3D Reconstruction of Non-Rigid Surfaces in Real-Time using Wedge Elements [*]

Antonio Agudo, Begoña Calvo, and J. M. M. Montiel

Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, Spain
{aagudo,bcalvo,josemari}@unizar.es

Abstract. We present a new FEM (Finite Element Method) model for the 3D reconstruction of a deforming scene using as sole input a calibrated video sequence. Our approach extends the recently proposed 2D thin-plate FEM+EKF (Extended Kalman Filter) combination. Thin-plate FEM is an approximation that models a deforming 3D thin solid as a surface, and then discretizes the surface as a mesh of planar triangles. In contrast, we propose a full-fledged 3D FEM formulation where the deforming 3D solid is discretized as a mesh of 3D wedge elements. The new 3D FEM formulation provides better conditioning for the rank analysis stage necessary to remove the rigid boundary points from the formulation. We show how the proposed formulation accurately estimates deformable scenes from real imagery even for strong deformations. Crucially we also show, for the first time to the best of our knowledge, NRSfM (Non-Rigid Structure from Motion) at 30Hz real-time over real imagery. Real-time can be achieved for our 3D FEM formulation combined with an EKF resulting in accurate estimates even for small size maps.

1 Introduction

The FEM+EKF combination —Finite Element Method to code the deformation model, which is embedded within an Extended Kalman Filter Bayesian estimator— has proven to be able to provide a sequential solution for Non-Rigid Structure from Motion (NRSfM). Agudo *et al.* [1] propose a simplified 2D thin-plate FEM to code the scene as a deformable surface. In contrast, we propose a full-fledged 3D FEM formulation.

General FEM methods have to discretize a solid into elements defined by the 3D location of their nodes. In our proposal, the elements are wedge-shaped defined by 6 nodes. As the observed solids are opaque, we have to tackle the difficulty of only observing one face of the solid. We propose to extrapolate the solid structure just from its visible face by extrusion. The visible side is modelled as a mesh of triangles, and the mesh is extruded in the normal direction up to a predefined thickness h. The nodes on the hidden side cannot be observed but they define the deformation model. In order to consider their effect, they have to be marginalized out by means of the Schur's complement.

The stiffness matrix K is the key to the FEM formulation. Theoretically, K's rank deficiency is 6 because displacements corresponding to rigid body motion [2] do not result in deformation, and hence can happen if no forces are acting on the solid. The thin-plate formulation, due to its approximate nature, does not result in K with 6 rank deficiency [1], so it has to be forced by means of Singular Value Decomposition (SVD). In contrast, our 3D FEM formulation produces K matrices which are 6 rank deficient, showing its theoretical superiority. The 3D FEM can be easier to combine with other FEM models available for the observed structure, such as biomechanical models which are often modelled using 3D elements.

In our approach, all the images are processed, since the estimates are so close a rather simple low cost FEM deformation model is valid. The overall result is that 3D FEM+EKF is able to achieve real-time performance as is experimentally shown. It is also our contribution to provide the first —to the best of our knowledge— experimental validation of a NRSfM system performing in real-time at 30Hz frame-rate using real image sequence.

2   Related Work

NRSfM computes deformable 3D structures from a sequence of images acquired with a monocular camera. It is an ill-posed problem, so additional smoothing constraints or priors are necessary. A relevant class of NRSfM methods are based on closed form factorization, the time varying 3D structure is coded as a linear combination of basis shapes as proposed by Bregler *et al.* [3] for orthographic cameras. Paladini *et al.* [4] propose the first sequential approach for solving NRSfM based on a sequential version of the factorization method over a sliding window. The authors do not report real-time performance.

Bundle Adjustment (BA) has been applied also to solve shape basis approaches to NRSfM. BA can additionally incorporate temporal and spatial smoothness priors both on the deformations and motion [5, 6]. Torresani *et al.* [7] introduce an expectation maximization probabilistic linear dynamic model coding deformation weight as Gaussians. Reported experiments, compared with respect to closed form, exhibit better noise rejection and improved accuracy.

