Force-based Representation for Non-Rigid Shape and Elastic Model Estimation

Antonio Agudo and Francesc Moreno-Noguer

Abstract—This paper addresses the problem of simultaneously recovering 3D shape, pose and the elastic model of a deformable object from only 2D point tracks in a monocular video. This is a severely under-constrained problem that has been typically addressed by enforcing the shape or the point trajectories to lie on low-rank dimensional spaces. We show that formulating the problem in terms of a low-rank force space that induces the deformation and introducing the elastic model as an additional unknown, allows for a better physical interpretation of the resulting priors and a more accurate representation of the actual object's behavior. In order to simultaneously estimate force, pose, and the elastic model of the object we use an expectation maximization strategy, where each of these parameters are successively learned by partial M-steps. Once the elastic model is learned, it can be transfered to similar objects to code its 3D deformation. Moreover, our approach can robustly deal with missing data, and encode both rigid and non-rigid points under the same formalism. We thoroughly validate the approach on Mocap and real sequences, showing more accurate 3D reconstructions than state-of-the-art, and additionally providing an estimate of the full elastic model with no a priori information.

Index Terms—Non-Rigid Structure from Motion, 3D Reconstruction, Expectation Maximization, Elastic Model, Force Space.

✦

1 INTRODUCTION

THe aim of the Non-Rigid Structure from Motion (NRSfM) is to simultaneously recover the camera motion and to reconstruct the 3D shape of a deformable object from 2D point observations in a monocular video. It is known to be a severely under-constrained problem, since many different shapes can have very similar image observations. The problem is even more challenging when observations contain 2D noise or missing data (see examples of missing points due to self-occlusion in Fig. 1). In order to make this ambiguous problem solvable, it is standard to assume that objects do not arbitrarily deform their shape, and that they obey certain 'statistical' rules. Along this idea, early approaches extended the rigid factorization algorithm [50] to the non-rigid domain [11], [16], [52], and approximated the shape by a linear combination of basis estimated on-the-fly. Alternatively, other approaches have represented the temporal evolution of each object point through a set of pre-defined trajectories [9], [41], [54]. However, both these constraints are statistical priors that do not have a direct physical interpretation.

In this paper, we introduce a new constraint based on a low-rank force prior. This prior has a direct physical interpretation, as it models the interaction between the object and the underlying forces that deform it, which are modulated by the elastic model of the material. Our rationale is that if certain deformation patterns can be observed, this is because the underlying forces producing these deformations also obey some patterns.

• The authors are with the Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Barcelona, 08028, Spain. Email: {aagudo, fmoreno}@iri.upc.edu.
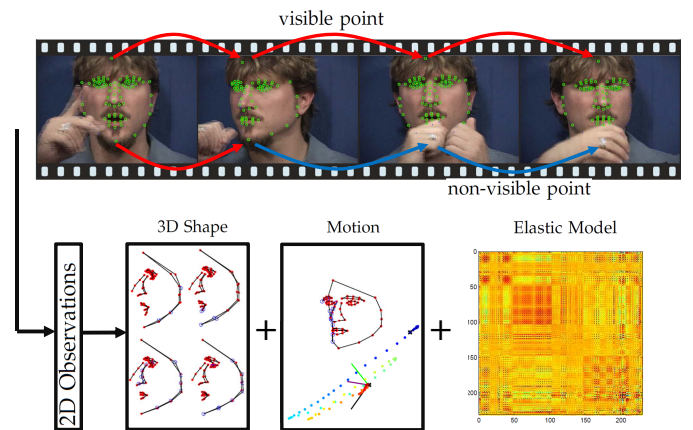

Fig. 1. **Force-based NRSfM.** The proposed algorithm allows to simultaneously recover the 3D non-rigid shape, camera motion and the full elastic model, from a sequence of 2D point trajectories acquired with a monocular camera. Self-occlusions (see the track annotated as "non-visible") can also be naturally handled with our formulation.

