

Finite Element based Sequential Bayesian Non-Rigid Structure from Motion

Antonio Agudo Begoña Calvo J. M. M. Montiel
 Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, Spain
 {aagudo, bcalvo, josemari}@unizar.es

Abstract

Navier’s equations modelling linear elastic solid deformations are embedded within an Extended Kalman Filter (EKF) to compute a sequential Bayesian estimate for the Non-Rigid Structure from Motion problem. The algorithm processes every single frame of a sequence gathered with a full perspective camera. No prior data association is assumed because matches are computed within the EKF prediction-match-update cycle.

Scene is coded as a Finite Element Method (FEM) elastic thin-plate solid, where the discretization nodes are the sparse set of scene points salient in the image. It is assumed a set of Gaussian forces acting on solid nodes to cause scene deformation. The EKF combines in a feedback loop an approximate FEM model and the frame rate measurements from the camera, resulting in an efficient method to embed Navier’s equations without resorting to expensive non-linear FEM models.

Classical FEM modelling has implied an interactive identification of boundary points to constrain the scene rigid motion, in this work this dissatisfying prior knowledge is no longer needed. The scene and camera rigid motion are combined in a unique pose vector and the estimation is coded relative to the camera. Additionally, the deforming effect of the Gaussian forces on the thin-plate is computed by means of the Moore-Penrose pseudoinverse of the FEM stiffness matrix.

The proposed algorithm is validated with three real sequences gathered with hand-held camera observing isometric and non-isometric deformations. It is also shown the consistency of the EKF estimation with respect to ground truth computed from stereo.

1. Introduction

Given a full perspective camera undergoing an unknown trajectory while observing an also unknown rigid scene (Fig. 1(a)). Both the camera trajectory \mathbf{x}_v and the scene 3D structure can be estimated, up to scale factor, just from the sole input of the image sequence gathered by the cam-

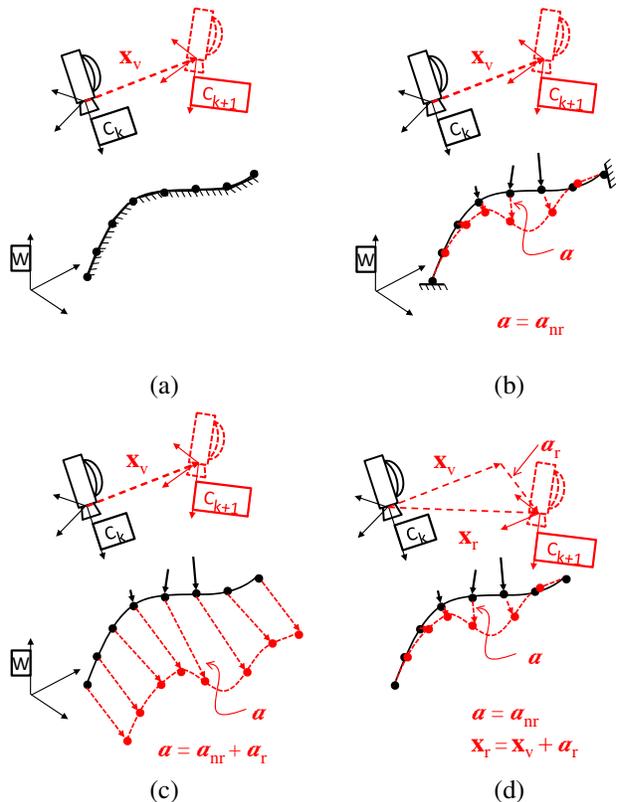


Figure 1. (a) A mobile camera observing a rigid scene. (b) A mobile camera observing a non-rigid scene deformed under the action of a Gaussian set of forces; boundary points prevent scene rigid displacements \mathbf{a}_{nr} . (c) Removing boundary points allows rigid motion for both the camera and the scene structure, on top of the scene non-rigid deformation. (d) Rigid motions for the scene and the camera are combined in the pose, \mathbf{x}_r . The scene non-rigid deformation is coded in \mathbf{a}_{nr} .

era, this is the classical rigid visual SLAM (Simultaneous Localisation And Mapping) problem. Extended Kalman Filter (EKF) based approaches [6, 7] were the first computing SLAM in real-time, more recently keyframe based approaches [15] are yielding more accurate and efficient SLAM algorithms. In any case, Agudo *et al.* in [1] recently proved that EKF SLAM can successfully deal with scenes

combining rigid and non-rigid points recovering both the camera trajectory and scene structure (Fig. 1(b)). The deformable scene is coded by means of Finite Element Method (FEM) modelling, it is assumed a set of Gaussian deforming forces acting on the non-rigid surface, resulting in the scene incremental displacement \mathbf{a} due to the non-rigid displacement \mathbf{a}_{nr} . In this case, rigid boundary points – corresponding to boundary conditions– are necessary for the proposed FEM formulation. However, identifying which scene points are boundary points is a too strong prior knowledge. It is our main contribution to remove the boundary points from the EKF-FEM formulation.

If boundary points are removed from the formulation, at each estimation step both the camera and the scene can undergo a rigid motion, \mathbf{x}_v and \mathbf{a}_r respectively (Fig. 1(c)). Additionally the scene can suffer a non-rigid deformation \mathbf{a}_{nr} .