In contrast to global models, the piecewise modelling, proposed by Varol *et al.* [8], can more accurately code strong deformations by means of a composition of multiple local deformations. Piecewise methods rely on common features shared between patches to enforce spatial consistency and create a continuous global surface. Taylor *et al.* [9] propose a triangle soup assuming as rigid each triplet of neighbour points. Fayad *et al.* [10] propose quadratic models for each patch.

Template-based methods assume that 2D-to-3D correspondences can be established with a reference image in which the shape is known a priori. These approaches assume isometric deformation for the structure [11]. They have been extended to non-isometric deformations in [12].

Most of the NRSfM methods assume orthographic cameras. Regarding data association can tolerate partial observation of scene points, in any case, the

reported experiments rely on given data association instead of embedding the matching within the sequence processing.

FEM was applied in monocular computer vision by Ilić and Fua [13], where an expensive non-linear model accurate for large deformations focused on 1D beam like structures was introduced. The formulation includes the forces and needs to identify boundary points in the image, resulting in a robust and accurate tracking method. In [14] a physical elastic solid model is coded by means of the boundary element method for 2D deformable object tracking, it is not necessary to identify boundary points in the image because 3 rank deficiency is enforced. Agudo *et al.* in [15] extend the EKF-SLAM method [16] to non-rigid scenes, they propose 2D thin-plate FEM model to embed Navier's equations within the EKF estimation. The method can combine both rigid and non-rigid points to estimate 3D reconstruction of deformable scenes and 3D camera trajectory using full perspective cameras. However, prior knowledge about scene point classification as boundary or non-boundary is mandatory. Recently in [1], the boundary points identification prior has been removed, resulting in a method comparable to other NRSfM approaches. FEM models can cope both with isometric and non-isometric scenes without assuming any scene deformation mode. The method is piecewise quadratic and can be considered close to template-based because the shape at rest —not necessarily planar— is needed before coping with the deforming scene. Template-based methods register every image with respect to the initial 3D template, while FEM+EKF sequentially compares the current frame with respect to the 3D scene structure estimated at the previous step.

3 3D FEM Formulation

The partial differential equations for the 3D elasticity problem, known as Navier's equations [2], model the deformation of the solid Ω, with a boundary Γ, under 3D external forces. These equations can be expressed in terms of displacements as:

$$\frac{E}{2(1+\nu)(1-2\nu)}\nabla(\nabla\cdot\boldsymbol{u}) + \frac{E}{2(1+\nu)}\nabla^2\boldsymbol{u} + \boldsymbol{F} = \boldsymbol{0}\ \ in\ \ \Omega, \tag{1}$$

with $\boldsymbol{u}$ the 3D displacement vector, $\boldsymbol{F}$ the body force applied to the solid, and the material properties of an isotropic linearly elastic solid: the Young's modulus E and the Poisson's ratio ν. $\nabla\cdot(\bullet)$ is the divergence operator, $\nabla(\bullet)$ represents the gradient operator and $\nabla^2(\bullet)$ represents the Laplacian operator. The boundary conditions for Eq. (1) can be expressed as a displacement vector, $\boldsymbol{u} = \bar{\boldsymbol{u}}$, over the boundary Γ_u.

The thin-plate formulation is based on a 2D version of Eq. (1) to model forces and displacements within the plane (membrane effect), and on the Kirchhoff plate to model forces and displacements off the plane (bending effect). The discretized scene elements Ω_e are planar triangles [15, 1]. In contrast, our 3D FEM formulation uses Eq. (1) directly to model the solid as a single layer 3D wedge element. The continuous displacements are approximated as $\boldsymbol{u} = \mathbf{N}\boldsymbol{a}$, being $\mathbf{N}$ the matrix of shape functions, then Eq. (1) is rewritten as a sparse linear

system:

$$\boldsymbol{K}\,\boldsymbol{a} = \boldsymbol{f}\,, \tag{2}$$

where $\boldsymbol{K}$ is the global stiffness matrix, $\boldsymbol{a}$ is the 3D nodal displacements vector and $\boldsymbol{f}$ is the 3D nodal forces vector. The global $\boldsymbol{K}$ results from the assembling of elemental stiffness matrices $\boldsymbol{K}^e$. The elemental matrix $\boldsymbol{K}^e$, in the normalized element, in natural coordinates (ξ, η, ζ), is computed as:

$$\boldsymbol{K}^e = \frac{1}{E} \int_{-1}^{1} \int_{0}^{1} \int_{0}^{1-\xi} \mathbf{B}^\top \mathbf{D} \mathbf{B}\,|\mathbf{J}|\;d\eta\,d\xi\,d\zeta\,, \tag{3}$$

where $\mathbf{B}$ is the strain-displacement matrix for a 3D bilinear wedge element, it depends on the shape functions $\mathbf{N}$. $\mathbf{D}$ is the constitutive matrix which encodes the material properties [2]. $\mathbf{J}$ is the transformation from the normalized to the actual wedge element geometry. E is factorized out of the stiffness matrix to ease the tuning (see Sec. 4). The integrals in Eq. (3) are computed numerically [17].

3.1   3D Wedge Element Definition

We have proposed to model the solid as a single layer of wedge shaped elements. Being an opaque solid, the camera can only detect one surface side, we will call it the visible side. We have to generate the FEM mesh from only the visible side nodes. We propose to extrude each visible side node i along its normal unit vector $\mathbf{d}_i$, up to a fixed thickness h (see Fig. 1(left)). The normal is estimated as a weighted average of the normals of the mesh triangles having the considered node $\mathbf{y}_i$ as a vertex (see Fig. 1(right)). Each normal vector weights according to the corresponding triangle area computed by means of the cross product:

$$\mathbf{d}_i = \frac{\sum_{j=1}^{k}(\mathbf{m}_{ij} - \mathbf{y}_i) \times (\mathbf{m}_{ij+1} - \mathbf{y}_i)}{\left\|\sum_{j=1}^{k}(\mathbf{m}_{ij} - \mathbf{y}_i) \times (\mathbf{m}_{ij+1} - \mathbf{y}_i)\right\|}\,, \tag{4}$$

where $\mathbf{m}_{ij} \in \{\mathbf{m}_{i1}, \mathbf{m}_{i2}, \dots, \mathbf{m}_{ik}\}$ are the neighbour nodes defining the k Delaunay triangles to which the node $\mathbf{y}_i$ belongs to.

3.2   FEM Linear System Solution

As we assume a single layer of wedge elements, the displacements and forces in linear system Eq. (2) can be reordered to group those corresponding to the visible side, $\boldsymbol{a}^v$, $\boldsymbol{f}^v$, and those corresponding to the hidden side $\boldsymbol{a}^h$, $\boldsymbol{f}^h$:

$$\begin{bmatrix} \boldsymbol{K}_{vv} & \boldsymbol{K}_{vh} \\ \boldsymbol{K}_{vh}^\top & \boldsymbol{K}_{hh} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}^v \\ \boldsymbol{a}^h \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}^v \\ \boldsymbol{f}^h \end{bmatrix}. \tag{5}$$

