Good Vibrations: A Modal Analysis Approach for Sequential Non-Rigid Structure from Motion

Antonio Agudo[1] Lourdes Agapito[2] Begoña Calvo[1] J. M. M. Montiel[1]
[1]Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, Spain
[2]Department of Computer Science, University College London, UK

Abstract

We propose an online solution to non-rigid structure from motion that performs camera pose and 3D shape estimation of highly deformable surfaces on a frame-by-frame basis. Our method models non-rigid deformations as a linear combination of some mode shapes obtained using modal analysis from continuum mechanics. The shape is first discretized into linear elastic triangles, modelled by means of finite elements, which are used to pose the force balance equations for an undamped free vibrations model. The shape basis computation comes down to solving an eigenvalue problem, without the requirement of a learning step. The camera pose and time varying weights that define the shape at each frame are then estimated on the fly, in an online fashion, using bundle adjustment over a sliding window of image frames. The result is a low computational cost method that can run sequentially in real-time.

We show experimental results on synthetic sequences with ground truth 3D data and real videos for different scenarios ranging from sparse to dense scenes. Our system exhibits a good trade-off between accuracy and computational budget, it can handle missing data and performs favourably compared to competing methods.

1. Introduction

The combined inference of 3D scene structure and camera motion from monocular image sequences, or Structure from Motion (S*f*M), is one of the most active areas in computer vision. The last decade has seen significant progress towards real-time recovery of camera pose and 3D shape for a sparse set of salient points [14] and even per-pixel real-time dense reconstructions [18]. While S*f*M is now considered to be a mature field, these methods cannot be applied to scenes undergoing non-rigid deformations.

Non-Rigid Structure from Motion (NRS*f*M) addresses this limitation and methods from this field are now capable of creating accurate 3D reconstructions of moving and deformable objects with striking results [10]. The underlying principle behind most approaches is to model deformations using a low-rank shape [6, 29, 4, 20, 19, 7, 10] or trajectory [3, 13] basis. However, NRS*f*M methods remain behind their rigid counterparts when it comes to real-time performance. The reason behind this is that they are typically limited to batch operation where all the frames in the sequence are processed at once, after their acquisition, preventing them from *on-line* and *real-time* performance. Only recently, have NRS*f*M methods been extended to sequential processing [19, 2]. However they remain slow [19] or do not scale to the use of a large number of points [19, 2].

In this paper we push monocular NRS*f*M forward towards real-time operation by proposing an online algorithm to recover the 3D non-rigid shape and pose of strongly deforming surfaces under realistic real-world assumptions: our system can deal with significant occlusions, non-isometric deformations (stretching) and can be used with dense data. We use a linear combination of mode shapes (estimated using continuum mechanics) to model the non-rigid shape. Our approach works in two stages. Firstly, we compute the shape at rest, and estimate the mode shapes solving a simple eigen-value problem. Equipped with this low-rank deformable shape basis, the system continues operation in a sequential manner where the only parameters to estimate per-frame are the camera pose and a small number of basis coefficients, optimised using Bundle Adjustment (BA) over a sliding window.

2. Related work

NRS*f*M is an ill-posed problem unless additional constraints, such as smoothness priors, are considered. Bregler *et al.*'s seminal work [6] was proposed as an extension of Tomasi and Kanade's factorization algorithm [28] to the non-rigid case. Their key insight was to model time-varying shape as a linear combination of an unknown shape basis under orthography. Different optimisation schemes have since been proposed to include temporal and spatial smoothness priors [29]. BA has been used at the core of

the optimisation of the pose, shape basis and coefficients as it allows to incorporate both motion and deformation priors [8, 4]. In [20] a solution is proposed that recovers motion matrices that lie on the correct motion manifold where the metric constraints are exactly satisfied. More recently, low-rank shape and local smoothness priors have been combined within a variational approach to produce per-pixel dense vivid reconstructions [10]. Piecewise models were proposed to encode more accurately strong local deformations. Piecewise planar [31], locally rigid [27] or quadratic [9] approaches rely on common features shared between patches to enforce global consistency. Russell *et al*. [22] proposed a formulation to automate the best division of the surface into local patches. The Finite Element Method (FEM) formulation proposed in [2, 1] injects surface continuity to the stretchable triangles used as local elements.

The alternative strand of methods known as template-based [24, 23, 16] rely on correspondences between the 2D points in the current image and a reference 3D shape which is assumed to be known in advance. The unknown shape is encoded as a linear combination of deformation modes learnt in advance from a relatively large set of training data [24]. To avoid inherent ambiguities, additional shape constraints are required such as inextensibility. More recently, the exclusive use of inextensibility constraints, without knowledge of the shape template, has been shown to be sufficient to perform non-rigid reconstruction [32].

Despite these advances, previous approaches to NRS*f*M typically remain batch and process all the frames in the sequence at once, after video capture. While sequential real-time S*f*M [14, 18] solutions exist for rigid scenes, online estimation of non-rigid shape from a single camera remains a challenging problem. Only recently, have sequential formulations emerged — Paladini *et al*. [19] proposed the first sequential NRS*f*M system based on BA over a sliding window. However, their approach did not achieve real-time performance and was only demonstrated on a small number of feature points. The first real-time, online solution to NRS*f*M [2, 1] combined an extended Kalman filter with FEM to build small maps of non-rigid scenes.

Our Contribution: In this work, we present a sequential solution to the NRS*f*M problem. We combine a physics-based model to define a number of meaningful deformation mode shapes, with a sequential BA framework to obtain a low computational cost system that can run in real-time. We are able to reconstruct highly extensible surfaces without the need for a pre-trained model.

