FEM Models to Code Non-Rigid EKF Monocular SLAM

Antonio Agudo
RoPeRT-I3A
Universidad de Zaragoza, Spain
aagudo@unizar.es

Begoña Calvo
GEMM-I3A
Universidad de Zaragoza, Spain
bcalvo@unizar.es

J. M. M. Montiel
RoPeRT-I3A
Universidad de Zaragoza, Spain
josemari@unizar.es

Abstract

Bayesian EKF (Extended Kalman Filter) is cross-fertilized with Navier's equations solid deformation modeling to compute 3D non-rigid structure from monocular camera motion. The method operates with a projective camera and autonomously computes –for every sequence frame– both the geometry and the matches. The combination results in an sequential efficient method to code the rich available physical priors, particularly relevant for medical monocular endoscope sequences such as those observing the abdominal cavity.

The scene is modeled as a planar triangular mesh where each triangular element is modeled as a thin-plate. Navier's equations are solved numerically by means of FEM (Finite Element Method), being the FEM nodes the 3D points of the estimated sparse structure. Despite the assumed elastic model is only valid for small deformations, the eventually large scene deformation is accurately computed, because the EKF extracts partial measurements of the actual scene deformation at frame rate.

Ground truth is computed from a real sequence gathered with hand-held stereo camera. The observed non-rigid scene is a silicone cloth fixed on a stretcher. It is deformed under the action of an unknown force applied on the silicone surface. It is shown how the estimation resulting from applying the proposed algorithm a monocular sequence is statistically consistent with the ground truth.

1. Introduction

Since the initial proposal in [8], EKF (Extended Kalman Filter) monocular SLAM (Simultaneous Localization And Mapping) methods have proven valid for computing both scene structure and camera motion in real time at video rate. These methods successfully implement Bayesian estimator from image sequences, the prior that the camera moves smoothly according to the laws of dynamics is coded by means of a constant velocity model, regarding the scene, it is assumed as perfectly rigid. The rigidity assumption

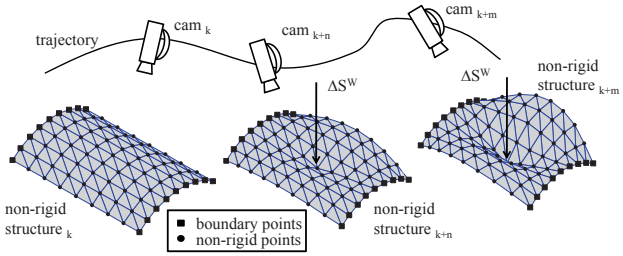

Figure 1. A moving camera observes a non-rigid structure. The nodes where the external forces are acting and their magnitude are unknown. Boundary points undergo a rigid motion with respect to the camera. It is prior knowledge what nodes are boundary points.

enforced by means of JCBB (Joint Compatibility Branch and Bound) [16], as shown in [7], or by means of 1-point RANSAC [6], has proven a key factor for robust performance.

We propose to extend Bayesian EKF sequential processing to estimate the structure and motion when observing a non-rigid scene by a hand-held monocular camera, eventually in real time. We aim to process medical endoscope sequences of body cavities such as the abdominal cavity. The cavity walls can be modeled as plates –thin solids–, and we assume a set of unknown magnitude forces acting on the surface Fig. 1. We propose to code 3D structure as a linear elastic solid following the Navier's equations. Despite the proposed elastic model is a low cost one, only exact for *small deformations*, a eventually *large scene deformation* can be accurately estimated, because the EKF is able to combine the available measurements of the deformed structure at frame rate with the elastic model.

Being a monocular sequence, the structure at rest has to be initially estimated by processing an initial subsequence without any non-rigid deformation. Navier's equations solution needs boundary conditions, in our case they are a set of known points to have a rigid motion. The Navier's equations are numerically solved by FEM (Finite Element Method).

The actual parameters defining the material elastic prop-

erties and the plate thickness are unknown. We propose to factor out these parameters from the state equation and use them to normalize the deforming external forces in order to ease the EKF tuning.

A set of salient scene points, the map, are selected to be estimated. We will refer to the map points as nodes because they are the nodes defining the discretization used to formulate the FEM solution to the Navier's equations.

Real imagery experimental validation is provided for a hand-held camera observing silicone cloth fixed in a stretcher while an unknown force acts on the cloth surface. The boundary conditions i.e. the identity of boundary points not undergoing non-rigid motion are known as priors. The computed structure and motion are shown to be compatible with a stereo ground truth.

2. Related Work

NRSfM (Non-Rigid Structure from Motion) computes, per each image in the sequence, both camera location and 3D structure when the image was taken. Being an under-constrained problem, additional smoothing constraints are necessary. Most of the approaches are based in the seminal work by Bregler *et al*. [4] where the time varying structure is coded as a time varying linear combination of predefined shapes. The camera model is orthographic. Factorization enforcing orthonormality closed form solution is computed by means of SVD. Factorization methods have been extended for the perspective case by [23, 13]. However, closed form solutions are reported to be noise sensitive by [3, 21].

Bundle adjustment (BA) has been applied to solve shape basis approaches to NRSfM. BA can additionally incorporate temporal and spatial smoothness priors both on the deformations and motion [9, 1]. Torresani *et al*. in [21] introduce a probabilistic linear dynamic model coding deformation weight as Gaussians solved by Expectation Maximization.

Methods based on linear shape basis have shown poor performance when dealing with large scene deformations. To cope with this limitation Fayad *et al*. in [11] replace the linear model by a quadratic deformation global model.

In contrast to previous global methods, our proposal is based on triangular elements where the consistency has been enforced to build global surfaces. Several proposals for local methods have been done. Rabaud and Belongie in [18] exploit the concept of locally smooth manifold learning. Varol *et al*. in [22] propose planar models. Fayad *et al*. in [10] propose quadratic models. Taylor *et al*. in [20] propose isometric models where local distance is preserved. In our proposal the local elements are modeled according to Navier's equations within a FEM approach.