The intuition behind our approach is described in Fig. 2. Let us consider N points on the object, which is deformed under the action of external forces. Following continuum mechanics, the relation between the acting forces and the deformation field can be characterized by an elastic model. Regarding the force space, we can fully define it by $3N$ independent forces, whose combination allows mapping the shape from a rest configuration to a wide variety of arbitrary arrangements. However, only a few of these forces, conforming a low-rank force space, are indeed necessary to represent realistic deformations. Based on this idea, we propose a new formulation of the NRSfM problem in which, given 2D point tracks, we estimate camera trajectory and force parameters
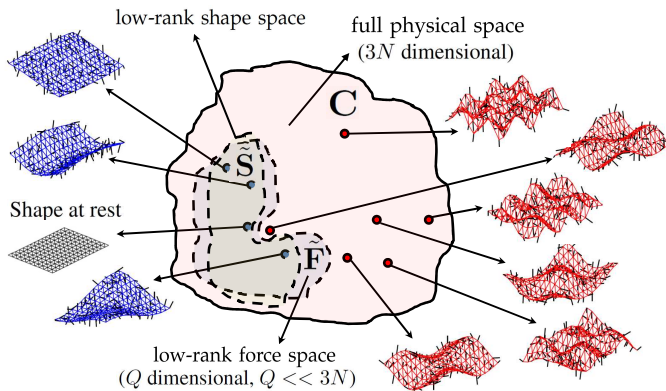
Fig. 2. **Intuition behind our approach.** The non-rigid shape with N vertexes can be encoded in terms of its underlying elastic model $\mathbf{C}$ (compliance matrix) and the force field $\mathbf{f}$ acting on it. In turn, the full force field can be approximated by a low-rank basis $\tilde{\mathbf{F}}$. In this work, we simultaneously learn both the elastic model and the low-rank force space, while recovering shape and camera motion. The figure shows the full force-space and its corresponding shapes in red together with force vectors. The low-rank force and the corresponding shapes are shown in blue; and a tentative shape subspace $\tilde{\mathbf{S}}$ in green.

(and consequently shape). Even though reasoning on the force space introduces the compliance matrix as new unknown, we propose an Expectation Maximization (EM) strategy, initialized with the shape obtained from a rigid shape from motion, that allows to simultaneously solve for all parameters. By thorough testing on Mocap and real sequences we show that our formulation yields more accurate reconstructions than state-of-the-art methods, while providing more physical insights in terms of the elastic model. Interestingly, we will also show the connection of the force prior we propose with previous shape, trajectory and shape-trajectory models, turning these, into physical priors too.

A preliminary version of this work was presented in [6], in which we showed our approach to be suitable for simultaneously recovering shape, pose and elastic properties from 2D trajectories of non-rigid points. In this paper, we extend the method in such a way it can reason from the observation of both rigid and non-rigid points. This is a typical situation encountered in practice (e.g., in a non-rigid tissue attached to a rigid body, a flag on a mast or in the regions of a human face). Additionally, besides extending the battery of results to further emphasize the advantages of our approach, we also perform experiments in which the estimated elastic model is transferred between different objects.

2 RELATED WORK

The inherent ambiguity of the NRSfM problem is commonly tackled by constraining the shape to lie on a low-rank space spanned by a set of deformation modes [11], [52], [59]. This is further constrained by enforcing spatial [52] or temporal [11], [20] shape smoothness, by imposing the 3D shapes to be closely aligned [31], [32],

or by means of a union of low-rank shape subspaces [61]. Alternatively, low-rank shape constraints can be imposed through global [23] or local [22], [44] quadratic models over a rest configuration. [18] proposed a direct rank minimization of the 3D shape matrix to impose temporal smoothness, that was later used in [24] for dense reconstruction. These constraints were also exploited in the variational version of the problem [25].