The combined effect of the camera and scene rigid motion is coded in a unique *pose* vector \mathbf{x}_r and the non-rigid deformation is coded in \mathbf{a}_{nr} (Fig. 1(d)). We propose to code both the pose and the non-rigid scene with respect to the last camera location, following the robocentric formulation, proposed by Castellanos *et al.* in [5].

2. Related work

Non-Rigid Structure from Motion (NRSfM) computes time varying 3D structure and pose from the sole camera-gathered image sequence, it is an ill-posed problem so additional smoothing constraints are necessary. A seminal work was proposed by Bregler *et al.* [4] based on a low rank shape model in which the time varying 3D structure is coded as a linear combination of basis shapes. Paladini *et al.* in [18] propose a sequential version of the factorization method over a sliding window. These methods need to detect all the scene points in all the processed images, and orthographic cameras are assumed.

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 [2, 8]. Torresani *et al.* in [22] introduce an expectation maximization probabilistic linear dynamic model coding deformation weight as Gaussians. Reported experiments, compared to closed form, exhibit better noise rejection and improved accuracy. Also based on global optimization, Fayad *et al.* in [10] replace the linear model by a quadratic global model, achieving good results when dealing with large scene deformations.

In contrast to global models, piecewise models can code more accurately strong deformations composed of multiple local deformations. Piecewise methods rely on common features shared between patches to enforce spatial consistency and create a continuous global surface. Varol *et al.* in [23] propose planar piecewise patches. Fayad *et al.* in [9]

propose quadratic models. Taylor *et al.* in [21] propose a triangle soup being each triangle assumed rigid.

Template-based methods, [19, 20] propose to compute correspondences between the current image and a reference image in which the 3D shape —the structure at rest— is known. The 3D structure is coded as a triangular mesh. Like in all other methods, both temporal and spatial smoothing are applied. These methods can deal with projective cameras.

Previous optimization methods –except template-based– still relay mostly in an orthographic camera. Regarding data association can tolerate partial observation of scene points in the images but data association has to be given as prior. More recently Moreno-Noguer and Porta [17] code the isometric scene as a combination of prior known deformation modes with a perspective camera.

In computer vision, physics-based models have been proposed for recovering model deformations from images [14, 16]. Ilić and Fua in [13] propose an expensive non-linear FEM model for large deformations focused on 1D beam like structures. The formulation includes the forces and boundary conditions, resulting in a robust and accurate 2D tracking algorithm. Greminger and Nelson propose in [12] coded an elastic solid by means of the Boundary Element Method (BEM) for 2D deformable object tracking; by enforcing a 3 rank deficiency, boundary conditions are removed from the formulation.

Recently Agudo *et al.* in [1] propose linear FEM thin-plate model to embed Navier’s equations within the EKF estimation. They can deal with full perspective cameras and compute the data association, every single frame in the sequence is processed. Thanks to the Navier FEM model they can cope both with isometric and non-isometric scenes without assuming any scene deformation mode. The method is piecewise and can be considered close to template-based because the structure at rest is needed to cope with the deforming scene. Template-based methods register every image with respect to the initial 3D template. EKF-FEM compares the current frame with respect to the 3D scene structure estimated, after processing the initial structure at rest and all the previous frames in the image sequence. The method can combine both rigid and non-rigid points, but prior knowledge about scene point classification as boundary or non-boundary is mandatory. Every non-rigid point is coded by means of a 6 d.o.f vector. In the current work, we build on previous proposal but boundary points are removed from the formulation, additionally the points are coded only in 3 d.o.f, with the corresponding efficiency increase.

3. Thin-Plate Formulation

The 3D displacements in an isotropic linear elastic solid Ω under 3D external forces are modelled by the steady

state Navier's equations Eq. (1) and the boundary conditions Eq. (2) [24, 3]:

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

$$a_i = \bar{a}_i \text{ on } \Gamma. \quad (2)$$

Equations are coded according to Einstein's index convention. Γ is the solid boundary. a_i is the displacement vector. f_i is the volumetric force vector. λ and G are the Lamé parameters defining the material elastic properties:

$$\lambda = \frac{\nu E}{(1 + \nu)(1 - 2\nu)}, \quad G = \frac{E}{2(1 + \nu)}, \quad (3)$$

where E is the Young's modulus and ν the Poisson's ratio.

Navier's equations are approximated by means of FEM. FEM discretizes the solid in finite elements Ω_e (Fig. 2 (a)) defined by their nodes. The displacements are computed for every node solving the sparse linear system:

$$\mathbf{K} \mathbf{a} = \mathbf{f} \quad (4)$$

where \mathbf{K} is the global stiffness matrix, \mathbf{a} is the nodal displacements vector and \mathbf{f} is the nodal forces vector. The continuous displacement field is approximated by interpolation from the nodal displacements by means of the shape functions.