Forces might be acting on both solid faces, however without loss of generality, we can reduce any force acting on the hidden side to an equivalent one on the visible side, then we can safely assume zero forces acting on the hidden side, $\boldsymbol{f}^h = \mathbf{0}$. Hence Schur's complement can be applied to relate $\boldsymbol{a}^v$, $\boldsymbol{f}^v$:
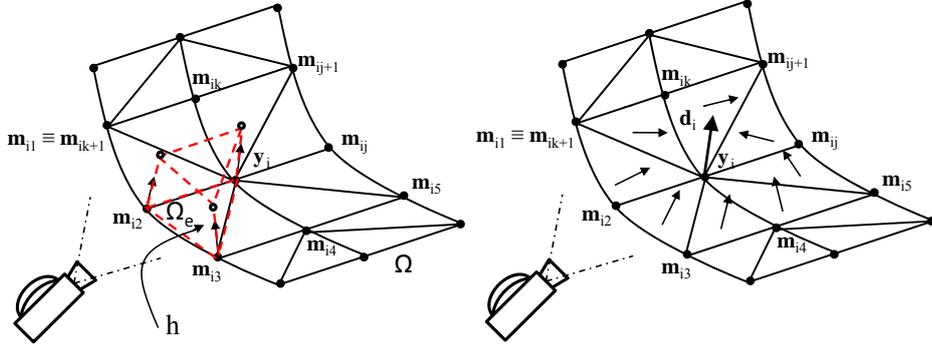
Fig. 1. Left: Extruded wedge element. **Right:** Extrusion normal unit vector $\mathbf{d}_i$, estimated as a weighted average of the normals of the mesh triangles having the considered node $\mathbf{y}_i$ as a vertex.

$$\left[\boldsymbol{K}_{vv} - \boldsymbol{K}_{vh}\boldsymbol{K}_{hh}^{-1}\boldsymbol{K}_{vh}^{\top}\right]\boldsymbol{a}^v = \boldsymbol{f}^v \ . \tag{6}$$

In the absence of boundary conditions, the previous linear system is under-constrained. The full affine solution space can be computed as [1, 18]:

$$\boldsymbol{a}^v = \boldsymbol{a}_h + \boldsymbol{a}_p, \qquad \mathcal{K}\,\boldsymbol{a}_h = \boldsymbol{0}, \qquad \boldsymbol{a}_p = \mathcal{K}^+\,\boldsymbol{f}^v, \tag{7}$$

with $\mathcal{K} = \boldsymbol{K}_{vv} - \boldsymbol{K}_{vh}\boldsymbol{K}_{hh}^{-1}\boldsymbol{K}_{vh}^{\top}$. We have verified that $\boldsymbol{K}$ has 6 null singular values as it is theoretically expected. However, $\mathcal{K}$ has 6 singular values significantly smaller than the rest but only 3 of them can be considered null up to numerical accuracy. We attribute this deviation from the theoretical value to the matrix inversion included in the Schur's complement that introduces numerical round-off errors. The particular solution $\boldsymbol{a}_p$ is computed by means of the Moore–Penrose pseudoinverse $\mathcal{K}^+$ enforcing the matrix $\mathcal{K}$ rank to $3p - 6$, with p the number of nodes.

4   Sequential Approach to NRSfM

The 3D FEM formulation is embedded within the EKF sequential Bayesian estimation cycle. We use a cameracentric EKF-based formulation to estimate the relative pose of the deformable scene with respect to the camera. The 3D FEM formulation free of rigid boundary points permits the displacement of all map points in correlated form by means of the pseudoinverse stiffness matrix $\mathcal{K}^+$ and it is embedded in the EKF prediction step.

The state vector $\mathbf{x}_k^{C_k} = \left(\mathbf{x}_{rk}^{C_k\top}, \mathbf{y}_k^{C_k\top}\right)^{\top}$ is composed of the camera pose $\mathbf{x}_{rk}^{C_k}$ and the locations of the p map points $\mathbf{y}_k^{C_k} = \left(\mathbf{y}_{1k}^{C_k\top}, \dots, \mathbf{y}_{pk}^{C_k\top}\right)^{\top}$, coded as cameracentric, i.e. all of them are referenced with respect to the camera frame

(C_k superindex). $\mathbf{x}_{rk}^{C_k}$ is composed of the pose and the corresponding velocity vectors. The dynamics of the pose includes the combined effects of both the camera and scene rigid motions [1] it is assumed to be a constant velocity model.

