$$

One of the most important advantages of using a non-linear minimization scheme to minimize image reprojection error is that any prior information available about the nature of object being observed that has an effect on the values that the deformation matrices Γ_i , Ω_i and Λ_i can take may be incorporated into the cost function. Some possible constraints or explicit values of the parameters were discussed in Section 4.2. We have added a regularization term that enforces smoothness constraints on the frame-wise 3D shapes.

Initialization: Bundle adjustment methods rely on the initial estimates being close to the global minimum to avoid falling into local solutions. In our minimization problem we must provide adequate initial values for the rotation matrices, the coefficients of the 3 deformation matrices and the matrix containing the shape at rest. The rotations are initialized using the rigid motion given by the Tomasi-Kanade [TK91] rigid factorization of the sequence. The deformation matrices Γ , Ω and Λ are initialized to represent a rigid shape, i. e. $\Gamma_i = I_{3 \times 3}$, $\Omega_i = O_{3 \times 3}$ and $\Lambda_i = O_{3 \times 3}$ for every $i = 1 \dots f$. In this work we assume that the sequence starts with the body at rest for some initial frames and Tomasi and Kanade’s rigid factorization is used on those frames to obtain an initial estimate of the 3D shape matrix \mathbf{S} . Finding an initialization for the 3D shape at rest for general sequences requires further research. Since our formulation assumes that the 3D shape is aligned and centered with the principal deformation axes an estimate of these must be provided. We assume the axes of deformation are aligned with

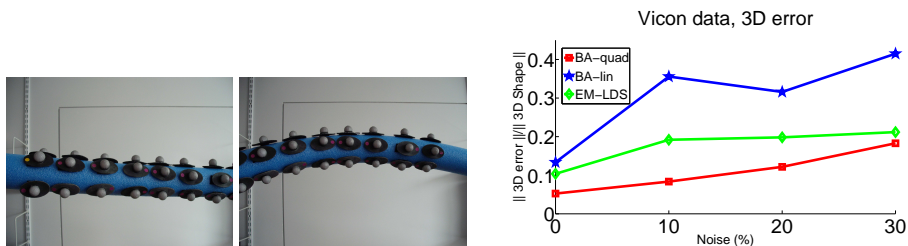


Figure 3: Left: Examples of the woggle used in the experiments with motion capture data. Right: Average 3D error plot for experiments that converged to a valid solution for increasing levels of noise. The experiments show that BA-Quad outperforms EM-LDS and BA-Lin.

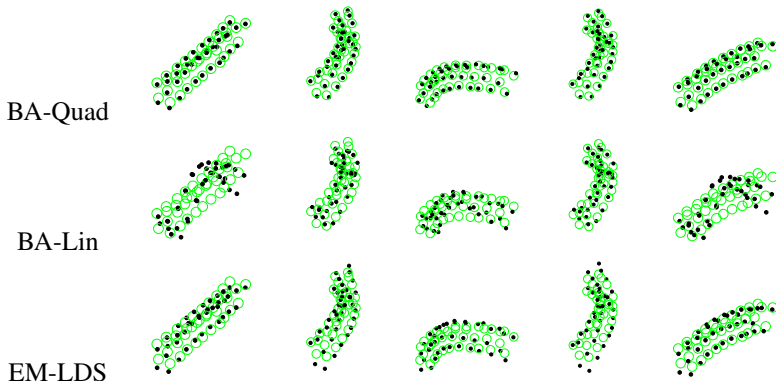


Figure 4: Example frames of the 3D reconstructions of the real woggle mocap data sequence obtained with the 3 different algorithms.

the principal axes of inertia which we compute using SVD on the 3D coordinates of the initial estimate of the shape at rest and the deformation centre at the centroid of the object. In the sequences we have analysed in our experiments the axes and centre of deformation remain fixed and therefore we do not estimate the parameters. However, they could be easily added to the cost function and optimized in the bundle adjustment process.