In [1] it is proposed a thin-plate formulation to reduce the full 3D solid mechanic problem in a 2D problem, using the 2D plane stress model for membrane effect –displacement within the plane– and the Kirchhoff–Love plate model for bending effect –displacement off the plane. \mathbf{a}_i and forces \mathbf{f}_i in each node i are 6 d.o.f vectors defined as:

$$\mathbf{a}_i = (u_i \ v_i \ w_i \ \theta_{xi} \ \theta_{yi} \ \theta_{zi})^\top, \quad (5)$$

$$\mathbf{f}_i = (f_{ui} \ f_{vi} \ f_{wi} \ \Theta_{xi} \ \Theta_{yi} \ \Theta_{zi})^\top. \quad (6)$$

In this paper, we propose a novel simplification additional to the thin-plate formulation where $(\theta_{xi}, \theta_{yi}, \theta_{zi})$ rotations and moments $(\Theta_{xi}, \Theta_{yi}, \Theta_{zi})$ are removed. For each node only translational displacements can be detected, in fact the rotations effects eventually are detected also as translations. Similarly the effect of an acting moment can be modelled as a force combination. Additionally to the half cut down in the state size, the stiffness matrix conditioning is also improved.

In [1] the \mathbf{K} is assembled from \mathbf{K}^e that is assembled from:

$$\mathbf{K}_{ij}^{mb} = \begin{pmatrix} \mathbf{K}_{Tij} & \vdots & \mathbf{K}_{TRij} \\ \mathbf{K}_{TRij} & \vdots & \mathbf{K}_{Rij} \end{pmatrix}. \quad (7)$$

while in our proposal rows and columns corresponding to the rotational displacements and moments are removed and \mathbf{K} is assembled from \mathbf{K}^{e*} , that is assembled from submatrix \mathbf{K}_{Tij} Eq. (7).

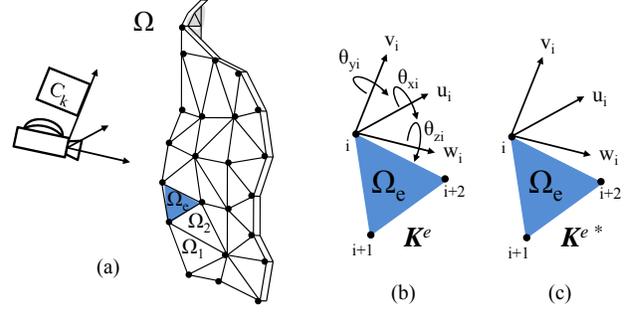


Figure 2. (a) Thin-plate discretization. Detail for the displacements and forces acting on a mesh node i . (b) according to [1]. (c) according to our proposal.

4. FEM Linear System Solution

A common FEM problem statement assumes two kinds of nodes: boundary and non-boundary. Displacements for boundary nodes are null Eq. (2) what provides additional constraints in order to have a unique solution for Eq. (4), so given forces acting on non-boundary nodes corresponding displacements can be uniquely computed.

It is our goal to remove the constrained boundary point displacements from the formulation, what implies that linear system Eq. (4) becomes under-constrained. In that case, the full affine space solution can be computed as [11]:

$$\mathbf{a} = \mathbf{a}_p + \mathbf{a}_h \quad (8)$$

$$\mathbf{K} \mathbf{a}_h = \mathbf{0} \quad (9)$$

$$\mathbf{a}_p = \mathbf{K}^+ \mathbf{f} \quad (10)$$

where \mathbf{a}_h is the homogeneous solution, a vector space dimension equal to the \mathbf{K} rank deficiency. \mathbf{a}_h is the set displacements vectors compatible with no force acting. As \mathbf{K} comes from a FEM formulation, \mathbf{a}_h corresponds to a rigid transformation, so the homogeneous solution should be dimension 6, corresponding to the 6 d.o.f. of the 3D rigid body motion. The particular solution, \mathbf{a}_p is computed by means of the Moore–Penrose pseudoinverse \mathbf{K}^+ .

Experimentally, it has been verified that \mathbf{K} rank deficiency is only 3, up to the numerical accuracy, instead of the theoretical 6. In any case, the first 3 non-null singular values are smaller than the rest. We attribute this excess in rank to the thin-plate approximation with respect to the 3D solid FEM formulation. For this reason, the 6 rank deficiency has to be enforced. The 6 rank deficiency enforcement has proven to be a key factor for proper experimental performance.

Stiffness matrix \mathbf{K} rank is enforced to $r = 3n - 6$, where n is the total number of nodes, by means of a Singular Value Decomposition (SVD) exploiting the fact that \mathbf{K} is symmetric:

$$\mathbf{K} = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^\top. \quad (11)$$

Hence the pseudoinverse is computed as:

$$\mathbf{K}^+ = \mathbf{U}_r \mathbf{\Sigma}^+ \mathbf{U}_r^\top, \quad (12)$$

where \mathbf{U}_r containing the first r columns of \mathbf{U} and $\mathbf{\Sigma}^+$ is a diagonal matrix composed of the first r singular values inverses:

$$\mathbf{\Sigma}^+ = \text{diag} \left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_r} \right). \quad (13)$$

5. Cameracentric non-rigid EKF

This section is devoted to embedding the FEM free of boundary points modelling for non-rigid scenes within the sequential EKF estimation.

As no point in the scene can be assumed as static, only the relative pose of the non-rigid scene with respect to the camera can be estimated. For this reason, we propose a robocentric EKF-based formulation [5], hence on we call it cameracentric.