The state equation for the deformable scene map points is:

$$\mathbf{g}_y = \mathbf{y}_{k+1}^{C_k} = \mathbf{y}_k^{C_k} + \mathcal{K}_k^+ \Delta \mathbf{S}^C, \tag{8}$$

where the incremental non-rigid displacement is modelled as the particular solution for the FEM linear system Eq. (7). The pseudoinverse stiffness matrix $\mathcal{K}_k^+ \left(\hat{\mathbf{y}}_{k-1|k-1}^{C_k} \right)$ depends on the current structure geometry estimate.

The vector of normalized forces $\Delta \mathbf{S}^C$ causes recursively an incremental deformation at each step. We assume $\Delta \mathbf{S}^C$ follows a Gaussian with zero-mean, with $\mathbf{Q_y}$ its covariance. The normalized forces are defined as:

$$\Delta \mathbf{S}_i^C = \frac{1}{E} \left(\Delta f_{xi}^C, \Delta f_{yi}^C, \Delta f_{zi}^C \right)^\top, \tag{9}$$

where E has been factorized out from the stiffness matrix in order to allocate the material tuning parameters within the state noise. In contrast to [1], h thickness parameter cannot be factored out from $\boldsymbol{K}$ and hence cannot be allocated within the normalized forces.

We propose to tune $\mathbf{Q_y}_i$ as a diagonal matrix, where the standard deviation codes the normalized forces magnitude. The scene is coded as rigid when $\mathbf{Q_y}_i = \mathbf{0}$. Additionally, h and ν have to be defined so as to compute the $\boldsymbol{K}^e$ in Eq. (3). Quasi-incompressible materials are assumed tuning $\nu = 0.499$.

5   Experiments

A Matlab code has been developed to validate the proposed 3D FEM algorithm on real 320×240@30Hz image sequences. The first experiment corresponds to a non-isometrically deforming silicone. Sequence and ground truth taken from [1]. A second experiment corresponds to a waving camera observing an isometrically deforming paper sheet.

We have also developed an optimised C++ code to show the 3D FEM+EKF real-time performance. We process the paper sequences for a 35 feature map to reach real-time at 30Hz on an Intel Core i7 processor at 2GHz based laptop.

For all experiments[1], we assume that the initial sequence frames correspond to a mobile camera observing a rigid scene in order to estimate the structure at rest. We use the proposed method just tuning $\mathbf{Q}_y = \mathbf{0}$ (Sec. 4) to estimate 3D structure at rest.

5.1   Multiply Deformed Silicone Sequence

Data association was automatically computed applying the general matching algorithm. The coded labels for the markers are not considered for the data