A more recent approach to provide additional constraints has been template based methods [19, 17]. They propose to compute correspondences between the current image and a reference image in which the 3D shape is known. The 3D structure is coded as a triangular mesh. On one hand temporal consistency has been proposed as an additional constraint. On the other hand geometrical constraints such as developable surfaces, smooth surfaces (global and local smoothness) or distance constraints are also applied.

The physics based FEM applied to non-rigid structure detection can be tracked back to [15], tomography imagery is used as input, in our proposal we also use FEM models but a computer vision projective camera provides the input. More recently, and closer to us, is the work by Ilić and Fua [14] using FEM for the first time in computer vision. They proposed an expensive non-linear model accurate for large deformations, they focus on beam like structures, modeled as 1D FEM, including in the formulation forces and boundary conditions, resulting in a robust and accurate tracking method. We similarly exploit FEM models but we tackle a 2D scene, discretized as a triangular mesh. In contrast our model is low cost and only it is valid for small deformations, but combined with the EKF and digesting all the images of video sequence is able to accurately estimate large scene deformations.

Most of the above mentioned methods assume the matches as available from a previous tracking process, and then it is computed the structure from motion in batch mode, simplified camera models are frequent. In contrast, our proposal is purely sequential –providing an estimate for the sequence frames– and both estimation and matching are coupled in a sequential Bayesian approach for a general full perspective camera, making the most of the priors that Navier's equations can feed in the problem.

3. Non-Rigid Structure FEM Modeling

Given a linear elastic solid, Ω, the steady state Navier's equations Eq. (1) and the boundary conditions Eq. (2) [24, 25] model the solid deformation. Both equations use Einstein's index notation.

$$(\lambda + G)a_{j,ij} + Ga_{i,jj} + f_i = 0 \;\; in \;\; \Omega, \tag{1}$$

$$a_i = \underline{a}_i \;\; in \;\; \Gamma \tag{2}$$

where Γ is the solid boundary. a_i is the displacement vector. f_i is the volumetric force. λ and G are the Lamé parameters that define the material elastic properties, both of them are defined in terms of the Young's modulus, E, and the Poisson's ratio, ν, being $\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)}$ and $G = \frac{E}{2(1+\nu)}$.

To code the geometry we assume the solid is a plate, modeled by its middle plane, because it has a small thickness (Fig. 3). We approximate the continuous curvature surface by a planar triangular mesh. Each triangular element is defined by its 3 vertexes denominated nodes. Each triangular mesh element is transformed to a normalized triangle in natural coordinates (ξ, η) (Fig. 2). It used the stan-

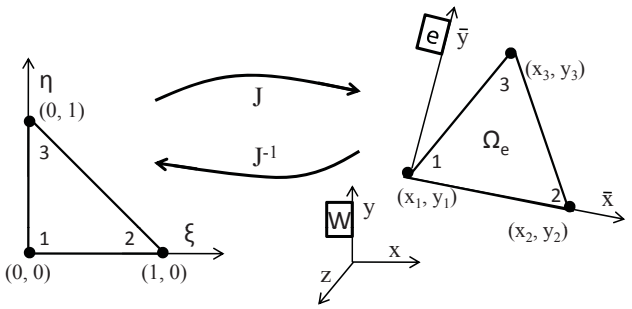

Figure 2. Normalized triangle in natural coordinates (left) and geometric transformation to real triangle (right).

dard approach [24] where the geometry within the normalized triangle is coded by the nodal linear shape functions $N_j\,(\xi,\eta)\quad j=1,2,3$.

Navier's equations are solved numerically, applying FEM, resulting in a linear equation system:

$$\boldsymbol{K}^W\,\mathbf{a}^W=\mathbf{f}^W \tag{3}$$

where $\boldsymbol{K}^W$ is the global stiffness matrix (the superindex ^{W} denotes world: global coordinates). $\mathbf{a}^W$ is the nodal displacement vector –displacements for every vertex–, and $\mathbf{f}^W$ is the nodal force vector –forces acting in every vertex–. FEM methods compute $\boldsymbol{K}^W$ by assembling matrices corresponding to every triangle in the discretization. The rest of the section is devoted to computing the elemental stiffness matrix. For readability, the assembly process is not described.

The elemental stiffness matrix, $\bar{\boldsymbol{K}}^e$ (the bar denotes coordinates in local reference, the superindex ^{e} denotes elemental), in its planar triangular domain Ω_e is defined as (Fig. 2):

$$\bar{\boldsymbol{K}}^e=\int_0^1\int_0^{1-\xi}\mathbf{B}^\top\mathbf{D}\mathbf{B}\,|\mathbf{J}|\,d\eta\,d\xi \tag{4}$$

where $\mathbf{B}$ is composed of the shape function derivatives. $\mathbf{D}$ is the behaviour matrix depending on the material elastic properties. $|\mathbf{J}|$, the Jacobian of the transformation from the natural to the local coordinates. The integral in Eq. (4), is computed applying Hammer's numerical integration for triangles.