5.1. State Vector Definition

The state vector:

$$\mathbf{x}_k^{C_k} = \left(\mathbf{x}_{rk}^{C_k \top}, \mathbf{y}_k^{C_k \top} \right)^\top, \quad (14)$$

is composed of the camera pose $\mathbf{x}_{rk}^{C_k}$ and the n map point locations $\mathbf{y}_k^{C_k} = \left(\mathbf{y}_{1k}^{C_k \top}, \dots, \mathbf{y}_{nk}^{C_k \top} \right)^\top$, coded as camera-centric, i.e. all of them are expressed in the camera frame coordinate system using C_k superscript.

$\mathbf{x}_{rk}^{C_k}$ models the pose that includes the combined effects of the camera and scene rigid motions in 6 d.o.f. (Fig. 1(d)). It is proposed to be modelled as a constant velocity model so the state vector is composed of the pose and the corresponding velocity vectors:

$$\mathbf{x}_{rk}^{C_k} = \left(\mathbf{r}_k^{C_k \top}, \mathbf{q}_k^{C_k \top}, \mathbf{v}_k^{C_k \top}, \omega_k^{C_k \top} \right)^\top, \quad (15)$$

where $\mathbf{r}_k^{C_k}$ is the translation, $\mathbf{q}_k^{C_k}$ is the quaternion representing orientation, $\mathbf{v}_k^{C_k}$ and $\omega_k^{C_k}$ are linear and angular velocities. We assume that linear and angular accelerations \mathbf{a}^C and α^C affect the pose, producing at each step an impulse of linear velocity, $\mathbf{V}^C = \mathbf{a}^C \Delta t$, and angular velocity $\Omega^C = \alpha^C \Delta t$, with a zero-mean Gaussian distribution being $\mathbf{Q}_{\mathbf{x}_r}$ its covariance.

The state equation for the pose is:

$$\mathbf{g}_r = \begin{pmatrix} \mathbf{r}_{k+1}^{C_k} \\ \mathbf{q}_{k+1}^{C_k} \\ \mathbf{v}_{k+1}^{C_k} \\ \omega_{k+1}^{C_k} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_k^{C_k} + (\mathbf{v}_k^{C_k} + \mathbf{V}^C) \Delta t \\ \mathbf{q}_k^{C_k} \times \mathbf{q} \left((\omega_k^{C_k} + \Omega^C) \Delta t \right) \\ \mathbf{v}_k^{C_k} + \mathbf{V}^C \\ \omega_k^{C_k} + \Omega^C \end{pmatrix}, \quad (16)$$

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

The state equation for the non-rigid scene is:

$$\mathbf{g}_y = \mathbf{y}_{k+1}^{C_k} = \mathbf{y}_k^{C_k} + \mathbf{K}_k^+ \Delta \mathbf{S}^C. \quad (17)$$

where, the structure non-rigidity is coded by means of the *compliance* matrix $\mathbf{K}_k^+ \left(\hat{\mathbf{y}}_{k-1|k-1}^{C_k} \right)$, that depends on the current structure geometry estimate $\hat{\mathbf{y}}_{k-1|k-1}^{C_k}$. The incremental non-rigid displacement caused by the Gaussian set of forces, \mathbf{a}_{nr} (Fig. 1 (d)), is proposed to be modelled as the particular solution Eq. (10) to the FEM linear system.

The vector of *normalized forces*, $\Delta \mathbf{S}^C$, is causing recursively at each step an incremental deformation. We assume $\Delta \mathbf{S}^C$ follows a zero-mean Gaussian distribution being \mathbf{Q}_y its covariance.

The normalized forces:

$$\Delta \mathbf{S}_i^C = \frac{1}{Eh} \left(\Delta f_{xi}^C, \Delta f_{yi}^C, \Delta f_{zi}^C \right)^\top, \quad (18)$$

are defined to concentrate most the material tuning parameters in the state noise vector, being h the surface thickness. However, \mathbf{K} still keeps a dependency on a h^2 factor, because the h influence cannot be completely factorized out.

On the one hand, we propose to tune \mathbf{Q}_{y_i} as a diagonal matrix, where the standard deviation codes the tangential deformation, measured in length units, caused by 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 h^2 codes this anisotropy.

It should be noted that Eqs. (16) and (17) are state transition equations, where both the camera and scene states are dynamic. It contrasts with the classical rigid SLAM case where only the camera –a small fraction of the state– is dynamic.

5.2. EKF Formulation

To sum up, the state equations are Eq. (16) and Eq. (17), and the corresponding Jacobians for the EKF are:

$$\mathbf{F}_k = \begin{pmatrix} \frac{\partial \mathbf{g}_r}{\partial \mathbf{x}_r} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{g}_y}{\partial \mathbf{y}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{I} \Delta t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{q}_{k+1}^{C_k}}{\partial \mathbf{q}_k^{C_k}} & \mathbf{0} & \frac{\partial \mathbf{q}_{k+1}^{C_k}}{\partial \omega_k^{C_k}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}, \quad (19)$$

$$\mathbf{G}_k = \begin{pmatrix} \frac{\partial \mathbf{g}_r}{\partial \mathbf{n}} \\ \frac{\partial \mathbf{g}_y}{\partial \mathbf{n}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} \Delta t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \mathbf{q}_{k+1}^{C_k}}{\partial \Omega^C} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{K}_k^+ \end{pmatrix}, \quad (20)$$