---

[1] Videos of the experimental results on website `http://webdiis.unizar.es/~aagudo`
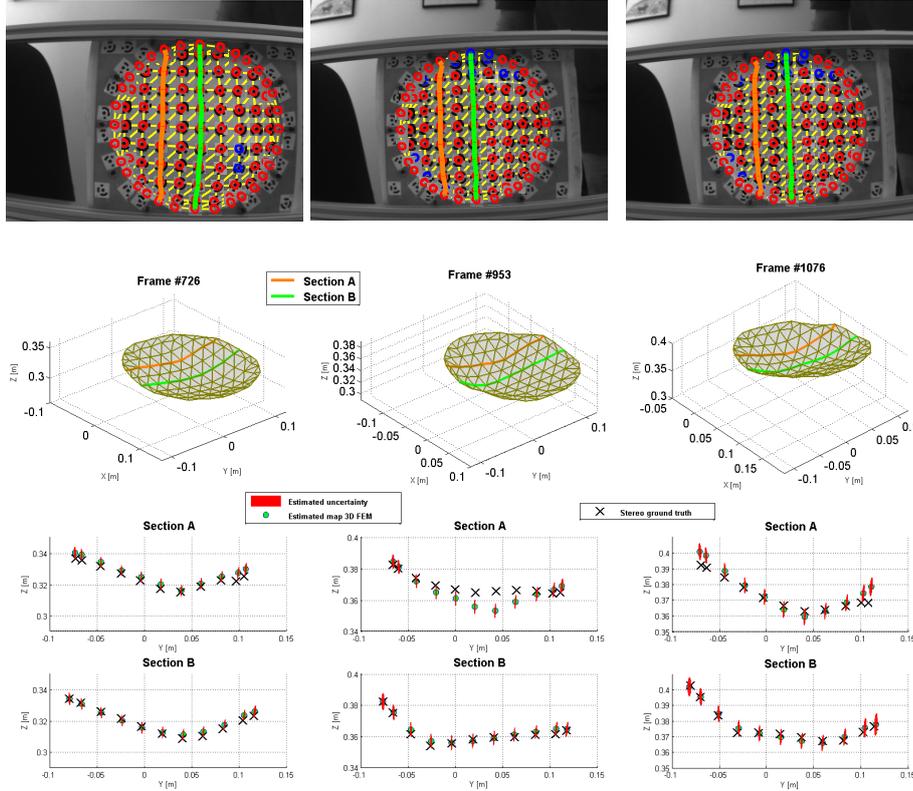
Fig. 2. 3D FEM estimated structure for the multiply deformed silicone. **Top:** Images with overlaid deformed 3D mesh and elliptical matching acceptance regions: red (predicted&matched), blue (predicted&not matched). **Middle:** General view of the 3D reconstructed deformed scene. **Bottom:** Two cross sections of the reconstructed surface; these represent the estimated points with the 95% acceptance regions according to the estimated covariances, and the stereo ground truth.

association algorithm. Not all the points were detected in all the images. In the video `silicone_sequence.avi`, it can be noticed how non-detection results in map point covariance increase. In any case, non-matched map point locations are updated indirectly from other map point observations via the covariance matrix correlation terms.

The estimated structure provides both the map point locations and the corresponding covariance. The covariance codes the estimation error. The ground truth is included in the 95% acceptance region for the estimated map points (see Fig. 2). We can conclude after a quantitative comparison that the proposed method is accurate, because the ellipses are small, and mostly exact, because the ellipses include the ground truth. It should be noticed that the processed scene does not fulfill the isometric deformation but we can handle it successfully.
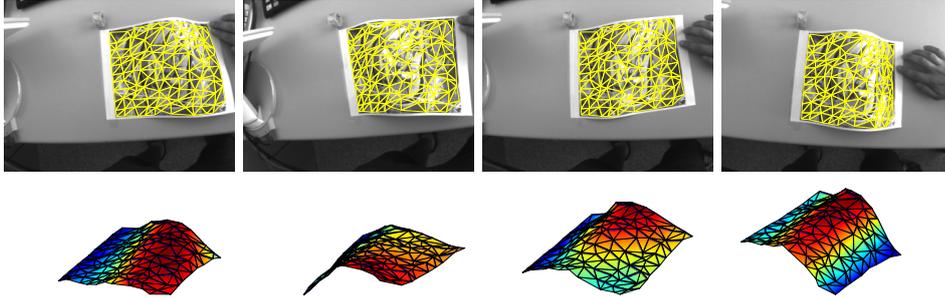
Fig. 3. Estimated structure for the **strong deformation paper**. **Top:** Deformed 3D mesh overlaid on the image sequence. **Bottom:** General view of the 3D deformable surface.

5.2   Deforming Paper Sheet

To display performance under extreme deformations, we used a real image sequence in which there is a strong deformation of a paper sheet observed by a waving camera (see Fig. 3 and `strong_deformation.avi`). FAST interest points in the first image define the map points, the triangular mesh is just the corresponding Delaunay triangulation. Every triangle yields a different stiffness because of its different geometry.

Fig. 3 shows selected frames and the corresponding estimated structure. The tuning, very thin solid and quasi-incompressible material, results in the ability to deal with an isometric scene as a consequence of the generality of our proposal.