$\mathbf{J}$ is the Jacobian matrix of the transformation from natural (ξ,η) to local $(\bar{x},\bar{y})$ coordinates. For our linear shape functions:

$$\mathbf{J}=\begin{pmatrix}\frac{\partial\bar{x}}{\partial\xi} & \frac{\partial\bar{y}}{\partial\xi}\\ \frac{\partial\bar{x}}{\partial\eta} & \frac{\partial\bar{y}}{\partial\eta}\end{pmatrix}=\begin{pmatrix}\bar{x}_2-\bar{x}_1 & \bar{y}_2-\bar{y}_1\\ \bar{x}_3-\bar{x}_1 & \bar{y}_3-\bar{y}_1\end{pmatrix}. \tag{5}$$

In our case, the nodal displacements for the j node are:

$$\bar{\mathbf{a}}_j^{mb}=\begin{pmatrix}\bar{u}_j & \bar{v}_j & \bar{w}_j & \theta_{\bar{x}j} & \theta_{\bar{y}j} & \theta_{\bar{z}j}\end{pmatrix}^\top, \tag{6}$$

where $(\bar{u}_j,\bar{v}_j)$ are the $\bar{x},\bar{y}$ displacements due to the membrane contribution. The membrane effect is approximated by a linear shape function within the element. $\bar{w}_j$, $\theta_{\bar{x}j}$ and $\theta_{\bar{y}j}$ are due to the bending contribution they represent $\bar{z}$ displacement, $\theta_{\bar{x}j}$ rotation, and $\theta_{\bar{y}j}$ respectively. The bending effect is approximated by a quadratic shape function (DKT element [2]) within the element. The superindexes mb, m and b denote membrane and bending, membrane only, and bending only contributions respectively. The relation between the nodal forces $\bar{\mathbf{f}}^e$ and the nodal displacements $\bar{\mathbf{a}}^e$ in each element is:

$$(\bar{\boldsymbol{K}}^e)^{mb}\,(\bar{\mathbf{a}}^e)^{mb}=(\bar{\mathbf{f}}^e)^{mb}. \tag{7}$$

The matrix $(\bar{\boldsymbol{K}}^e)^{mb}$ is formed by assembling $\bar{\boldsymbol{K}}_{ij}^{mb}$ $i,j=1,2,3$, corresponding to node pairs i,j, membrane and bending contributions ($\varepsilon\approx 0$):

$$\bar{\boldsymbol{K}}_{ij}^{mb}=\begin{pmatrix}\bar{\boldsymbol{K}}_{ij}^{m} & \vdots & \mathbf{0} & \vdots & \mathbf{0}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ \mathbf{0} & \vdots & \bar{\boldsymbol{K}}_{ij}^{b} & \vdots & \mathbf{0}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ \mathbf{0} & \vdots & \mathbf{0} & \vdots & \varepsilon\end{pmatrix}. \tag{8}$$

Next, $\mathbf{B}$ is defined in terms of $\mathbf{B}_j^m$ $j=1,2,3$, membrane contribution per node that have to be assembled, and, $\mathbf{B}^b$, the bending contribution:

$$\mathbf{B}_j^m=\frac{1}{|\mathbf{J}|}\begin{pmatrix}J_{22}N_{j,\xi}-J_{12}N_{j,\eta} & 0\\ 0 & J_{11}N_{j,\eta}-J_{21}N_{j,\xi}\\ J_{11}N_{j,\eta}-J_{21}N_{j,\xi} & J_{22}N_{j,\xi}-J_{12}N_{j,\eta}\end{pmatrix}, \tag{9}$$

$$\mathbf{B}^b=\frac{1}{|\mathbf{J}|}\begin{pmatrix}J_{22}\mathbf{N}_{x,\xi}^\top-J_{12}\mathbf{N}_{x,\eta}^\top\\ J_{11}\mathbf{N}_{y,\eta}^\top-J_{21}\mathbf{N}_{y,\xi}^\top\\ J_{11}\mathbf{N}_{x,\eta}^\top-J_{21}\mathbf{N}_{x,\xi}^\top+J_{22}\mathbf{N}_{y,\xi}^\top-J_{12}\mathbf{N}_{y,\eta}^\top\end{pmatrix}, \tag{10}$$

where $N_{j,\xi}=\frac{\partial N_j}{\partial\xi}$, $N_{j,\eta}=\frac{\partial N_j}{\partial\eta}$ and $\mathbf{N}_{x,\xi}^\top$, $\mathbf{N}_{y,\xi}^\top$ contain the shape function derivatives for the DKT element [2].

Next $\mathbf{D}$ matrix is defined in terms of the membrane and bending contributions. It depends on the plate thickness h and elastic parameters Young's modulus E and the Poisson's ratio ν:

$$\mathbf{D}^m=\frac{Eh}{1-\nu^2}\begin{pmatrix}1 & \nu & 0\\ \nu & 1 & 0\\ 0 & 0 & \frac{1-\nu}{2}\end{pmatrix}, \tag{11}$$

$$\mathbf{D}^b=\frac{Eh^3}{12(1-\nu^2)}\begin{pmatrix}1 & \nu & 0\\ \nu & 1 & 0\\ 0 & 0 & \frac{1-\nu}{2}\end{pmatrix}. \tag{12}$$

$\boldsymbol{K}^W$ is assembled from the $\boldsymbol{K}^e$ set. The $\boldsymbol{K}^e$ set is assembled from $\boldsymbol{K}_{ij}^{mb}$. $\boldsymbol{K}_{ij}^{mb}$ has to be transformed from local

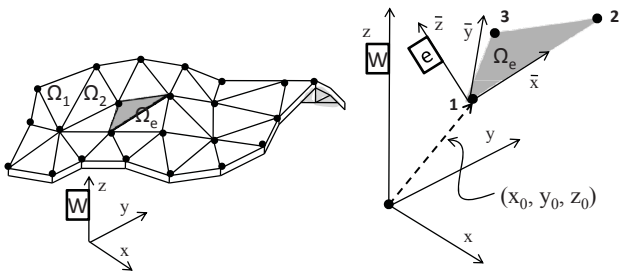

Figure 3. Solid discretize by linear triangular elements. World and local coordinates.