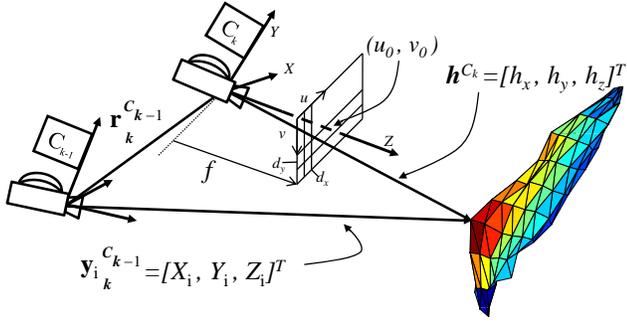


Figure 3. Measurement equation in cameracentric EKF estimation.

where $\mathbf{n} = (\mathbf{a}^{C^T} \quad \alpha^{C^T} \quad \Delta \mathbf{S}^{C^T})^T$ is the state noise vector.

5.3. Measurement Equation

Each observed feature imposes a constraint between the camera location and the corresponding map feature (Fig. 3). The observation of a point $\mathbf{y}_i^{C_{k-1}} = (X_i, Y_i, Z_i)^T$ defines a ray coded by a directional vector $\mathbf{h}^{C_k} = (h_x \quad h_y \quad h_z)^T$ in the camera frame C_k :

$$\mathbf{h}^{C_k} = \mathbf{R}_{k-1}^{C_k} \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} - \mathbf{r}_k^{C_{k-1}}, \quad (21)$$

where $\mathbf{R}_{k-1}^{C_k}$ is the rotation matrix corresponding to $\mathbf{q}_k^{C_{k-1}}$. The camera does not directly observe \mathbf{h}^{C_k} but its projection in the image according to the pinhole model. Projection to a normalized retina and then camera calibration is applied:

$$\mathbf{h} = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u_0 - \frac{f}{d_x} \frac{h_x}{h_z} \\ v_0 - \frac{f}{d_y} \frac{h_y}{h_z} \end{pmatrix}, \quad (22)$$

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.

5.4. Data Association

Data association is embedded within the EKF prediction-update loop:

1. A texture patch around every map point is stored at point creation.
2. EKF prediction step provides a prediction estimate in the image for every map point along with its innovation covariance.

parameter (units)	Silicone	Paper Bending	Paper Flag
h (m)	0.0015	0.0001	0.0001
ν	0.499	0.499	0.499
$\frac{\Delta f}{Eh}$ (m) (std.)	$1.5 \cdot 10^{-5}$	$2.0 \cdot 10^{-9}$	$2.0 \cdot 10^{-9}$

Table 1. Parameter tuning.

3. Prediction and innovation covariance define a gated elliptical acceptance region in the image where is exhaustively search by normalized correlation with the corresponding map point texture patch. In our experiments, the significance level is fixed at 95% (Fig. 4)
4. Update correction for the estimates and their corresponding covariances are computed from the matched map points. It is worth noting that all map points are updated even those not matched in the image.

5.5. Computational Cost

EKF estimation, if all the map points are detected in all the images, is $O(n^3)$, where n is the state vector size. The proposed algorithm combining EKF and FEM needs the SVD computation on top of the standard EKF operation. Symmetric SVD is an $O(12n^3)$ algorithm [11] so the total complexity of the algorithm is kept at same $O(n^3)$ order. It has to be noted that stiffness matrix is sparse so in a practical implementation an important computation overhead reduction can be achieved if this sparsity pattern is exploited.

6. Experimental Results

The proposed free boundary point algorithm has been experimentally validated with real image sequences $320 \times 240@30\text{Hz}$ in three real sequences¹. Data association is completely automatic after selecting the non-rigid feature set in the first image. In all the experiments, the first sequence frames correspond to a mobile camera observing a static scene in order to estimate the structure at rest by means of standard rigid EKF SLAM.

The first experiment (Sec. 6.1, Fig. 4) corresponds to a silicone on a stretcher. The tuning allows to code the scene non-isometric deformation. A quantitative comparison with respect to fixed boundary points and ground truth is provided.

The other two experiments correspond to a deforming piece of paper (Sec. 6.2, Figs. 5, 6). In both cases, the scene deformation is modelled as isometric –geodesic distance between mesh points is constant– just by tuning (Table 1): the thin-plate thickness h is low and the normalized forces $\frac{\Delta f}{Eh}$ are low, so the corresponding deformation is approximately isometric.

¹Videos of the experimental results can be found on website <http://webdiis.unizar.es/~josemari>

6.1. Multiply Deformed Silicone Sequence

The purpose is to quantitatively validate the computed structure and pose with respect to ground truth. The proposed free boundary point algorithm is compared with respect to the fixed boundary point algorithm [1].

A waving hand-held stereo rig observes a silicone cloth fixed on a stretcher while four fingers elastically deform the silicone surface (see Fig. 4 and `silicone.avi`). The stereo pair is used to compute scene ground truth at 640×480 resolution for selected frames. Every frame in the left camera monocular sequence, at half resolution 320×240 , is processed by the proposed algorithm.