3. Continuum Mechanics Deformation Model

A common way to model non-rigid shapes in computer vision is to represent the 3D shape as a linear combination of shape basis [6, 29, 4, 24, 20, 19, 16, 7, 10]. In this work, we propose a method to compute the shape basis from a continuum mechanics physics-based model of the scene. We first review some concepts from dynamics in continuum mechanics that will lead to our formulation of the estimation of the deformable shape basis.

3.1. Undamped Free Vibrations Analysis

The dynamic behaviour of a deformable solid has been widely studied in mechanical engineering [5]. Numerical methods are mandatory to approximately solve the partial differential equations modelling this behaviour, with FEM being the most common approach.

A continuous object Ω is discretised into a number of *finite elements* Ω_e (Fig. 1). Each element is defined by the 3D location its nodes. Hence the geometry of the undeformed solid is encoded as $\mathbf{S} \in \mathbb{R}^{3\times p}$ where the matrix columns are the 3D coordinates defining the location of the p discretisation nodes:

$$\mathbf{S} = \begin{bmatrix} X_1 & X_2 & \dots & X_p \\ Y_1 & Y_2 & \dots & Y_p \\ Z_1 & Z_2 & \dots & Z_p \end{bmatrix}. \tag{1}$$

The fundamental equations of structural dynamics are elaborate versions of Newtonian mechanics formulated as force balance statements. The governing force balance equations for *undamped free vibrations* –without energy dissipation or external forces– can be written as:

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{0} \tag{2}$$

where $\mathbf{M}$ and $\mathbf{K}$ are the global mass and stiffness $3p \times 3p$ matrices respectively, $\mathbf{u}$ is a $3p \times 1$ 3D nodal displacement vector and the motion at each node j-th is specified by means of $\mathbf{u}_j = (\Delta X_j, \Delta Y_j, \Delta Z_j)^\top$. Derivatives with respect to t are abbreviated by superposed dots, i.e. $\ddot{\mathbf{u}}(t) \equiv \frac{d^2\mathbf{u}(t)}{dt^2}$. In this equation and in the absence of external loads, the internal elastic forces $\mathbf{K}\mathbf{u}$ balance the negative of the inertial forces $\mathbf{M}\ddot{\mathbf{u}}$ and it can be interpreted as assigning a certain mass and certain stiffness between nodal points. The equation is linear and homogeneous, and its solution is a linear combination of exponentials modulated by the mode shapes.

3.2. Stiffness and Mass Matrices

This section is devoted to the computation of the matrices $\mathbf{K}$ and $\mathbf{M}$. We discretise the surface of the observed solid into m linear triangular elastic elements with $\mathcal{E}$ connectivity. We compute the stiffness matrix $\mathbf{K}$ by means of a model for thin-plate elements. The deformation is modelled as a combination of plane-stress and Kirchhoff's plate, using the free-boundary conditions matrix for linear elastostatics as proposed in [2]:

$$\mathbf{K} = \mathop{\mathbf{A}}_{e=1}^{m} \int_{\Omega_e} \mathcal{T}^\top \mathbf{B}_e^\top \, \mathbf{D} \, \mathbf{B}_e \mathcal{T} \, d\Omega_e \tag{3}$$

where $\mathbf{B}_e$ is the strain-displacement matrix defined in terms of the shape function derivatives. $\mathbf{D}$ is the constitutive matrix depending on the material elastic properties: Young's modulus E and Poisson's ratio ν. We assume near incompressible materials $\nu \approx 0.5$. $\mathcal{T}$ is the local-to-global displacement transformation matrix while $\mathbf{A}$ represents the assembly process of elemental matrices.

In order to model the mass matrix $\mathbf{M}$, we consider two scenarios [5]. First we assume a *distributed mass* within element:

$$\mathbf{M}_d = \mathop{\mathbf{A}}_{e=1}^{m} \int_{\Omega_e} \mathbf{N}_e^\top \, \rho \, \mathbf{N}_e \, d\Omega_e \tag{4}$$

where $\mathbf{N}_e$ is the interpolation matrix containing the linear shape functions and ρ is the material density (mass-density). We have assumed for simplicity that the density is constant, i.e., $\frac{d\rho}{dt} = 0$, and we also rely on the fact that the mass-density is the same for all elements.

The second scenario assumes *lumped mass* at the nodes, leading to the mass matrix being computed as:

$$\mathbf{M}_l = \mathop{\mathbf{A}}_{e=1}^{m} \frac{\rho h A_e}{3} \, \mathbf{diag}\big([1\,1\,1\,1\,1\,1\,1\,1\,1]^\top\big) \tag{5}$$

preserving the total element mass $\sum_a \tilde{M}^e_{aa} = \int_{\Omega_e} \rho \, d\Omega_e$ where $\tilde{M}^e_{aa}$ is the mass per component. For simplicity, the surface thickness h is the same for all elements. A_e represents the element area.

3.3. Modal Analysis

According to the structural engineering FEM analysis [5], the deformed object at a given sample time can be approximated as a linear combination of some mode shapes which can be computed as a generalized eigenvalue problem from the undamped free vibration dynamics Eq. (2). It is worth noting, for the linear case, that the mode basis does not depend on the applied forces but just on the $\mathbf{K}$ and $\mathbf{M}$ matrices computed from the shape at rest.

Modal analysis is standard in structural engineering, it has been also applied in computer vision for motion analysis to track and recover the heart motion [21, 17], and in [25, 26] for non-rigid 2D tracking. Modal analysis was used to decoupling the equilibrium equations by obtaining a closed-form solution. In this work, we propose to use modal analysis with a soup of elastic triangles with unknown material properties (E, ρ) and without knowledge of the boundary conditions. We could exploit them when available, typically Dirichlet constraints to fix points $\mathbf{u}_j = \mathbf{0}$. We directly use the mode basis without using the decoupled system.