to global coordinates (Fig. 3). The transformation $\mathbf{T}$ is applied:

$$\boldsymbol{K}_{ij}^{mb} = \mathbf{T}^{\top}\, \bar{\boldsymbol{K}}_{ij}^{mb}\, \mathbf{T}\,, \tag{13}$$

$$\mathbf{T} = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda} \end{pmatrix} \tag{14}$$

where $\mathbf{\Lambda}$ defines the transformation from local, $(\bar{x}_i, \bar{y}_i, \bar{z}_i)$, to global, (x_i, y_i, z_i), coordinates:

$$\begin{pmatrix} \bar{x}_i \\ \bar{y}_i \\ \bar{z}_i \end{pmatrix} = \mathbf{\Lambda} \begin{pmatrix} x_i - x_0 \\ y_i - y_0 \\ z_i - z_0 \end{pmatrix}, \tag{15}$$

$$\mathbf{\Lambda} = \begin{pmatrix} \cos(\bar{x}, x) & \cos(\bar{x}, y) & \cos(\bar{x}, z) \\ \cos(\bar{y}, x) & \cos(\bar{y}, y) & \cos(\bar{y}, z) \\ \cos(\bar{z}, x) & \cos(\bar{z}, y) & \cos(\bar{z}, z) \end{pmatrix} \tag{16}$$

where (x_0, y_0, z_0) are the per each triangular element local origin coordinates in the global frame. A different $\mathbf{\Lambda}$ has to be considered per each element and per each sample time, however indexes have been dropped for simplicity.

4. Coding Elastic Models in EKF

This section is devoted to combining the EKF with FEM for the sequential estimation.

4.1. State Vector Definition

As in standard EKF SLAM, we use a single joint state vector containing camera pose and feature estimates, with the assumption that the camera moves with respect to the structure. The whole state vector $\mathbf{x} = \left(\mathbf{x}_v^{\top}, \mathbf{y}^{\top}\right)^{\top}$ is composed of the camera state, $\mathbf{x}_v$ and all the structure nodes $\mathbf{y} = \left(\mathbf{y}_1^{\top}, \ldots, \mathbf{y}_n^{\top}\right)^{\top}$.

For the camera motion we propose the classical constant velocity model [8]. Camera state:

$$\mathbf{x}_v = \left(\mathbf{r}^{WC\top}, \mathbf{q}^{WC\top}, \mathbf{v}^{W\top}, \omega^{C\top}\right)^{\top}, \tag{17}$$

where $\mathbf{r}^{WC}$ is camera translation, $\mathbf{q}^{WC}$ is the quaternion representing orientation with respect to the world frame, $\mathbf{v}^{W}$ and ω^{C} are linear and angular velocities. We assume that linear and angular accelerations $\mathbf{a}^{W}$ and α^{C} affect the camera, producing at each step an impulse of linear velocity, $\mathbf{V}^{W} = \mathbf{a}^{W}\Delta t$, and angular velocity $\Omega^{C} = \alpha^{C}\Delta t$, with an zero-mean Gaussian distribution being $\mathbf{Q}_{\mathbf{x}_v}$ its covariance. The state equation for the camera is:

$$\mathbf{g}_v = \begin{pmatrix} \mathbf{r}_{k+1}^{WC} \\ \mathbf{q}_{k+1}^{WC} \\ \mathbf{v}_{k+1}^{W} \\ \omega_{k+1}^{C} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_{k}^{WC} + (\mathbf{v}_k^W + \mathbf{V}^W)\Delta t \\ \mathbf{q}_{k}^{WC} \times \mathbf{q}\left((\omega_k^C + \Omega^C)\Delta t\right) \\ \mathbf{v}_k^W + \mathbf{V}^W \\ \omega_k^C + \Omega^C \end{pmatrix}, \tag{18}$$

where $\mathbf{q}\left((\omega_k^C + \Omega^C)\Delta t\right)$ is the quaternion defined by the rotation vector $(\omega_k^C + \Omega^C)\Delta t$.

Our contribution is to code the structure as non-rigid by means of the compliance matrix Eq. (20) $\mathbf{C}_k\left(\hat{\mathbf{y}}_k^W\right)$, depending on the current undeformed structure geometry $\hat{\mathbf{y}}_k^W$. The acting deforming normalized force $\Delta\mathbf{S}^W$ is causing recursively at each step an incremental deformation, so next step, $\mathbf{C}_{k+1}$ will be computed over the deformed structure $\hat{\mathbf{y}}_{k+1}^W$. We assume $\Delta\mathbf{S}^W$ follows a zero-mean Gaussian distribution being $\mathbf{Q}_{\mathbf{y}}$ its covariance. So, the state equation for the structure:

$$\mathbf{g}_y = \mathbf{y}_{k+1}^W = \mathbf{y}_k^W + \mathbf{C}_k \Delta\mathbf{S}^W, \tag{19}$$

where $\mathbf{C}_k$ results from $\boldsymbol{C}_k$ after removing the rows and columns corresponding to rotations. Because we are interested only in middle plane nodes location where the rotation effect is null. $\boldsymbol{C}_k$ is defined to solve linear system Eq. (3) as:

$$\boldsymbol{C}_k = \boldsymbol{K}_k^{W\,-1}. \tag{20}$$

Linear elastic materials are characterized by E, ν and the thin-plate thickness h. We assume almost incompressible material and hence $\nu = 0.499$ is known. However both E and h are unknown. The external force for i node, $\left(\Delta f_{xi}^W, \Delta f_{yi}^W, \Delta f_{zi}^W\right)^{\top}$ is normalized as $\Delta\mathbf{S}_i^W$:

$$\Delta\mathbf{S}_i^W = \frac{1}{Eh}\left(\Delta f_{xi}^W, \Delta f_{yi}^W, \Delta f_{zi}^W\right)^{\top}, \tag{21}$$

to concentrate the unknown magnitudes in the state noise vector. However, the $\mathbf{C}_k$ compliance matrix still depends on a h^2 factor.