6 Experiments

Synthetic cylinder sequence

The 3D shape used in the synthetic sequences was generated to simulate a thin but long cylinder (similar to the object depicted in Figure 3) to allow strong bending motions. We then applied deformations of increasing maximum strength to the object, using deformation matrices Γ , Ω and Λ with random coefficients ranging in maximum magnitude from 0 to 1. We generated 50 random tests for each level of deformation, keeping all the other parameters unchanged. The 3D points were projected onto the image using an orthographic camera model. The sequences start with the object at rest for some initial frames.

We compare the results of our new algorithm based on the quadratic shape model (BA-Quad) with Torresani *et al.*'s state of the art algorithm [16] (EM-LDS) and with a Bundle Adjustment algorithm (BA-Lin) [2], both of which are based on the linear low-rank shape

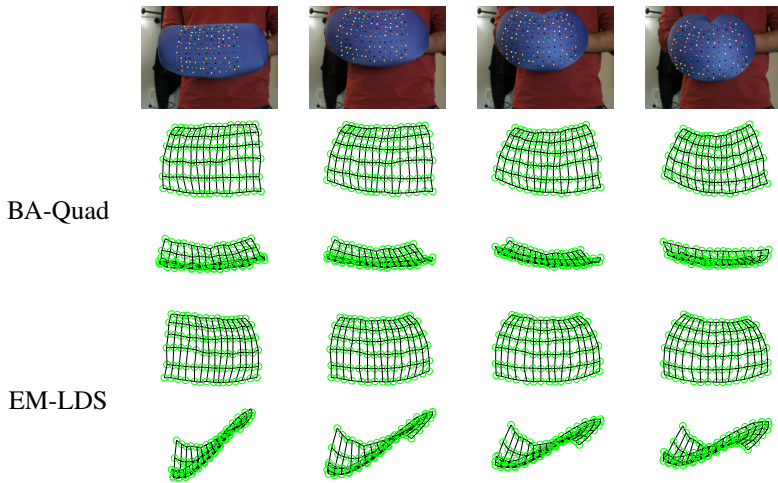


Figure 5: Top: Selected frames from the video sequence of a cushion bending and stretching. Rows 2 and 3: Front and top views of the 3D shapes for the selected frames using our new quadratic model. Rows 4 and 5: 3D reconstructions using EM-LDS

model. Both bundle adjustment algorithms were initialized using Tomasi and Kanade’s algorithm on the first 10 frames of the sequence, in which the shape was at rest. We computed the 3D error as the sum of the squared differences between the estimated 3D shapes and the ground truth divided by the norm of the shape. Figure 2 shows results of the average 3D error as well as box-plots for each of the algorithms. The average error plot was generated after removing the results from tests that failed to converge (outliers marked as red crosses in the box-plots). Our new algorithm outperforms the other methods in two important aspects. First, the box plots show a superior convergence rate to the other algorithms. With BA-quad 3.09% of all the tests converged while with EM-LDS as many as 8.91% of the tests failed to converge and 9.45% with BA-Lin. Secondly, amongst the tests that converged, the average error plot (Figure 2 top left) shows that the smallest 3D error was given by our new algorithm.

Experiments with real deformations from mocap data

In these experiments we used 3D motion capture data of a water woggle (or swimming noodle) which is a long and thin polystyrene cylinder that can undergo strong bending deformations. The 3D data was captured using a VICON system by tracking 30 markers. Figure 3 shows a few images of the object (with the markers) deforming. The 3D points were then projected onto an image sequence 676 frames long using an orthographic camera model. We evaluated the performance of the algorithm with respect to noise in the image measurements. Zero mean additive Gaussian noise was applied with standard deviation $\sigma = n \times s / 100$ where n is the noise percentage and s is defined as the diameter of the woggle in pixels. Noise levels of up to 30% were added. Figure 3 (right), shows the plot comparing the results obtained with our algorithm with those achieved using EM-LDS and BA-Lin. The plot depicts the 3D error averaged over 50 random runs after removing the results from tests that failed to converge showing an improved performance of the quadratic algorithm versus EM-LDS and BA-Lin. Figure 4 shows the ground truth (circles) and reconstructed 3D shapes (dots) for

five frames of the sequence in the absence of noise using the three different algorithms where our new algorithm shows consistent improved reconstructions.