From the system Eq. (2), we can compute the undamped free vibration response of the structure caused by a disturbance with respect to the shape at rest by solving the generalized eigenvalue problem in ω^2:

$$\left(\mathbf{K} - \omega_k^2 \mathbf{M}\right) \psi_k = \mathbf{0} \tag{6}$$

where $\left\{\psi_k, \omega_k^2\right\}, k = 1 \ldots 3p$ are the *mode shapes* (eigenvectors) and *frequencies* (eigenvalues) respectively. Each eigenmode ψ_k is a $3p \times 1$ vector and is composed of the displacements for all the p nodes defining the discretisation. We compute the normalized to length one modes $\|\psi_k\|_2 = 1$ in order to satisfy the orthonormality conditions $\psi_k^\top \mathbf{M}^{-1} \mathbf{K} \psi_i = \omega_k^2$ and $\psi_k^\top \psi_i = \delta_{ki}$ where δ_{ki} is the Kronecker's delta.

3.4. Mode Shapes: Analysis and Selection

In the case of non-boundary conditions, i.e. rigid points, modal analysis yields $3p$ orthonormal modes –provided $\mathbf{K}$ and $\mathbf{M}$ are symmetric positive definite– (see some examples in Fig. 1). To analyse the mode shapes we sort them according to the energy needed to excite to each mode – *frequency spectrum*–, from lower to higher frequency. We can approximately identify three practical mode families, instead of two proposed in [25], corresponding to:

Rigid motion modes. Theoretically the first 6 frequencies should be zero, because they correspond to 6 d.o.f. rigid body motions. However in practice, due to the thin-plate approximation [2], only the first 4 frequencies are zero up to numerical error. These 4 rigid motion modes are excluded from the mode basis when coding non-rigid deformations.

Bending modes. Bending, out-of-plane deformations, are mainly represented by the modes in the interval $[5, p+4]$ (see Fig. 1). These modes can represent elastic bending deformations (with low stretching in-plane). Moreover, selecting a few of the first bending modes provides an accurate mode basis to model bending deformations.

Stretching modes. Stretching deformations can be modelled as a linear combination of the modes in the interval $[p+5, 2p]$. Similarly selecting only the first stretching modes provides a mode basis to accurately represent the stretching in-plane deformations.

The rest of the mode shapes $[2p+1, 3p]$ — the higher frequencies — do not correspond to physical deformations but to artifacts due to the discretisation process. Note that we separate the intermediate modes proposed in [25] into bending and stretching modes for the 3D case.

To sum up, any non-rigid displacement $\mathbf{u}$, can be spanned by a mode basis selecting only $\{k_1, \ldots, k_r\}, r << 3p$ mode shapes. For notational simplicity, it is assumed that the r selected modes are renumbered $k = 1, \ldots, r$, so any displacement vector can be approximated by a linear combination of the mode shapes:

$$\mathbf{u} \approx \sum_{k=1}^{r} \gamma_k \psi_k \tag{7}$$

where γ_k are weights to obtain a lower dimensional representation. Expression (7) can be rewritten as $\mathbf{u} = \left(\mathbf{I}_3 \otimes \mathbf{\Gamma}^\top\right) \mathbf{\Psi}$, where $\mathbf{I}_3$ is a 3×3 identity matrix and $\otimes$

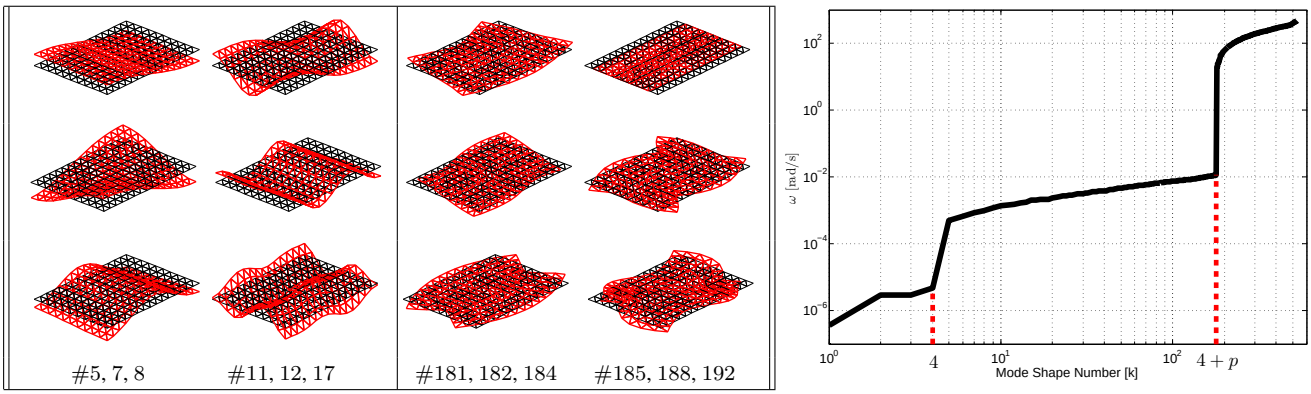

Figure 1. **Left:** Representation of some non-rigid mode shapes of a plate surface discretised into triangular elements with #176 nodes. We show the effect of adding the mode shape (red mesh) to the shape at rest (black mesh) using arbitrary weight. Note that the effect of subtracting can be obtained using the opposite weight. First and second column: bending mode shapes. Third and fourth column: membrane mode shapes. **Right:** Eigenfrequencies ω_k for the previous black mesh in logarithmic scale. Best viewed in color.