For tuning, on the one hand we propose to tune $\Delta\mathbf{S}_i^W$ as a diagonal matrix, where the standard deviation value codes the tangential deformation, measured in length units, when applied the typical tangential force. On the other hand, if the typical force is applied normal to the surface, the deformation will be bigger than the tangential deformation, approximately proportional to $\frac{1}{h^2}$. So we use h^2 to code this anisotropy.

The covariance $\mathbf{Q}_{\mathbf{y}}$ governs the magnitudes of the forces acting on the structure. A null $\mathbf{Q}_{\mathbf{y}}$ codes a rigid scene.

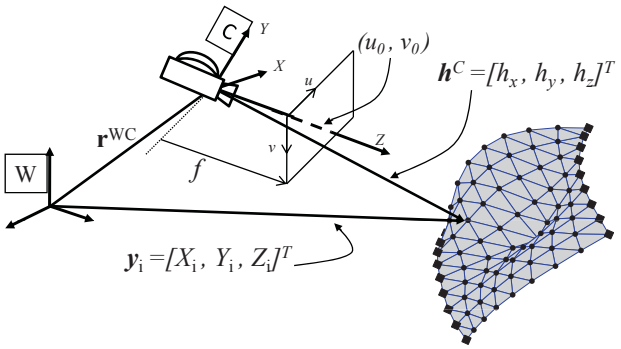

Figure 4. Measurement equation.

4.2. Coding Priors about Scene: Forces and Boundary Conditions

The general Eq. (19) assumes 3D forces acting on every structure node. It also assumes as known which scene points are boundary points and hence their motion is rigid with respect to the camera. The forces acting on boundary points are null.

Stiffness matrix $\boldsymbol{K}_k^W$ is defined by the node connectivity and the identity of the boundary points, it is mainly a sparse band diagonal matrix. If the proper boundary points are defined, $\boldsymbol{K}_k^W$ is invertible.

The normalized force vector, $\Delta\mathbf{S}^W$:

$$\Delta\mathbf{S}^W = \begin{pmatrix} \Delta\mathbf{s}^{W\top} & \mathbf{0} \end{pmatrix}^\top, \tag{22}$$

$$\Delta\mathbf{s}^W = \begin{pmatrix} \Delta\mathbf{s}_1^{W\top} & \dots & \Delta\mathbf{s}_i^{W\top} & \dots & \Delta\mathbf{s}_p^{W\top} \end{pmatrix}^\top \tag{23}$$

is a zero vector except in the first $3p$ components, $\Delta\mathbf{s}^W$, corresponding to the non-rigid nodes where the forces are acting. p is the number of non-rigid nodes, n is the total number of nodes. So the product $\mathbf{C}_k\Delta\mathbf{S}^W = \mathbf{A}_k\Delta\mathbf{s}^W$ where $\mathbf{A}_{3n\times 3p}$ is a submatrix formed by selecting the elements in $\mathbf{C}_k$ acting on non null $\Delta\mathbf{S}^W$. So Eq. (19) is simplified to:

$$\mathbf{g}_y = \mathbf{y}_{k+1}^W = \mathbf{y}_k^W + \mathbf{A}_k\Delta\mathbf{s}^W. \tag{24}$$

4.3. Initialization

We are assuming that the moving camera first observes the structure at rest –because no force is acting on the 3D structure–, so the 3D structure at rest can be estimated. Afterwards, the scene model is switched to non-rigid Eq. (24).

The 3D structure at rest is computed using the classical EKF monocular SLAM. Map points are initialized in inverse depth and then converted to Euclidean XYZ coding. The 3D structure at rest is considered to be accurately estimated when most of the scene points are switched to XYZ [5] coding, then scene model is switched to non-rigid. Boundary points have to be identified prior to non-rigid model switching.

In the experimental section it is shown that once the 3D structure at rest is estimated, the non-rigid model can deal both with deforming and non deforming scenes.

4.4. EKF Formulation

To sum up, the state equations are Eq. (18) and (24) and the corresponding Jacobians for the EKF are:

$$\mathbf{F}_k = \begin{pmatrix} \frac{\partial \mathbf{g}_v}{\partial \mathbf{x}_v} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{g}_y}{\partial \mathbf{y}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{I}\Delta t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{q}_{k+1}^{WC}}{\partial \mathbf{q}_k^{WC}} & \mathbf{0} & \frac{\partial \mathbf{q}_{k+1}^{WC}}{\partial \omega_k^C} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{25}$$

$$\mathbf{G}_k = \begin{pmatrix} \frac{\partial \mathbf{g}_v}{\partial \mathbf{n}} \\ \frac{\partial \mathbf{g}_y}{\partial \mathbf{n}} \end{pmatrix} = \begin{pmatrix} \mathbf{I}\Delta t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{q}_{k+1}^{WC}}{\partial \Omega^C} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_k \end{pmatrix}, \tag{26}$$

where $\mathbf{n} = (\mathbf{a}^{W\top} \;\; \alpha^{C\top} \;\; \Delta\mathbf{s}^{W\top})^\top$ is the state vector noise.

4.5. Measurement Equation

Each observed feature imposes a constraint between the camera location and the corresponding map feature (Fig. 4). The observation of a point $\mathbf{y}_i = (X_i, Y_i, Z_i)^\top$ defines a ray coded by a directional vector $\mathbf{h}^C = \begin{pmatrix} h_x & h_y & h_z \end{pmatrix}^\top$ in the camera frame:

$$\mathbf{h}^C = \mathbf{R}^{CW} \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} - \mathbf{r}^{WC} \Bigg), \tag{27}$$

where $\mathbf{R}^{CW}$ is the rotation matrix corresponding to $\mathbf{q}^{WC}$. The camera does not directly observe $\mathbf{h}^C$ but its projection in the image according to the pinhole model. Projection to a normalized retina and then camera calibration is applied:

$$\mathbf{h} = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u_0 - \frac{f}{d_x}\frac{h_x}{h_z} \\ v_0 - \frac{f}{d_y}\frac{h_y}{h_z} \end{pmatrix}, \tag{28}$$

where (u_0, v_0) is the camera principal point, f is the focal length and (d_x, d_y) is the pixel size. Finally, a distortion model has to be applied to deal with real camera lenses. In this work we have used the standard two parameter distortion model from photogrammetry.