5.3   Real-Time Performance

We have implemented the algorithm in C++. Additionally, we have implemented a specialized algorithm for symmetric matrix SVD, to reduce the computational cost from $O\left(21n^3\right)$ to $O\left(12n^3\right)$. We tested the code on the paper sequences (see Fig. 4 and `real_time_sequence.avi`). Given the $O\left(n^3\right)$ complexity, we have limited the map size to 35 features so as to achieve real-time rates at 30Hz. Thus, we show experimentally that the 3D FEM+EKF combination can produce accurate estimates even with small size maps. Note that other methods in the literature use bigger maps, about a hundred points for FEM+EKF in [1], or several hundreds for batch methods in [10]. Taylor *et al.*'s method [9] is only valid for bigger maps. In addition, we must emphasize that our proposal does not use as input the 2D tracking data, but it is computed automatically.

Fig. 5 shows the total cycle time budget identifying: map management and matching, EKF update and EKF prediction. Two phases can be distinguished: rigid initialization and non-rigid estimation. At initialization the map points are coded in inverse depth [16] (6 parameters per point), the prediction cost is negligible and the dominant cost is the cubic update. During the non-rigid estimation map points are coded in Euclidean XYZ (3 parameters per point)
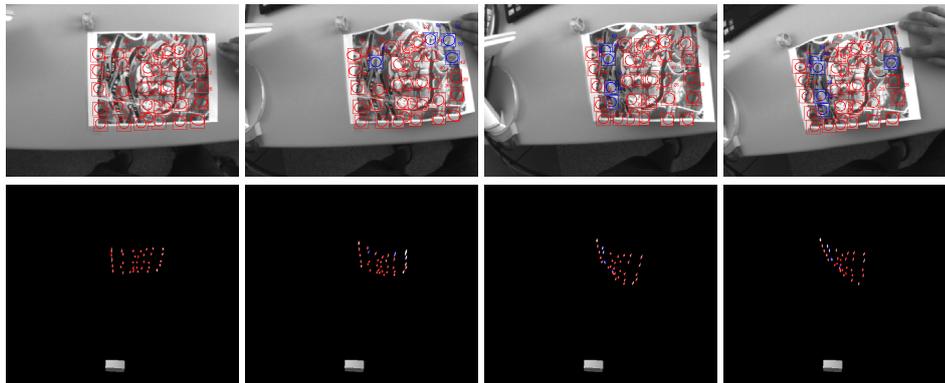
Fig. 4. Estimated structure for the **strong deformation in real-time** at 30Hz. **Top:** Images including elliptical search regions: red (predicted&matched) and blue (predicted&not matched). **Bottom:** General view of the 3D reconstructed deformed scene.

with the corresponding reduction in the update stage. However, due to the FEM modelling, the prediction stage becomes relevant due to its cubic cost. An increase in the matching during the non-rigid stage is due to the increase in the prediction ellipse size with the corresponding increase in the search time.

6   Conclusions

The 3D FEM+EKF combination has been experimentally validated —for the first time to the best of our knowledge— to be a valid method for sequential real-time at frame-rate NRSfM. It has to be stressed that the 3D FEM+EKF formulation deals with full perspective cameras, and it does not assume prior data association because it is computed within the EKF prediction-match-update cycle. NRSfM can be solved for a low number of scene points, what can be comparatively advantageous for scenes with low texture content.

To reduce the effect of the non-linearities, we are forced to process every single frame in the sequence. In any case, as the EKF handles a dynamic state, the state size does not increase with the number of processed images because we do not have to explicitly represent the scene at previous time steps. All the estimation memory is accumulated in the actual scene estimate and its corresponding covariance matrix. We believe that this is one of key elements to achieve real-time performance.