represents the Kronecker's product. We use a transformation matrix $\mathbf{\Psi} \in \mathbb{R}^{3r\times p}$, which has the 3D displacement coordinates of p points using r mode shapes as:

$$\mathbf{\Psi} = \begin{bmatrix} \bar{\psi}_1 \\ \vdots \\ \bar{\psi}_r \end{bmatrix} = \begin{bmatrix} \Delta X_{11} & \Delta X_{12} & \dots & \Delta X_{1p} \\ \Delta Y_{11} & \Delta Y_{12} & \dots & \Delta Y_{1p} \\ \Delta Z_{11} & \Delta Z_{12} & \dots & \Delta Z_{1p} \\ \vdots & \vdots & & \vdots \\ \Delta X_{r1} & \Delta X_{r2} & \dots & \Delta X_{rp} \\ \Delta Y_{r1} & \Delta Y_{r2} & \dots & \Delta Y_{rp} \\ \Delta Z_{r1} & \Delta Z_{r2} & \dots & \Delta Z_{rp} \end{bmatrix}, \quad (8)$$

and another vector $\mathbf{\Gamma}$ for the frequency of vibration as:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1 & \dots & \gamma_k & \dots & \gamma_r \end{bmatrix}^\top. \quad (9)$$

4. Sequential NRS*f*M

We propose to use the shape basis resulting from modal analysis to represent the non-rigidly deforming scene. This section is devoted to describe the details of our sequential approach to NRS*f*M.

4.1. Proposed Deformation Model

We approximate the non-rigid shape at each instant as a linear combination of mode shapes. Hence the estimation of the 3D structure at each frame f comes down to estimating the corresponding weight vector $\mathbf{\Gamma}_f$. The deformed structure at frame f can be written as $\mathbf{S}_f = \mathbf{S} + \mathbf{u}_f = \mathbf{S} + \left(\mathbf{I}_3 \otimes \mathbf{\Gamma}_f^\top\right)\mathbf{\Psi}$.

Note that the mode shapes do not depend on the material properties (E, ρ). For the same shape at rest, the normalized modes are the same irrespective of (E, ρ). Different (E, ρ) would produce the same deformation but with different amplitude, which can be absorbed into the deformation weights.

We assume an orthographic camera model, where the projection of p points onto image frame f is expressed as:

$$\mathbf{W}_f = \begin{bmatrix} u_{f1} & u_{f2} & \dots & u_{fp} \\ v_{f1} & v_{f2} & \dots & v_{fp} \end{bmatrix} = \Pi\mathbf{Q}_f(\mathbf{S}_f + \mathbf{T}_f), \quad (10)$$

where Π is the 2×3 orthographic camera matrix, $\mathbf{Q}_f$ is the 3×3 rotation matrix and $\mathbf{T}_f$ stacks p copies of a 3×1 translation vector and $\mathbf{R}_f = \Pi\mathbf{Q}_f$ are the first two rows of a full rotation matrix. Without loss of generality, we can register all the measurements to their image centroid [28]. Considering $\mathbf{S}_f = \mathbf{S} + \left(\mathbf{I}_3 \otimes \mathbf{\Gamma}_f^\top\right)\mathbf{\Psi}$, we can write the projection equation (10) for all frames using a unique matrix where all the unknown weights γ_{fk} are stacked:

$$\mathbf{W} = \begin{bmatrix} \mathbf{R}_1 & & \\ & \ddots & \\ & & \mathbf{R}_f \end{bmatrix} \left(\begin{bmatrix} \mathbf{S} \\ \vdots \\ \mathbf{S} \end{bmatrix} + \begin{bmatrix} \mathbf{I}_3 \otimes \mathbf{\Gamma}_1^\top \\ \vdots \\ \mathbf{I}_3 \otimes \mathbf{\Gamma}_f^\top \end{bmatrix} \begin{bmatrix} \bar{\psi}_1 \\ \vdots \\ \bar{\psi}_r \end{bmatrix} \right). \quad (11)$$

4.2. Non-linear Optimisation

Recall that the deformation modes $\mathbf{\Psi}$ are computed in a previous step following the modal analysis described in Section 3 and only requiring an estimate of the shape at rest $\mathbf{S}$. Therefore, the sequential NRS*f*M problem is reduced to the estimation of per-frame camera motion $\mathbf{R}_i$ and deformation weights $\mathbf{\Gamma}_i$. This involves the estimation of a very small number of parameters r to encode the shape at each frame, which leads to a low computational cost system that can potentially run in real-time.

We use a sliding temporal window approach as proposed in [19], to perform BA [30] on the last $\mathcal{W}$ frames. The model parameters are estimated by minimizing the image reprojection error of all the observed points ϱ ($\mathcal{M}$ is the set of visible points ϱ) over all frames in the current temporal window $\mathcal{W}$ by means of the following cost function