5. Experimental Results

The proposed method has been validated observing a synthetic silicone cloth (Fig. 5). As the silicone is textureless, artificial markers have been painted on its surface. The silicone has been placed in a circular stretcher to fix the boundary conditions. The silicone is observed with a mobile stereo rig at 30Hz, its resolution is 640×480. The silicone suffers noticeable deformation resulting from the action of a force applied in its central point. For testing the proposed monocular algorithm the right camera sequence, at half resolution 320×240 is processed. To provide a quantitative comparison, the stereo pair at full resolution is used to produce a ground truth for both non-rigid structure and camera motion at 10 selected key camera locations. The ground truth has been computed using commercial photogrammetric software to process the selected stereo pairs.

The sequence is composed of 2699 images. The first 899 correspond to a non-deforming structure, because no force is applied. After image #900 the silicone is being pushed with a stick handled by a person. The full sequence is processed according to the proposed algorithm. At frame 360 all the points are switched to XYZ coding and hence the 3D structure at rest is regarded as estimated. To fix the monocular unknown scale factor, the distance from the camera to a selected boundary point is fixed to its ground truth value. A measurement equation where the selected point is observed with a minute measurement error ϵ, is included in the EKF estimation. Once the scaled 3D structure at rest is estimated, the non-rigid model is applied. It has to be noted that between frames 360 and 899, the estimation assumes a non-rigid scene despite the actual images correspond to a rigid scene, however, the non-deforming scene is accurately estimated (see Fig. 7 and the accompanying video[1]).

Fig. 6 shows a general perspective view of the reconstructed structure and the corresponding ground truth. Figure 7 details the structure evolution in a cross section. In frame #2 it can be seen the quite uncertain initialized structure from just one image. #475 and #899 correspond to the estimation when no force is applied. Next frames show the evolution corresponding to the deforming scene. It is worth noting how the boundary points covariance is small because they are coded as rigid points. It is also remarkable the consistency between estimation and the ground truth, both for the structure and for the camera trajectory (Fig. 8). The estimation is consistent, and the estimated covariance is small so the quantitative comparison reports an accurate non-rigid estimation.

Regarding computing budget, there is room for further research, but the additional cost with respect to EKF is just the inversion of a sparse matrix to compute the compliance matrix $\mathbf{C}_k$. It is roughly double in size the covariance matrix, so the cost is comparable to that of the EKF, so frame rate real time is achievable for maps about hundred points.

[1] http://webdiis.unizar.es/%7Ejosemari/4dmod11.avi

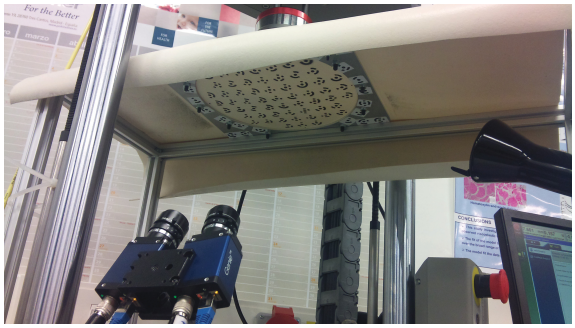

Figure 5. Mobile stereo rig observing a silicone cloth fixed in a circular stretcher.

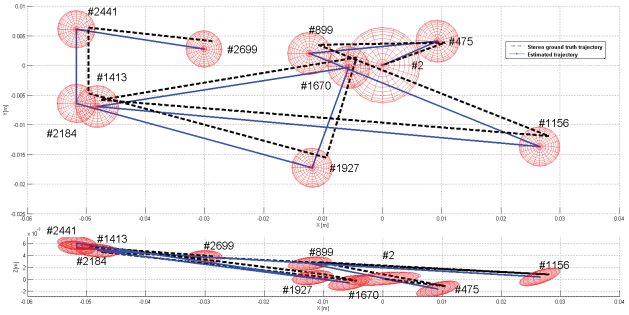

Figure 8. Estimated camera trajectory and its corresponding covariance, compared wrt. ground truth. Top view (above), lateral view (below).

6. Conclusions and Future Work

It has been shown how a rigorous coding of Navier's equations by FEM numerical method is an adequate tool to include physical priors in the EKF to solve a NRSfM problem, being a method that can eventually perform in real time. The experimental validation shows that estimated structure is consistent –both the structure and motion– with respect to the ground truth. Given the small covariances and the consistency we can conclude that the experiments provide a quantitative estimation of the method accuracy.

It has to be stressed that we have used a full perspective camera, and the method computes both the NRSfM problem and also the matching for all the frames in a video sequence.

The elastic model considered is the simplest possible, linear elastic defined in just in terms of E and ν, and for the case of small deformations. Despite the simple model a large deformation problem has been solved, due to the sequential estimation that corrects the estimated deformation at frame rate, resulting in an accurate estimation at a low computational cost.