A new FEM model based on 3D wedge elements have been developed. The novel model is theoretically superior to previous thin-plate FEM formulation [1]. FEM-based NRSfM can be rather effective in considering the rich priors that accurate FEM models can provide. This is particularly relevant for the case of medical images where it is realistic to have available accurate FEM models of the observed scene. This is the goal of our future work.
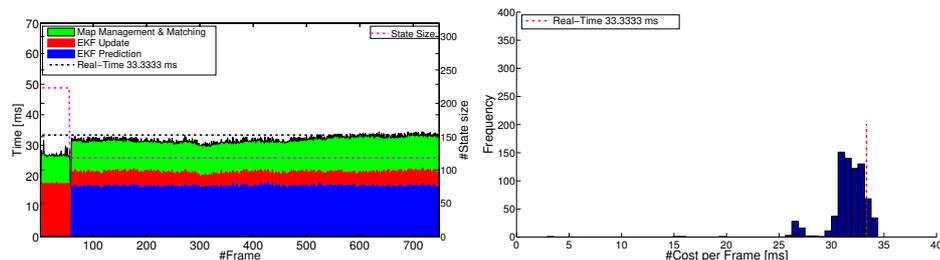
Fig. 5. Left: Real-time computation budget for the strong deformation sequence. Rigid initialization until frame#60 when non-rigid estimation starts. Two scale plots: left-y axis time, right y-axis state size n. **Right:** Per cycle time histogram.

References

1. Agudo, A., Calvo, B., Montiel, J.M.M.: Finite element based sequential bayesian non-rigid structure from motion. In: CVPR. (2012)
2. Zienkiewicz, O.C., Taylor, R.L.: The finite element method. Vol. 1: Basic formulation and linear problems. McGraw-Hill, London (1989)
3. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3D shape from image streams. In: CVPR. (2000)
4. Paladini, M., Bartoli, A., Agapito, L.: Sequential non rigid structure from motion with the 3D implicit low rank shape model. In: ECCV. (2010)
5. Bartoli, A., Gay-Bellile, V., Castellani, U., Peyras, J., Olsen, S., Sayd, P.: Coarse-to-fine low-rank structure-from-motion. In: CVPR. (2008)
6. Del Bue, A., Llado, X., Agapito, L.: Non-rigid metric shape and motion recovery from uncalibrated images using priors. In: CVPR. (2006)
7. Torresani, L., Hertzmann, A., Bregler, C.: Nonrigid structure-from motion: estimating shape and motion with hierarchical priors. PAMI (2008) 878–892
8. Varol, A., Salzmann, M., Tola, E., Fua, P.: Template-free monocular reconstruction of deformable surfaces. In: ICCV. (2009)
9. Taylor, J., Jepson, A.D., Kutulakos, K.N.: Non-rigid structure from locally-rigid motion. In: CVPR. (2010)
10. Fayad, J., Agapito, L., Del Bue, A.: Piecewise quadratic reconstruction of non-rigid surfaces from monocular sequences. In: ECCV. (2010)
11. Salzmann, M., Fua, P.: Reconstructing sharply folding surfaces: A convex formulation. In: CVPR. (2009)
12. Moreno-Noguer, F., Salzmann, M., Lepetit, V., Fua, R.: Capturing 3D stretchable surfaces from single images in closed from. In: CVPR. (2009)
13. Ilić, S., Fua, P.: Non-linear beam model for tracking large deformation. In: ICCV. (2007)
14. Greminger, M.A., Nelson, B.J.: Deformable object tracking using the boundary element method. In: CVPR. (2003)
15. Agudo, A., Calvo, B., Montiel, J.M.M.: FEM models to code non-rigid EKF monocular SLAM. In: Workshop on Dynamic Shape Capture and Analysis. (2011)
16. Civera, J., Davison, A.J., Montiel, J.M.M.: Inverse depth parametrization for monocular SLAM. IEEE Transactions on Robotics **24** (2008) 932–945
17. Stroud, A.H.: Approximate Calculation of Multiple Integrals. Prentice-Hall (1971)
18. Golub, G., Van Loan, C.: Matrix computations. Johns Hopkins Univ Pr (1996)