$\mathcal{A}(\mathbf{R}_i, \mathbf{\Gamma}_i)$:

$$\min_{\mathbf{R}_i,\mathbf{\Gamma}_i} \sum_{i=f-\mathcal{W}+1}^{f} \sum_{\varrho\in\mathcal{M}} \|\mathbf{W}_{i\varrho} - \mathbf{R}_i \left(\mathbf{S}_\varrho + \left(\mathbf{I}_3 \otimes \mathbf{\Gamma}_i^\top\right) \mathbf{\Psi}_\varrho\right) \|_{\mathcal{F}}^2$$
$$+ \lambda_q \sum_{i=f-\mathcal{W}+1}^{f} \|\mathbf{q}_i - \mathbf{q}_{i-1}\|_{\mathcal{F}}^2 + \lambda_\gamma \sum_{i=f-\mathcal{W}+1}^{f} \|\mathbf{\Gamma}_i^\top - \mathbf{\Gamma}_{i-1}^\top\|_{\mathcal{F}}^2 \quad (12)$$

where $\|\cdot\|_{\mathcal{F}}$ is the Frobenius norm and $\mathbf{R}_i$ are rotation matrices. These matrices $\mathbf{R}_i(\mathbf{q}_i)$ are parameterised using quaternions to guarantee orthonormality $\mathbf{R}_i\mathbf{R}_i^\top = \mathbf{I}_2$. We add temporal smoothness priors to penalize strong variations in both camera matrices $\mathbf{R}_i(\mathbf{q}_i)$ and weights $\mathbf{\Gamma}_i$. λ_q and λ_γ are regularisation weights determined empirically. The selected optimisation method is Levenberg-Marquardt.

Unlike other methods [6, 7] our formulation does not require all points to be tracked throughout the whole sequence. BA has the capability to deal with missing data since the cost function is evaluated only on visible points in each frame. To initialise the parameters for a new incoming frame, the camera pose is initialised as $\mathbf{R}_i = \mathbf{R}_{i-1}$ and the mode shape weights $\mathbf{\Gamma}_i$ are initialised assuming rigid motion $\mathbf{\Gamma}_i = \mathbf{\Gamma}_{i-1}$.

4.3. Estimation of the Shape at Rest

Our approach assumes that the shape at rest can be estimated similarly to [19, 9, 22, 2, 1], using a rigid factorization [15] on a few initial frames. We assume that the observed sequence contains some initial frames where the object is mostly rigid and does not deform substantially. Note that one of the challenges in sequential methods, including the rigid case [14], is the initialisation. Our shape at rest is a pair $\mathcal{S} = (\mathcal{N}, \mathcal{E})$ where $\mathcal{N} = (n_1, \ldots, n_p)$ is the set of 3D nodes and $\mathcal{E} = (e_1, \ldots, e_m)$ is the set of m elements. We use Delaunay triangulation to compute $\mathcal{E}$ after estimating $\mathbf{S}$ using the projection in the last rigid frame.

4.4. Computational Cost

Regarding the computational cost, two stages of the algorithm need to be considered:

Mode Shapes Computation (MSC). Mode shapes are computed before processing the video sequence. Two steps are needed. The first one is the inversion of $\mathbf{M}^{-1}\mathbf{K}$ to transform the generalized eigenproblem into a standard one. In the case of the lumped mass matrix in Eq. (5), the inverse computation cost is negligible, however in the case of the distributed mass Eq. (4) it requires the inversion of a $3p \times 3p$ matrix which is expensive even when the band-matrix pattern is exploited. As the accuracy of lumped vs. distributed mass is roughly the same, we always use lumped mass. The actual computation and assembly of both $\mathbf{K}$ and $\mathbf{M}$ has a $\mathcal{O}(p)$ complexity. The assembling cost is only significant for dense maps, where the process could be parallelized.

The second step is the computation of the mode shapes as eigenvectors. For computational efficiency, we propose to use orthogonal iteration with Ritz acceleration [12]. This returns the eigenvectors (mode shapes), sequentially, in ascending frequency order, hence the complexity scales with the number of computed modes. If the r mode shapes to be included in the basis correspond only to low frequencies (bending modes), the computation is efficient with a complexity scaling with $r + 4$. However, if both bending and stretching modes have to be computed, the complexity scales with $4 + r + p$, which can be expensive both in time and memory especially for dense meshes.

Non-linear Optimisation. The number of modes r is typically significantly smaller than the number of points in the scene shape p therefore the sliding window non-linear optimisation complexity is dominated by the computation of the Jacobian matrices, resulting in $\mathcal{O}\left(p\mathcal{W}^2r^2\right)$.

In conclusion, our method is sequential and can potentially achieve real-time performance at frame rate.

5. Experimental Results

In this section, we show experimental results on both synthetic and real sequences providing comparison with respect to state-of-the-art methods[1]. In all the experiments the error metric is defined as $e_{3\mathcal{D}} = \frac{1}{f}\sum_{i=1}^{f} \frac{\|\mathbf{S}_i - \mathbf{S}_i^{GT}\|_{\mathcal{F}}}{\|\mathbf{S}_i^{GT}\|_{\mathcal{F}}}$ where $\mathbf{S}_i$ is the 3D reconstruction and $\mathbf{S}_i^{GT}$ is the ground truth. Before computing this 3D error, the 3D reconstruction is aligned with the corresponding ground truth using Procrustes analysis over all frames.

5.1. Experiments on synthetic sequences

We propose two synthetic sequences of a deforming elastic plate with irregular and regular discretisation mesh respectively. Both sequences were generated with Abaqus[2] using material properties and nodal forces.

The method exhibits a nice accuracy vs. computational cost trade-off. Fig. 2 displays the consistent reduction of the error as more modes r are considered. As expected, the per frame BA time increases quadratically with the number of modes. Regarding the MSC computation time, it is defined by the number of nodes p, because the time is dominated by the computation and assembly of $\mathbf{K}$ and $\mathbf{M}$. Regarding the use of lumped mass vs. distributed mass, fig. 2 shows a negligible difference in accuracy. Since the computation time for the distributed mass can be quite high especially for a large number of nodes, we propose to always use the lumped mass model.