Next steps are a real time implementation and the pro-

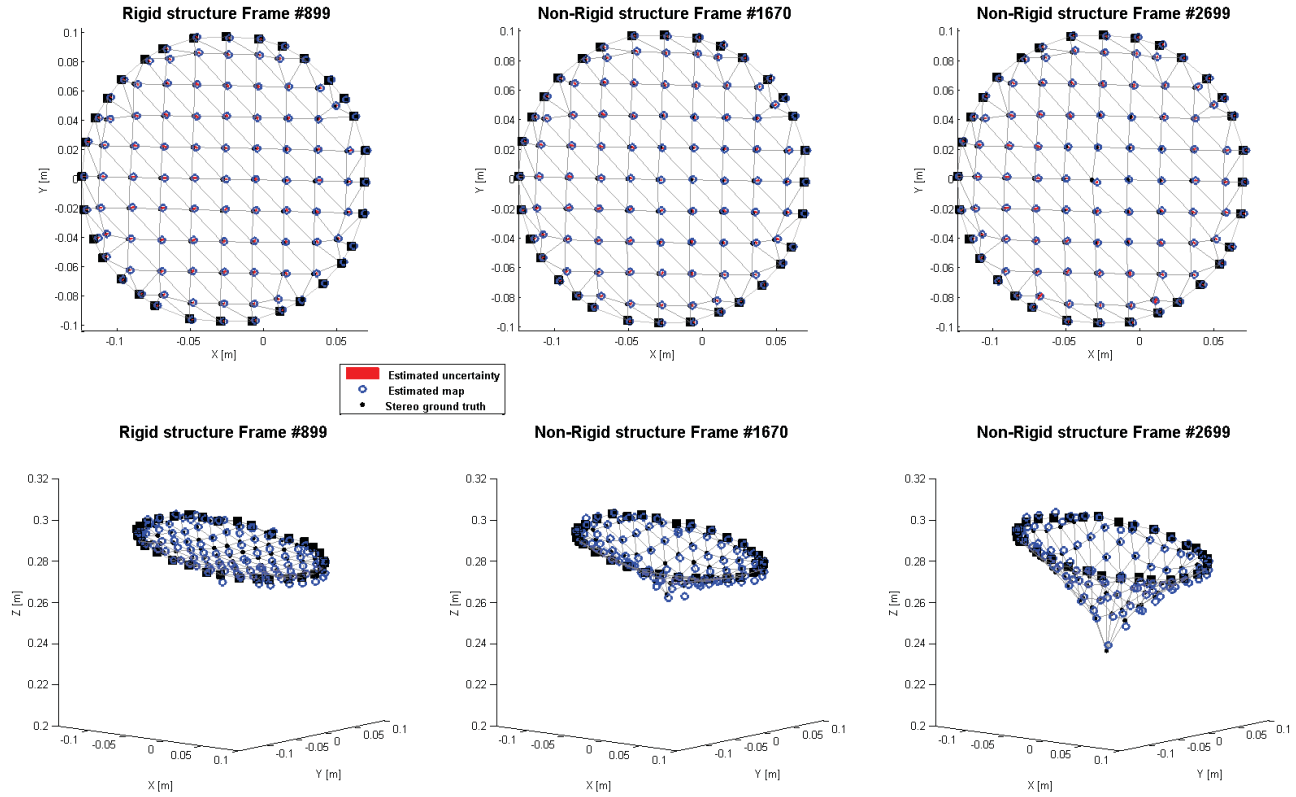

Figure 6. (left) frame #899 structure at rest, (center) frame #1670 deformed structure and (right) frame #2699 deformed structure. Black points and the mesh code the ground truth. Blue points code the estimated structure. Covariance, in red, is only represented in top view to ease visibility. Covariance in the structure at rest is almost imperceptible.

cessing of real medical image sequences, similar to those in [12] where the map points correspond to natural landmarks in the images, focusing in being able to register the scene non-rigid deformations and to perform robustly under motion clutter. The proposed algorithm is particularly well suited to this case because monocular endoscope observations are frequent and rich priors about the observed scene elastic properties are available.

Acknowledgments

This work was supported by the Spanish MICINN DIP2009-07130 and DPI2011-27939-C02-01 grants.

Thanks to J. Civera and Óscar G. Grasa for fruitful discussion and software assistance. Thanks to J. Romeo for assistance in experimental data acquisition.

References

[1] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd. Coarse-to-fine low-rank structure-from-motion. In *Conference on Computer Vision and Pattern Recognition*, 2008. 2

[2] J. L. Batoz, K. J. Bathe, and L. W. Ho. A study of three-node triangular plate bending elements. *Int. J. Num. Meth. Eng.*, 25:1771–1812, 1980. 3

[3] M. Brand. A direct method of 3D factorization of nonrigid motion observed in 2D. In *CVPR*, pages 122–128, 2005. 2

[4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *CVPR*, 2000. 2

[5] J. Civera, A. J. Davison, and J. M. M. Montiel. Inverse depth parametrization for monocular SLAM. *IEEE Transactions on Robotics*, 24(5):932–945, October 2008. 5

[6] J. Civera, O. G. Grasa, A. J. Davison, and J. M. M. Montiel. 1-Point RANSAC for EKF Filtering: application to real-time structure from motion and visual odometry. *Journal of Field Robotics*, 27(5):609–631, October 2010. 1

[7] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardós. Mapping large loops with a single hand-held camera. In *RSS*, 2007. 1

[8] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *ICCV*, 2003. 1, 4

[9] A. Del Bue, X. Llado, and L. Agapito. Non-rigid metric shape and motion recovery from uncalibrated images using priors. In *CVPR*, 2006. 2

[10] J. Fayad, L. Agapito, and A. Del Bue. Piecewise quadratic reconstruction of non-rigid surfaces from monocular se-