On the other hand, instead of constraining the shape, a number of approaches introduce restrictions on the trajectory of every object point using pre-defined bases [9], [54]. The problem was even further simplified in [41], where additional static points were used to independently solve for the camera motion, posing finally a linear problem. There have also been recent attempts to combine low-rank shape and trajectory spaces [27], [28], [47]. All these techniques are referred to as statistically-based methods, since the low-rank representations used to condition the problem are not physically grounded. Despite their popularity, one inherent limitation of these methods is that they can become very sensitive to the number of shape or trajectory modes, which needs to be carefully chosen to correctly model the deformation.

A better representation of the underlying dynamics involved in non-rigid deformations can be obtained through physically-grounded models [37], [46]. Force-based kinematics [7], [17], [45], inextensibility-based deformations [55], linear elastic models [35], [34], and numerical techniques based on Finite Element Methods (FEM) for tracking [58] or 3D reconstruction [2], are just a few examples of the renewed interest in physical models. A standard assumption when using physical models is to assume the deformation model, material properties and point connectivity to be known a priori [4], [10], [42]. On the other hand, there exist approaches in which the parameters ruling these models are learned from input data. For instance, displacement and force measurements allow recovering the Young's modulus [60] together with the Poisson's ratio [13]. In [48], these parameters are sequentially estimated from only image sequences. Elastic and viscosity properties are obtained, also from video, in [21]. More recently, material properties of fabrics moving under wind forces [15] or under small motions [19] are estimated from only video sequences. And vice-versa, applied forces can be recovered from 2D displacements and an estimate, up to scale, of the elastic parameters [2]. However, in all these approaches only small pieces of the full physical model (i.e., the complete stiffness matrix) are recovered. In contrast, we learn this matrix with no prior information, and without imposing isometric constraints [55].

In this paper we propose a new low-rank force model to simultaneously recover camera motion, 3D shape and the full elastic model of the object. Note that the latter is specially challenging, as it involves estimating a large number of parameters (a $3N \times 3N$ matrix, for an object with N points), and not just the material properties such as the Young's modulus or Poisson's ratio. Our

approach can also uniformly combine rigid and non-rigid points, if their typology is known a priori, to learn an elastic model that better reflects the true behavior of the object. Once the elastic model is learned, it can be transfered to similar objects for coding its behavior. We do all this from the sole input of 2D input tracks in a monocular video, which may even be corrupted by noise and missing data, and without the need of any training data. In addition, we link our physical model to previous shape, trajectory and shape-trajectory statistical approaches, giving them a physical interpretation, too.

3 LOW-RANK FORCE MODEL

A standard approach to reduce the ambiguity of the NRSfM problem involves representing the object in low dimensional spaces. Two subspaces have been considered so far, the shape and the trajectory ones, and one combination of them denoted as shape-trajectory. Although, both of them represent alternate ways of looking at the deformable shape, the force basis has a direct physical interpretation, and introduces into the whole equation the elastic model of the object. Before describing the new low-rank force space we propose, we review the previous formulations.

3.1 Low-rank Shape and Trajectory Space

Time-varying shapes can be represented by low-rank shape bases. These priors are computed using Principal Component Analysis (PCA) over training data [14], [38], applying modal [10], [42] or spectral [5] analysis over a rest configuration, or they are estimated on-the-fly [16], [25], [40], [52]. In particular, let us consider N 3D points on an object, being observed along T frames. If we denote by $\mathbf{x}_i^t = [x_i^t, y_i^t, z_i^t]^\top$ the 3D coordinates of the i-th point at time t, and by $\mathbf{s}^t = [(\mathbf{x}_1^t)^\top, \ldots, (\mathbf{x}_N^t)^\top]^\top$ the $3N$-dimensional representation of the shape, we can compactly write the time-varying shape as a $3N \times T$ matrix $\mathbf{S} = [\mathbf{s}^1, \ldots, \mathbf{s}^T]$. Every instant shape $\mathbf{s}^t$ may be approximated by linearly combining Q basis shapes $\tilde{\mathbf{s}}_q$:

$$\mathbf{s}^t = \sum_{q=1}^{Q} \psi_q^t \tilde{\mathbf{s}}_q = \tilde{\mathbf{S}} \boldsymbol{\psi}^t, \tag{1}$$

where $\boldsymbol{\psi}^t = [\psi_1^t, \ldots, \psi_Q^t]^\top$ are the coefficients for the shape at time t, and $\tilde{\mathbf{S}} = [\tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_Q]$ is a $3N \times Q$ matrix containing all basis shapes. By aggregating all coefficients into a $Q \times T$ matrix $\boldsymbol{\Psi} = [\boldsymbol{\psi}^1, \ldots, \boldsymbol{\psi}^T]$, we can finally write the factorization of the time-varying shape $\mathbf{S}$ as:

$$\mathbf{S} = \tilde{\mathbf{S}} \boldsymbol{\Psi}. \tag{2}$$

Alternatively, we may include a rest shape $\mathbf{s}_0$ in the subset of basis shapes [1], [52]. In that case, we would take $\hat{\mathbf{S}} = [\mathbf{s}_0, \tilde{\mathbf{s}}_1, \ldots, \tilde{\mathbf{s}}_Q]$, and the basis vectors $\tilde{\mathbf{s}}_i$ with $i = \{1, \ldots, Q\}$ would be interpreted as 3D displacements over $\mathbf{s}_0$, i.e., with coefficients $\hat{\boldsymbol{\psi}}^t = [1, \psi_1^t, \ldots, \psi_Q^t]^\top$. Note that although this formulation does not reduce the total

number of parameters to estimate, the use of the shape at rest helps to regularize the problem.

When representing the deformable shape in trajectory space [9], each time-varying point (i.e., the rows of $\mathbf{S}$) is represented in the linear space of trajectories spanned by a trajectory basis. This basis can be pre-defined in an object independent way, for instance, using the Discrete Cosine Transform (DCT). We can then factorize $\mathbf{S}$ as:

$$\mathbf{S} = \boldsymbol{\Phi} \tilde{\mathbf{T}}, \tag{3}$$

where $\tilde{\mathbf{T}}$ is a $Q \times T$ matrix of Q pre-defined basis trajectories, and $\boldsymbol{\Phi}$ is a $3N \times Q$ matrix of trajectory coefficients.

For completeness, we also consider the combined shape-trajectory model proposed in [28]. In this case, the time-varying shape matrix can be modeled with Q shape vectors in $\tilde{\mathbf{S}}$ and R trajectory ones in $\tilde{\mathbf{T}}$. The factorization can be written as:

$$\mathbf{S} = \tilde{\mathbf{S}} \boldsymbol{\Omega} \tilde{\mathbf{T}}, \tag{4}$$

where $\boldsymbol{\Omega}$ is a $Q \times R$ matrix of shape-trajectory coefficients. Note that the shape coefficients in Eq. (2) are now modeled in terms of trajectory basis in order to implicitly produce smooth 3D shape deformations.

3.2 Modeling Shapes in a Low-rank Force Space

We next derive the formulation of our physics-based low-rank force model to represent the shape. We draw inspiration on the Hooke's law, which states that the force needed to extend or compress a spring by a certain distance is proportional to that distance by a factor k, known as stiffness. This simple model can be generalized to 3D objects with mass and volume, resulting in complex systems of partial differential equations [12] that typically do not have an analytical solution and require from numerical approximations, such as those of FEMs. For instance, applying FEM over a shape at rest, made of N points and represented as a 3N-dimensional vector $\mathbf{s}_0$, yields the following linear system:

$$\mathbf{Ku} = \mathbf{f}, \tag{5}$$

where $\mathbf{K}$ is the $3N \times 3N$ stiffness matrix that maps the $3N$ displacement vector $\mathbf{u}$ into a $3N$-dimensional force field $\mathbf{f}$. The matrix $\mathbf{K}$ is usually built considering a number of physical characteristics, such as material elastic properties, the type of deformation (e.g., beam bending, stress plane) and the connectivity between the nodal points, which depends on the type of element discretization (e.g., triangular, wedge, tetrahedral). Additionally, unless providing boundary conditions, $\mathbf{K}$ is not full rank, i.e., $\text{rank}(\mathbf{K}) < 3N$.