We compare our results against state-of-the-art NRS*f*M methods, both batch and sequential algorithms. For batch

[1] Videos of the experimental results can be found on website `http://webdiis.unizar.es/~aagudo`

[2] Tool to model continuum solid mechanics `http://www.3ds.com/products-services/simulia/portfolio/abaqus/overview/`

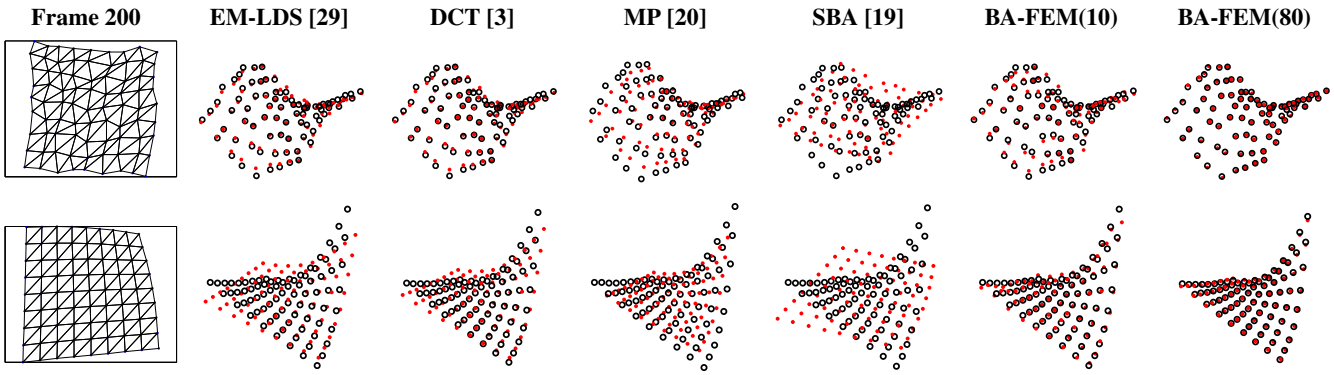

Figure 3. **Synthetic sequences.** Comparing BA-FEM using 10 and 80 bending mode shapes with respect to EM-LDS [29], DCT [3], MP [20] and SBA [19] for frame #200. Reconstructed 3D shape and the ground truth are showed with red dots and black circles respectively. **Top:** Results for synthetic sequence 1. **Bottom:** Results for synthetic sequence 2.

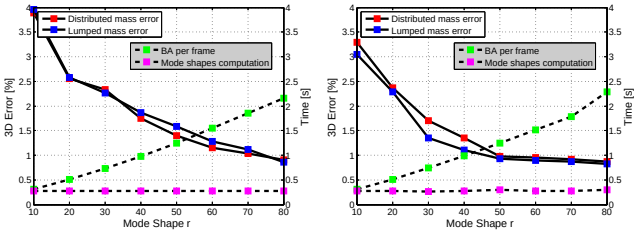

Figure 2. **Mean normalized error $e_{3\mathcal{D}}$ and run-time** with varying number of mode shapes for synthetic sequences. Both mass models are showed. Two scale plots: left-y axis mean normalized error $e_{3\mathcal{D}}$, right y-axis run-time in seconds **Left:** Results for synthetic sequence 1. **Right:** Results for synthetic sequence 2.

$e_{3\mathcal{D}}\%$	Syn. 1	Syn. 2	Batch	Sequential
EM-LDS	11.12(2)	10.92(2)	✓	
DCT	9.25(2)	11.81(5)	✓	
MP	12.42(2)	18.84(2)	✓	
SBA	14.03(16)	20.90(8)		✓
BA-FEM	3.89(10)	3.04(10)		✓
BA-FEM	0.86(80)	0.82(80)		✓

Table 1. **Quantitative comparison on synthetic sequences.** We show $e_{3\mathcal{D}}$ for EM-LDS [29], DCT [3], MP [20], SBA [19] and for our method BA-FEM using lumped mass with 10 and 80 mode shapes respectively. In all cases we have selected the number of shapes in the basis (in brackets) that gave the lowest $e_{3\mathcal{D}}$ error.

methods, we consider: EM-LDS [29], Metric Projections (MP) [20] and trajectory basis (DCT) [3]. We also compare with the Sequential BA (SBA) optimisation [19]. Table 1 summarizes the comparison the results for each method. The 3D reconstruction of a typical frame for each method is displayed in fig. 3. For our method we provide the results using two basis with $r = 10$ and $r = 80$ modes respectively. We can conclude that our method outperforms the state-of-the-art in terms of accuracy, with the additional advantage of being sequential. The low per frame computational cost in unoptimised Matlab code shows that the method can achieve real-time performance at frame rate.

We also use a synthetic dense face sequence with $p = 28{,}887$ points proposed in [10]. The scene is challenging for two reasons: the density of the data and the strong deformations that combine bending and stretching. If no boundary conditions are considered, the stretching modes appear in the $28{,}887^{th}$ position, making the computation unfeasible both because of computing time and memory storage. However, if boundary conditions are considered then the modes provided by the eigenvector method are reordered to firstly yield modes that combine both stretching and bending. Thanks to the ability of our method to handle boundary point priors, we can compute just the first $r = 5$ modes and experimentally show they are sufficient to encode the face deformations. In this case, boundary conditions correspond to face points that are considered to be rigid points by means of a connectivity analysis. We can achieve an $e_{3\mathcal{D}}$ error of 4.64% using our sequential approach, higher than the 2.60% for batch solution reported in [10] (see Fig. 4).

5.2. Experiments on real sequences

In this section, we evaluate our method using lumped mass on several existing datasets. Our performance is summarised in Table 2.