It has to be stressed that the matches have been computed automatically without using the codes associated with the markers. The coded markers are used only to compute the stereo ground truth. It has to be noted, that some of the points are not detected in some of the images, mainly because of the patch deformation that make the correlation based matching approach to produce false negatives. However, this matching lost does not affect the estimation process because all the structure points are correlated and their estimation is improved indirectly by the measurements corresponding to other successfully matched mesh points.

The EKF provides, *per frame*, an estimation for the structure along with its covariance. The quantitative comparison (Fig. 4) shows that for most of the structure points, the ground truth is within the 95% acceptance region, verifying the estimate consistency. The free boundary point case exhibits bigger covariance at boundary points than the fixed case as expected. No significant differences in consistency can be appreciated.

6.2. Paper Sequences

A static camera observes a deforming piece of paper (see Fig. 5 and `paper.bending.avi`). The map points correspond to FAST points detected in the first image, so they correspond to natural landmarks in the scene. The corresponding mesh is composed showing the ability to deal with the irregular triangles defined by the estimated 3D map points. FEM modelling is able to consider different stiffness for every triangle and then consider the joint effect in the total stiffness matrix after assembling.

A waving camera observes a deforming scene (see Fig. 6 and `flag_paper.avi`). As in the previous example, the map points are FAST points detected in the first image.

7. Conclusions and Future Work

It has been proposed to cross-fertilize FEM modelling with sequential Bayesian EKF estimation. Resulting in an algorithm with a distinctive set of qualities: full perspective camera, ability to estimate non-rigid structure and pose for every single frame in the sequence –potentially in real-time

at frame rate–. Unlike most approaches that assume data association as given prior, we compute data association; we do not need to observe all the points in all the images. Compared with methods based on deformation modes, we do not need to enumerate which are the allowed deformation bases. We do not need to constrain the deformations as isometric like in template-based methods. For a given type of scene we just tune two scene parameters: normalized forces and thickness.

The proposed simplistic but cheap FEM model is unable to model real solid deformations by itself. However the feedback structure of the EKF can integrate accurate scene measurements at frame rate, resulting in a system able to produce accurate estimations that exploit the laws of solid mechanic priors at a low computational cost, unlike other previous work in the field that considered more exact but too expensive FEM models. A second unsatisfactory dependency, when using FEM models, has been the boundary conditions, that normally imply identify rigid scene points. This dependency has been completely removed in the current work. So we can conclude that the advantages of FEM based models have been exploited and most of the disadvantages have been alleviated.

BA approaches need to estimate the full scene structure per each considered keyframe what increases complexity with the number of images. In contrast, our sequential approach only estimates the scene at the current frame thanks to dynamic state estimation capabilities of the EKF, avoiding the growth in complexity with the number of processed frames.

The obvious future work is to verify experimentally the real-time performance. In a longer term, this method can be quite profitable to deal with medical imagery where accurate FEM models are readily 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. Agudo, B. Calvo, and J. M. M. Montiel. FEM models to code non-rigid EKF monocular SLAM. In *1st IEEE Workshop on Dynamic Shape Capture and Analysis*, 2011. 1, 2, 3, 6
- [2] 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. 2
- [3] K. J. Bathe. *Finite element procedures in Engineering Analysis*. Prentice-Hall, 1982. 3