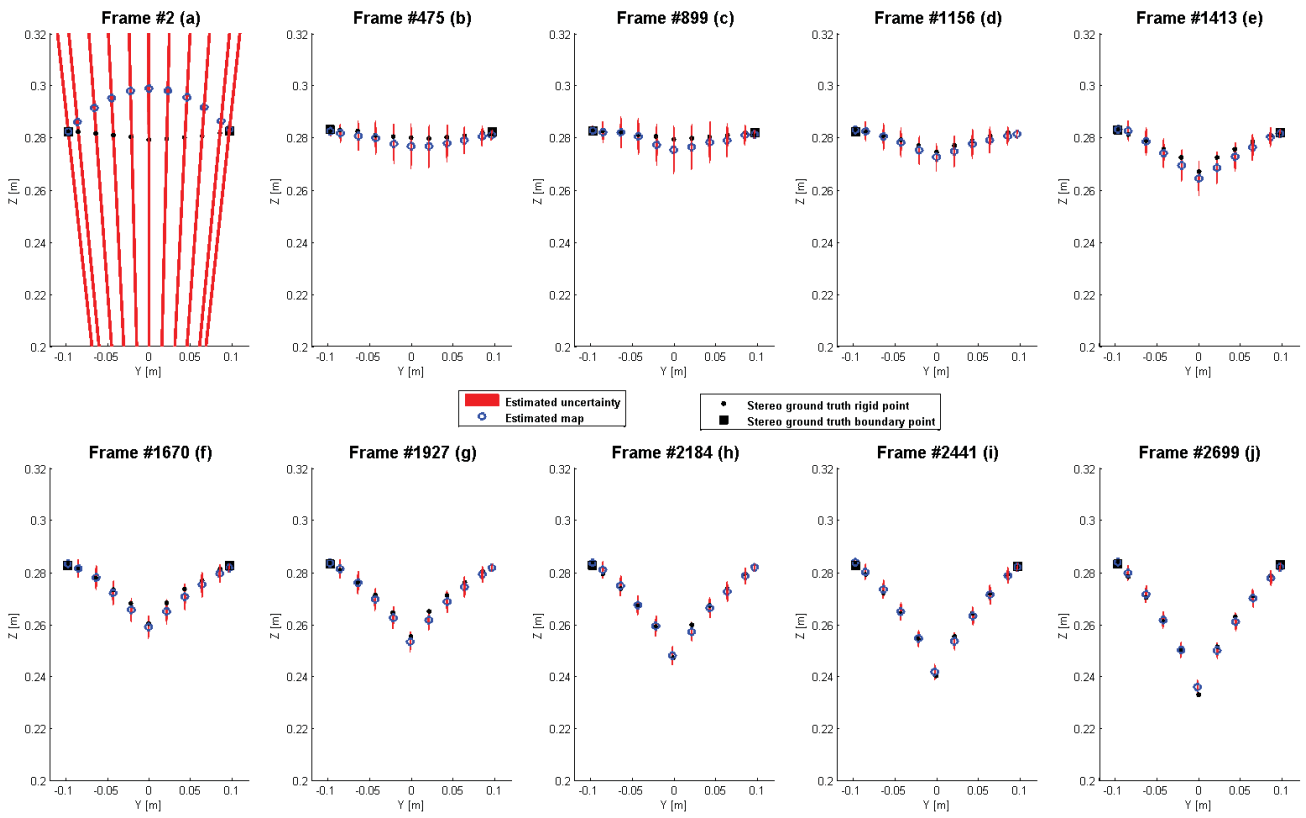

Figure 7. Estimation history cross section. It is represented the estimated structure in blue and 99% acceptance regions in red, and also the ground truth in black.

quences. volume 6314 of *Lecture Notes in Computer Science*, pages 297–310. Springer, 2010. 2

[11] J. Fayad, A. Del Bue, L. Agapito, and P. Aguiar. Non-rigid structure from motion using quadratic deformation models. In *British Machine Vision Conference, London*, 2009. 2

[12] O. G. Grasa, J. Civera, and J. M. M. Montiel. EKF monocular SLAM with relocalization for laparoscopic sequences. In *Conference on Robotics and Automation*, 2011. 7

[13] R. Hartley and R. Vidal. Perspective nonrigid shape and motion recovery. In *ECCV*, October 2008. 2

[14] S. Ilić and P. Fua. Non-linear beam model for tracking large deformation. In *ICCV*, October 2007. 2

[15] T. McInerney and D. Terzopoulos. A finite element model for 3D shape recognition and nonrigid motion tracking. In *International Conference on Computer Vision*, pages 518–523, 1993. 2

[16] J. Neira and J. D. Tardós. Data association in stochastic mapping using the joint compatibility test. *IEEE Transactions on Robotics*, 17(6):890–897, 2001. 1

[17] M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. In *International Journal of Computer Vision*, 2010. 2

[18] V. Rabaud and S. Belongie. Re-thinking non-rigid structure from motion. In *Conference on Computer Vision and Pattern Recognition*, June 2008. 2

[19] M. Salzmann, R. Urtasun, and P. Fua. Local deformation models for monocular 3D shape recovery. In *Conference on Computer Vision and Pattern Recognition*, 2008. 2

[20] J. Taylor, A. D. Jepson, and K. N. Kutulakos. Non-rigid structure from locally-rigid motion. In *Conference on Computer Vision and Pattern Recognition*, June 2010. 2

[21] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from motion: estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):878–892, 2008. 2

[22] A. Varol, M. Salzmann, E. Tola, and P. Fua. Template-free monocular reconstruction of deformable surfaces. In *IEEE International Conference on Computer Vision*, Kyoto, Japan, 2009. 2

[23] J. Xiao and T. Kanade. Uncalibrated perspective reconstruction of deformable structures. In *International Conference on Computer Vision*, 2005. 2

[24] O. C. Zienkiewicz and R. L. Taylor. *The finite element method. Vol. 1: Basic formulation and linear problems*. McGraw-Hill, London, 1989. 2, 3

[25] O. C. Zienkiewicz and R. L. Taylor. *The finite element method. Vol. 2: Solid and fluid mechanics, dynamics and non-linearity*. McGraw-Hill, London, 1989. 2