We use the sequence and tracks provided by [4] to report qualitative results on the actress sequence, which consists of 102 frames where an actress is talking and moving her head. In fig. 5, we show our 3D reconstruction obtained with $r = 10$ stretching mode shapes. Our results are comparatively similar to those reported in [19]. Similarly to [19], we use a rigid model on the first 30 frames to compute the rest shape.

We use the first 100 frames of a paper bending sequence proposed in [4] to display a qualitative evaluation of our method with respect to a significant fraction of missing data. We use the sparse tracking of 828 points obtained by dense tracking data reported in [11]. We process the sequence us-

Figure 4. **Face sequence.** Reconstruction of the dense face for selected frames: #30, #40, #79 and #95. **Top:** Ground truth 3D shapes. **Bottom:** Dense 3D reconstruction.

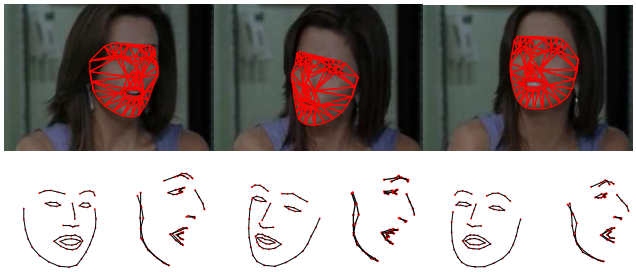

Figure 5. **Actress sequence. Top:** Selected frames #31, #48 and #84 with 3D reprojected mesh. **Bottom:** Original viewpoint and side views of the 3D reconstruction.

ing $r = 10$ bending mode shapes and 5 frames to compute the rest shape. Between frames #48 and #76, a 22% band of missing data simulating a strong self-occlusion is introduced. The performance does not degrade significantly (see fig. 6).

Finally, we evaluate our approach on the challenging deformations of a flag waving in the wind, proposed in [11], to show the performance of our method in the case of a dense map of $p = 9{,}622$ points. We have included a few initial frames corresponding to the camera observing the rigid shape. Fig. 7 shows comparison with respect to ground-truth. As the deformation contains little stretching, the first bending mode shapes can encode accurately the deforming scene. The trade-off accuracy vs. computational cost is displayed in fig. 7 (right). When the modes to be computed are only the first r, far less than the $3p$ total number of modes, then the MRC computing time is low.

6. Conclusions and Future Work

We have proposed to exploit the well-known FEM modal analysis to compute the mode shapes for sequential NRS*f*M. The resulting basis, combined with smoothness constraints without using additional distance constraints such as inextensibility, provides a competitive solution to the accuracy vs. per-frame computation time balance. For demanding deformations, such as the synthetic cases in Sec. 5.1, adding modes results in significant error reduction, whereas in quasi-isometric deformations like those displayed by the flag, most of the deformation can be explained with just a few modes. Regarding computational performance, some time has to be invested off-line to compute the mode basis. After this frame-to-frame real-time performance is possible. All our claims have been experimentally validated both on synthetic and real sequences showing a performance better or comparable to the state-of-the-art, with the additional advantage of our method being sequential, accurate and scalable. Our future work is oriented towards experimental validation on medical imaging, where accurate FEM biomechanical models are available. We expect that our method will be able to exploit the rich priors available and provide new avenues of research for the challenging use of NRS*f*M in medical applications.

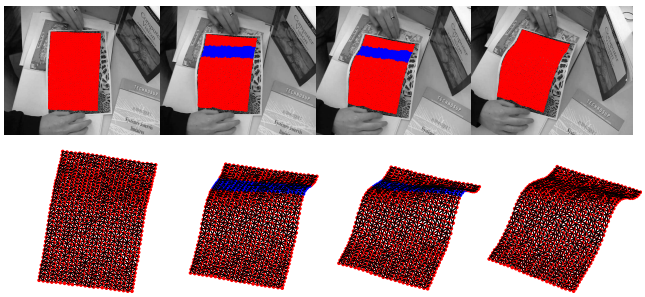

Figure 6. **Robustness with respect to self-occlusion. Top:** Selected frames (#25, #50, #75 and #100) with 3D reprojected mesh. Up to a 22% structured occlusion is simulated. **Bottom:** General view of the 3D reconstruction deformed scene. Blue points correspond to the structured occlusion.

	Size		Run-time (sec)		Error
Sequence	p	r	MSC	BA	e_{3D}%
Syn. 1	81	10-80	0.3	0.3-2.2	3.89-0.86
Syn. 2	81	10-80	0.3	0.3-2.2	3.04-0.82
Actress [4]	68	10	0.4	0.3	-
Bending [4]	828	10	60	0.85	-
Flag [11]	9,622	5-25	300	47-416	4.14-3.29
Face [10]	28,887	5	1540	60	4.64

Table 2. **Experiments summary.** The problem size is defined by the number of map nodes p and the mode shapes number r. Regarding computing time, both the MSC and the BA per frame time are displayed. For all cases, the sliding window size is $\mathcal{W} = 5$.

Acknowledgments

This work was partly funded by the MINECO projects DIP2012-32168 and DPI2011-27939-C02-01; by the ERC Starting Grant agreement 204871-HUMANIS; and by a scholarship FPU12/04886. The authors wish to thank R. Garg and M. Paladini for fruitful discussions.