Note that Eq. (5) allows computing the forces $\mathbf{f}$ that need to be applied onto every point of $\mathbf{s}_0$ to obtain a pre-defined displacement $\mathbf{u}$. However, we will regard this relation in the opposite direction, that is, we seek to compute the 3D displacement when the 3D acting forces are known. In this case, we will apply the relation $\mathbf{u} = \mathbf{Cf}$, where $\mathbf{C}$ is a $3N \times 3N$ compliance matrix.
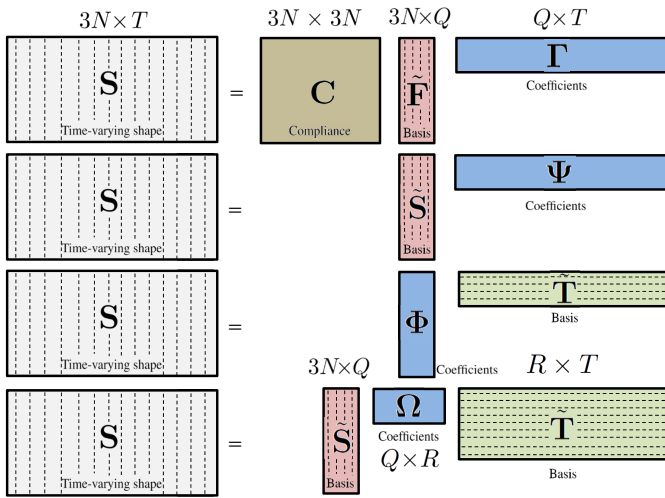
Fig. 3. **Equivalence between Shape, Trajectory and Force low-rank models.** The matrix $\mathbf{S}$ of temporal shapes shown on the left is approximated using four low-rank spaces: Force, Shape, Trajectory and Shape-Trajectory (from top to bottom). The dotted lines, represent the arrangement of the shapes and bases within the matrices. The low-rank force model (top row), incorporates the compliance matrix $\mathbf{C}$ to encode the full elastic physical model, which by direct comparison with the other three statistical sub-spaces, let us to give them also physical interpretation. Note that for the shape-trajectory model (bottom row), $Q \neq R$ as the number of trajectory vectors is usually larger than the number of shape bases.

When boundary conditions are known this matrix is computed as $\mathbf{C} = \mathbf{K}^{-1}$ [8], [53], and $\mathbf{C}$ is guaranteed to be a strictly positive-definite symmetric matrix. When boundary conditions are not available, we make use of the pseudoinverse, i.e., $\mathbf{C} = \mathbf{K}^{\dagger}$, but we can only assume $\mathbf{C}$ to be symmetric [2].

Once $\mathbf{C}$ is known, we can estimate a 3D displacement $\mathbf{u}$ for any 3D applied force vector $\mathbf{f}$, and therefore a new configuration of the object shape as:

$$\mathbf{s} = \mathbf{s}_0 + \mathbf{u} = \mathbf{s}_0 + \mathbf{C}\mathbf{f} = \mathbf{C}(\mathbf{K}\mathbf{s}_0 + \mathbf{f}) = \mathbf{C}(\mathbf{f}_0 + \mathbf{f}), \quad (6)$$

where $\mathbf{f}_0 = \mathbf{K}\mathbf{s}_0$ can be interpreted as the forces applied to keep the shape at rest. We can now expand this expression to account for all T sequence frames:

$$\mathbf{S} = \mathbf{C}[\mathbf{f}_0 + \mathbf{f}^1, \ldots, \mathbf{f}_0 + \mathbf{f}^T] = \mathbf{C}\mathbf{F}, \quad (7)$$

where $\mathbf{F}$ is a $3N \times T$ matrix made of the force fields along the sequence.