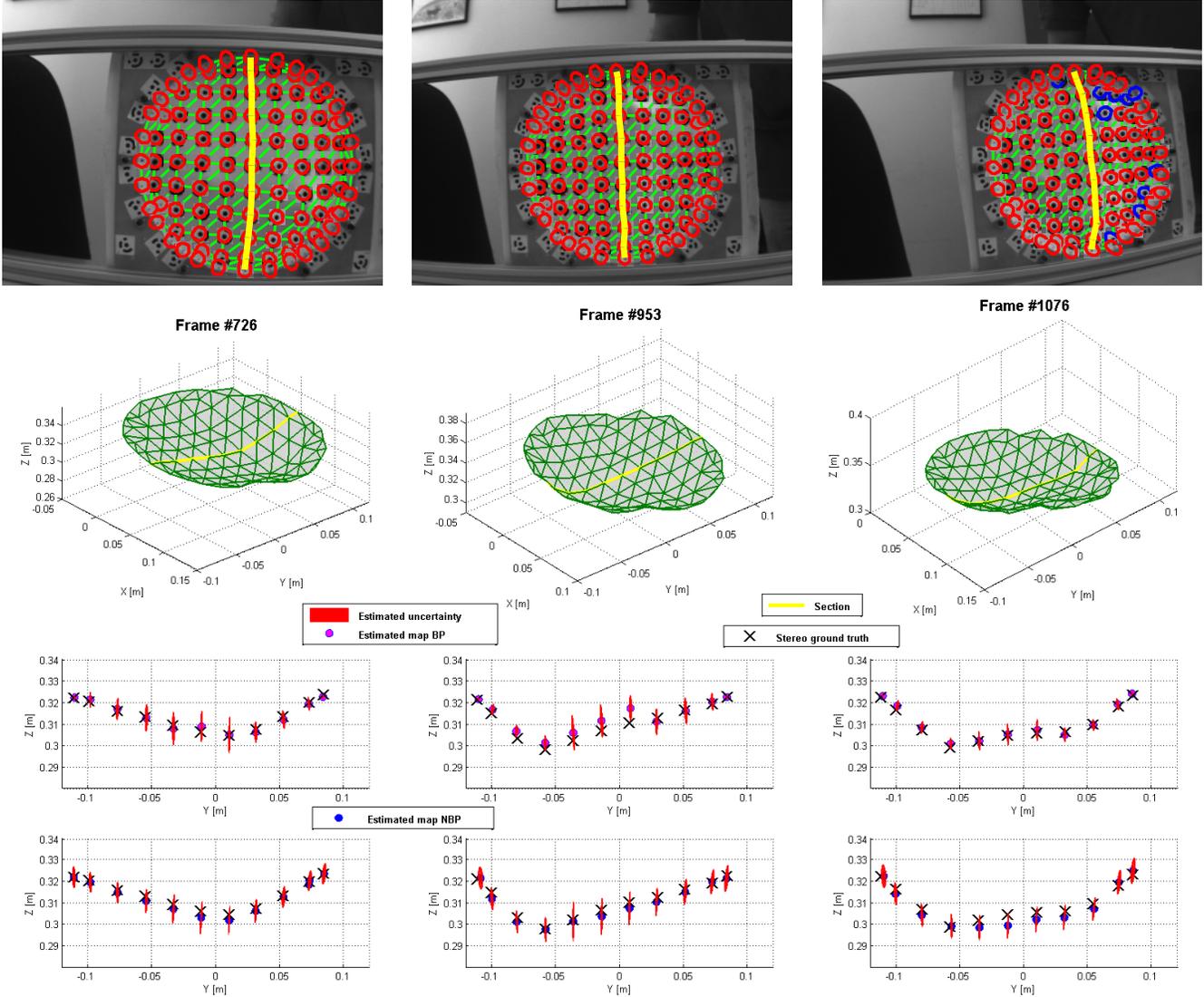


Figure 4. **Multiply deformed silicone sequence.** **Top:** Images gathered by the camera at three selected frames. Elliptical matching acceptance regions: red (predicted&matched), blue (predicted¬ matched). **Middle:** General view of the 3D reconstructed non-rigid scene. **Bottom:** Cross section of the reconstructed surface; it is represented the estimated points with the 95% acceptance regions according to the estimated covariances. Comparison of free boundary (below) with respect to fixed boundary (above). The estimation is mostly consistent because ground truth is within the acceptance region. Best viewed in color.

- [4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *CVPR*, 2000. 2
- [5] J. A. Castellanos, J. Neira, and J. D. Tardós. Limits to the consistency of EKF-based SLAM. In *5th IFAC Symposium on Intelligent Autonomous Vehicles*, 2004. 2, 4
- [6] J. Civera, A. J. Davison, and J. M. M. Montiel. Inverse depth parametrization for monocular SLAM. *TRO*, 2008. 1
- [7] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *ICCV*, 2003. 1
- [8] A. Del Bue, X. Llado, and L. Agapito. Non-rigid metric shape and motion recovery from uncalibrated images using priors. In *CVPR*, 2006. 2
- [9] J. Fayad, L. Agapito, and A. Del Bue. Piecewise quadratic reconstruction of non-rigid surfaces from monocular sequences. In *ECCV*, pages 297–310, 2010. 2
- [10] J. Fayad, A. Del Bue, L. Agapito, and P. Aguiar. Non-rigid structure from motion using quadratic deformation models. In *BMVC*, 2009. 2
- [11] G. Golub and C. Van Loan. *Matrix computations*, volume 3. Johns Hopkins Univ Pr, 1996. 3, 5
- [12] M. A. Greminger and B. J. Nelson. Deformable object tracking using the boundary element method. In *CVPR*, 2003. 2
- [13] S. Ilić and P. Fua. Non-linear beam model for tracking large deformation. In *ICCV*, October 2007. 2