Real experiments

Figure 5 shows a few frames of a real sequence of a cushion bending and stretching, in which 90 points were tracked. Front and top views of the reconstructions achieved using our new quadratic model and Torresani *et al.*'s EM-LDS algorithm are shown for comparison. Best results are obtained with the new Bundle Adjustment algorithm using the quadratic deformation model.

7 Conclusions

The main contribution of this paper is to introduce a physically grounded deformation model into the NRSfM formulation, which allows the 3D reconstruction of non-linear deformations viewed by an orthographic camera. The focus of this paper is to show examples where the proposed quadratic model provides a better representation in the case of strong physical non-linear deformations than algorithms based on the low-rank linear model. In our comparative tests, these algorithms have shown to have a higher rate of convergence failure and to provide less accurate 3D reconstructions than those obtained with our new quadratic algorithm. Future work includes using piecewise models to cope with higher dimensional deformations.

8 ACKNOWLEDGEMENTS

This work was partially funded by Fundacao para a Ciéncia e a Tecnologia (ISR/IST pluri-annual funding) through the POS_Conhecimento Program (include FEDER funds) and grant PTDC/EEA-ACR/72201/2006, and by the European Research Council under ERC Starting Grant agreement 204871-HUMANIS.

References

- [1] H. Aanæs and F. Kahl. Estimation of deformable structure and motion. In *Workshop on Vision and Modelling of Dynamic Scenes, Copenhagen, Denmark*, 2002.
- [2] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd. Coarse-to-Fine Low-Rank Structure-from-Motion. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska*, 2008.
- [3] M. Brand. Morphable models from video. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii*, volume 2, pages 456–463, December 2001.
- [4] M. Brand. A direct method for 3D factorization of nonrigid motion observed in 2D. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California*, pages 122–128, 2005.
- [5] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina*, pages 690–696, June 2000.
- [6] A. Del Bue, X. Lladó, and L. Agapito. Non-rigid metric shape and motion recovery from uncalibrated images using priors. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, New York, NY*, 2006.

- [7] A. Del Bue, F. Smeraldi, and L. Agapito. Non-rigid structure from motion using ranklet-based tracking and non-linear optimization. *Image and Vision Computing*, 25(3):297–310, March 2007.
- [8] J. Fayad, A Del Bue, L. Agapito, and P.M.Q. Aguiar. Human body modelling using quadratic deformations. In *7th EUROMECH Solid Mechanics Conference*, Lisbon, Portugal, 2009.
- [9] R. Hartley and R. Vidal. Perspective nonrigid shape and motion recovery. In *Proc. European Conference on Computer Vision*, pages 276–289, 2008.
- [10] R. I. Hartley and F. Schaffalitzky. Powerfactorization: an approach to affine reconstruction with missing and uncertain data. In *Australia-Japan Advanced Workshop on Computer Vision*, Adelaide, Australia, September 2003.
- [11] Matthias Müller, Bruno Heidelberger, Matthias Teschner, and Markus Gross. Meshless deformations based on shape matching. In *SIGGRAPH 2006*, volume 24, pages 471–478, New York, NY, USA, 2005. ACM.
- [12] Sang Il Park and Jessica K. Hodgins. Capturing and animating skin deformation in human motion. In *SIGGRAPH 2006*, pages 881–889, New York, NY, USA, 2006. ACM.
- [13] Vincent Rabaud and Serge Belongie. Re-thinking non-rigid structure from motion. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska*, pages 1–8, 2008.
- [14] 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.
- [15] L. Torresani, D. Yang, E. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii*, 2001.
- [16] L. Torresani, A. Hertzmann, and C. Bregler. Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 878–892, 2008.
- [17] 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.
- [18] J. Xiao and T. Kanade. Uncalibrated perspective reconstruction of deformable structures. In *Proc. 10th International Conference on Computer Vision, Beijing, China*, October 2005.
- [19] J. Xiao, J. Chai, and T. Kanade. A closed-form solution to non-rigid shape and motion recovery. *International Journal of Computer Vision*, 67(2):233–246, April 2006.