We can now introduce the proposed low-rank force model. As it has been previously done for the shapes and point trajectories, realistic distributions of applied forces can also be approximated by a reduced number of modes. To follow the parallelism with the previous section, we consider a basis made of Q force vectors, and represent our low-rank force field as a $3N \times Q$ matrix $\tilde{\mathbf{F}}$. The time-varying shape can then be written as:

$$\mathbf{S} = \mathbf{C}\tilde{\mathbf{F}}\mathbf{\Gamma}, \quad (8)$$

where $\mathbf{\Gamma} = [\boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^T]$ is a $Q \times T$ matrix of time-varying force coefficients.

3.3 Shape-Trajectory-Force Duality

A direct comparison of the low-rank shape, trajectory, shape-trajectory and force models defined in Equations (2), (3), (4) and (8), respectively, gives the equivalence between the four representations. And most importantly, it gives a relation between two models, the shape and trajectory ones, that have thus far been considered as statistical, and our new low-rank force model, directly derived from physical relations.

More specifically, considering the shape-force duality, we observe that $\tilde{\mathbf{S}} = \mathbf{C}\tilde{\mathbf{F}}$, that is, we can write the linear subspace of shapes in Eq. (2) in terms of force and elasticity parameters, and therefore, the statistical shape model does inherently encode physically-grounded properties. Similarly, we can establish a trajectory-force duality, and write that $\mathbf{\Phi} = \mathbf{C}\tilde{\mathbf{F}}$ and $\tilde{\mathbf{T}} = \mathbf{\Gamma}$. In this case, the low-rank force model is equivalent to the trajectory coefficients, and the low-rank trajectory bases, correspond to the force coefficients. Finally, for the shape-trajectory model we can devise two dualities. If the two subspaces have different rank (i.e., $Q \neq R$), we obtain $\tilde{\mathbf{S}} = \mathbf{C}\tilde{\mathbf{F}}$ and $\mathbf{\Omega}\tilde{\mathbf{T}} = \mathbf{\Gamma}$, that is, a shape-force duality. However, when the rank of both shape-trajectory subspaces are equal, (i.e., $Q = R$), we observe that $\tilde{\mathbf{S}}\mathbf{\Omega} = \mathbf{C}\tilde{\mathbf{F}}$ and $\tilde{\mathbf{T}} = \mathbf{\Gamma}$, yielding a trajectory-force duality. A comparison between low-rank models is displayed in Fig. 3.

Note that while the proposed approach has equal compaction power than shape and trajectory models, factorizing the low-rank space into $\tilde{\mathbf{F}}$ and $\mathbf{C}$, makes it possible to model a much wider range of object behaviors. This factorization, though, introduces an additional complexity in the learning process, as we need to discover all these terms from the sole input of 2D tracks. In the next section, we describe how we resolve this problem. Yet, when this is done, besides estimating shape, we are then also able to solve the inverse problem of estimating the forces necessary to obtain a specific shape configuration. This might be extremely useful, for instance in certain robotic applications that require force control for the manipulation of deformable objects, or in laparoscopy surgery.

4 LEARNING ELASTIC MODEL, SHAPE AND POSE

In this section we describe how we introduce the low-rank force space into the formulation of the NRSfM, and how we then simultaneously solve for the elastic model of the object, plus the shape and camera pose. Following the seminal work in [52] we will formulate the problem in a probabilistic manner, and assume that the 3D shape is drawn from some non-uniform PDF. This has been shown especially robust to avoid overfitting to noise, while allowing to resolve the inherent shape ambiguity.