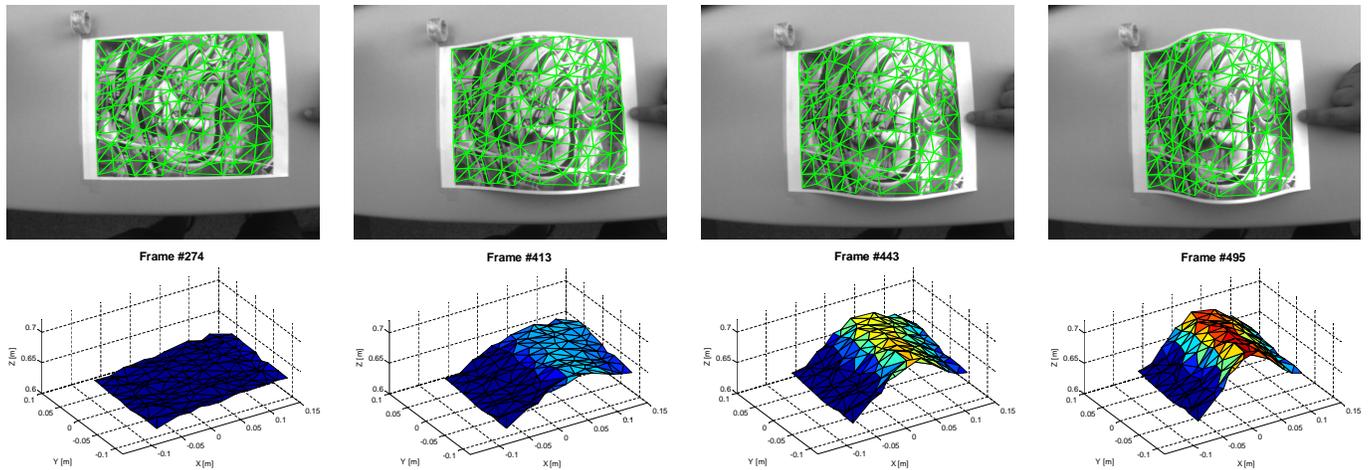


Figure 5. **Paper bending sequence.** **Top:** Images gathered by the camera at four selected frames 3D mesh has been reprojected. **Bottom:** General view of the 3D reconstructed deformed scene. Best viewed in color.

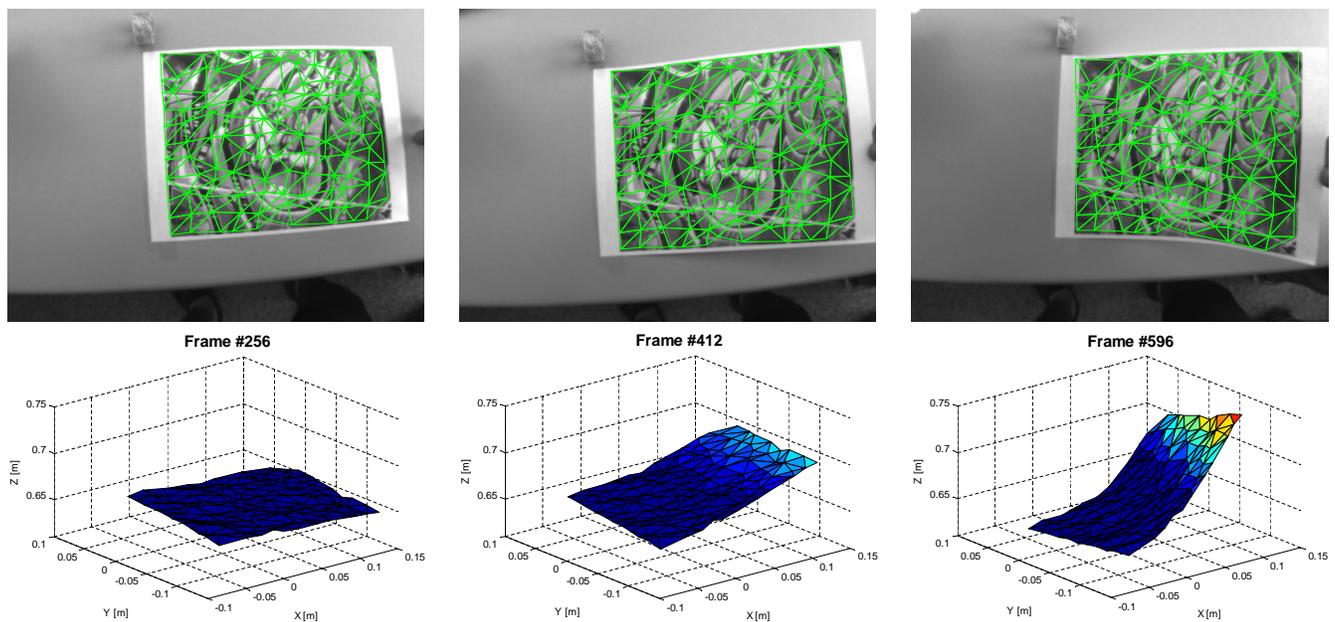


Figure 6. **Flag paper sequence.** **Top:** Images gathered by the camera at three selected frames 3D mesh has been reprojected. **Bottom:** General view of the 3D reconstructed deformed scene. Best viewed in color.

- [14] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1988. 2
- [15] G. Klein and D. W. Murray. Parallel tracking and mapping for small AR workspaces. In *ISMAR*, 2007. 1
- [16] D. T. D. Metaxas. Constrained deformable superquadrics and nonrigid motion tracking. *PAMI*, 1993. 2
- [17] F. Moreno-Noguer and J. M. Porta. Probabilistic simultaneous pose and non-rigid shape recovery. In *CVPR*, 2011. 2
- [18] M. Paladini, A. Bartoli, and L. Agapito. Sequential non rigid structure from motion with the 3D implicit low rank shape model. In *ECCV*, 2010. 2
- [19] M. Perriollat, R. Hartley, and A. Bartoli. Monocular template-based reconstruction of inextensible surfaces. *IJCV*, 2010. 2
- [20] M. Salzmann and P. Fua. Reconstructing sharply folding surfaces: A convex formulation. In *CVPR*, 2009. 2
- [21] J. Taylor, A. D. Jepson, and K. N. Kutulakos. Non-rigid structure from locally-rigid motion. In *CVPR*, 2010. 2
- [22] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid structure-from motion: estimating shape and motion with hierarchical priors. *PAMI*, 2008. 2
- [23] A. Varol, M. Salzmann, E. Tola, and P. Fua. Template-free monocular reconstruction of deformable surfaces. In *ICCV*, 2009. 2
- [24] 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. 3