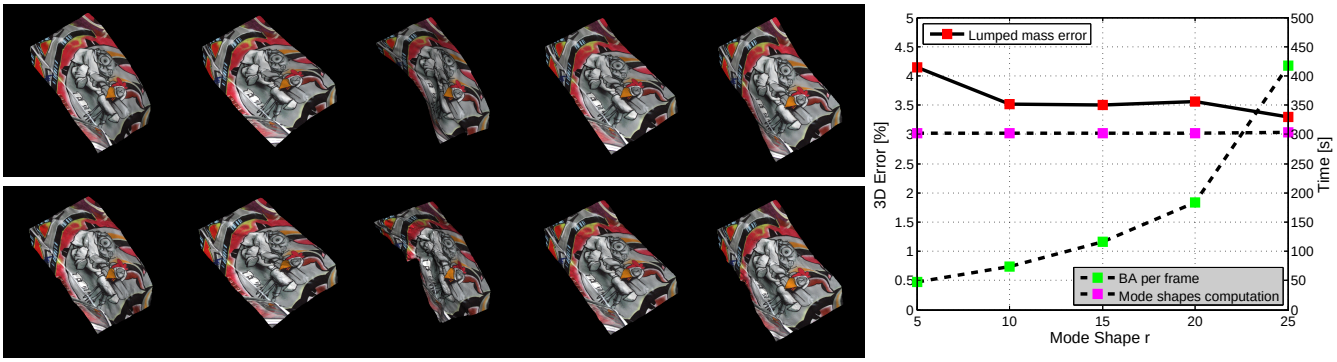

Figure 7. **Flag sequence.** Reconstruction of the dense 9,622 point flag at five selected frames: #10, #20, #30, #40 and #50. **Left:** Top: Rendered ground truth. Bottom: Rendered reconstruction. **Right:** Mean normalized error and run-time with varying modes number.

References

[1] A. Agudo, B. Calvo, and J. M. M. Montiel. 3D reconstruction of non-rigid surfaces in real-time using wedge elements. In *Workshop on NORDIA*, 2012.

[2] A. Agudo, B. Calvo, and J. M. M. Montiel. Finite element based sequential bayesian non-rigid structure from motion. In *CVPR*, 2012.

[3] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from motion. *TPAMI*, 33(7):1442–1456, 2011.

[4] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd. Coarse-to-fine low-rank structure-from-motion. In *CVPR*, 2008.

[5] K. J. Bathe. *Finite element procedures in Engineering Analysis*. Prentice-Hall, 1982.

[6] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *CVPR*, 2000.

[7] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure from motion factorization. In *CVPR*, 2012.

[8] A. Del Bue, X. Llado, and L. Agapito. Non-rigid metric shape and motion recovery from uncalibrated images using priors. In *CVPR*, 2006.

[9] J. Fayad, L. Agapito, and A. Del Bue. Piecewise quadratic reconstruction of non-rigid surfaces from monocular sequences. In *ECCV*, 2010.

[10] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *CVPR*, 2013.

[11] R. Garg, A. Roussos, and L. Agapito. A variational approach to video registration with subspace constraints. *IJCV*, 104(3):286–314, 2013.

[12] G. Golub and C. Van Loan. *Matrix computations*. Johns Hopkins Univ Pr, 1996.

[13] P. F. U. Gotardo and A. M. Martinez. Computing smooth time-trajectories for camera and deformable shape in structure from motion with occlusion. *TPAMI*, 33(10):2051–2065, 2011.

[14] G. Klein and D. W. Murray. Parallel tracking and mapping for small AR workspaces. In *ISMAR*, 2007.

[15] M. Marques and J. Costeira. Optimal shape from estimation with missing and degenerate data. In *WMVC*, 2008.

[16] F. Moreno-Noguer and J. M. Porta. Probabilistic simultaneous pose and non-rigid shape recovery. In *CVPR*, 2011.

[17] C. Nastar and N. Ayache. Fast segmentation, tracking and analysis of deformable objects. In *ICCV*, 1993.

[18] R. Newcome and A. J. Davison. Live dense reconstruction with a single moving camera. In *CVPR*, 2010.

[19] M. Paladini, A. Bartoli, and L. Agapito. Sequential non rigid structure from motion with the 3D implicit low rank shape model. In *ECCV*, 2010.

[20] M. Paladini, A. D. Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito. Factorization for non-rigid and articulated structure using metric projections. In *CVPR*, 2009.

[21] A. Pentland and B. Horowitz. Recovery of nonrigid motion and structure. *TPAMI*, 13(7):730–742, 1991.

[22] C. Russell, J. Fayad, and L. Agapito. Energy based multiple model fitting for non-rigid structure from motion. In *CVPR*, 2011.

[23] M. Salzmann and P. Fua. Reconstructing sharply folding surfaces: A convex formulation. In *CVPR*, 2009.

[24] M. Salzmann, R. Urtasun, and P. Fua. Local deformation models for monocular 3D shape recovery. In *CVPR*, 2008.

[25] S. Sclaroff and A. Pentland. Modal matching for correspondence and recognition. *TPAMI*, 17(6):545–561, 1995.

[26] H. Tao and T. S. Huang. Connected vibrations: A modal analysis approach for non-rigid motion tracking. In *CVPR*, 1998.

[27] J. Taylor, A. D. Jepson, and K. N. Kutulakos. Non-rigid structure from locally-rigid motion. In *CVPR*, 2010.

[28] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *IJCV*, 9(2):137–154, 1992.

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

[30] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Vision Algorithms: Theory and Practice*, pages 298–372, 2000.

[31] A. Varol, M. Salzmann, E. Tola, and P. Fua. Template-free monocular reconstruction of deformable surfaces. In *ICCV*, 2009.

[32] S. Vicente and L. Agapito. Soft inextensibility constraints for template-free non-rigid reconstruction. In *ECCV*, 2012.