Factor	Full	Shape	Traj.	Sha-Traj.	Force
Camera	$5T$	$5T$	$5T$	$5T$	$5T$
Basis	-	$3NQ$	-	$3NQ$	$3NQ$
Coefficients	-	QT	$3NQ$	QQ	QT
Model	$3NT$	-	-	-	$3N(3N+1)/2$
Total number of unknowns	$5T$ $+3NT$	$5T+3NQ$ $+QT$	$5T$ $+3NQ$	$5T+3NQ$ $+QQ$	$5T+3NQ+QT$ $+3N(3N+1)/2$

TABLE 1

Total number of unknowns that need to be estimated when considering the full model, or the low-rank models in shape, trajectory, shape-trajectory or force space. The results are represented in terms of the number of object points N, the number of frames T and the dimensionality Q of the low-rank subspace. For simplicity, for the shape-trajectory model, we consider the same rank for both subspaces, i.e., $Q = R$.

N	T	Q	Obs.	Full	Shape	Traj.	Sha-Traj.	Force
55	260	12	28,600	44,200	6,400	3,280	3,424	20,095
40	316	11	25,280	39,500	6,376	2,900	3,021	13,636
29	450	7	26,100	41,400	6,009	2,859	2,908	9,837
41	1,102	10	90,364	141,056	17,760	6,740	6,840	25,386

TABLE 2

Total number of unknowns that need to be estimated when considering the full and the low-rank models, for the combination of parameters N, Q and T we consider in the experimental section. The column "Obs." refers to the number of observed variables, $2NT$, corresponding to the 2D tracks of all N points along the T frames. For all cases, we consider $B = 0$, i.e., no boundary conditions.

4.1 Problem Formulation

Let us consider a deformable object with N points at a time instant t, represented by a $3N$ vector $\mathbf{s}^t$. Assuming an orthographic camera model, we can write the projection of the 3D points onto the image plane as a $2N$ vector $\mathbf{w}^t$:

$$\mathbf{w}^t = \mathbf{G}^t\mathbf{s}^t + \mathbf{h}^t + \mathbf{n}^t, \tag{9}$$

where $\mathbf{G}^t = \mathbf{I}_N \otimes \mathbf{R}^t$ has $2N \times 3N$ size, $\mathbf{I}_N$ is the N-dimensional identity matrix, $\mathbf{R}^t$ are the first two rows of a full rotation matrix, and $\otimes$ denotes the Kronecker product. Similarly, $\mathbf{h}^t = \mathbf{1}_N \otimes \mathbf{t}^t$ is a $2N$ vector resulting from concatenating N times a bidimensional translation vector $\mathbf{t}^t$, and $\mathbf{1}_N$ is a N-vector of ones. Finally, $\mathbf{n}^t$ is a $2N$ dimensional vector of Gaussian noise.

We can therefore define our problem as that of estimating, for $t = \{1, \ldots, T\}$, the shape $\mathbf{s}^t$ and camera pose parameters $\{\mathbf{R}^t, \mathbf{t}^t\}$, given the observation of point tracks $\mathbf{w}^t$ corrupted by noise $\mathbf{n}^t$. The total number of unobserved variables includes $3NT$ parameters for the shape and $5T$ parameters for the pose[1]. Estimating all these unknowns from the only $2NT$ noisy observations of the point tracks is clearly an ill-posed problem. We make the problem tractable by introducing our low-rank force model and encoding the time-varying shape as:

$$\mathbf{s}^t = \mathbf{s}_0 + \mathbf{u}^t = \mathbf{s}_0 + \mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\gamma}^t, \tag{10}$$

where $\mathbf{C}$ is the compliance matrix, $\tilde{\mathbf{F}}$ are the low-rank force vectors, and $\boldsymbol{\gamma}^t$ are the corresponding force coefficients at frame t. The projection Eq. (9) becomes:

$$\mathbf{w}^t = \mathbf{G}^t(\mathbf{s}_0 + \mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\gamma}^t) + \mathbf{h}^t + \mathbf{n}^t. \tag{11}$$

Note that using the low-rank force model introduces a new challenge to the problem, which is that besides having to estimate the variables involved in a standard NRSfM problem (i.e., pose, shape basis and shape coefficients, or equivalently in our framework, pose, force basis and force coefficients), we now need to learn the full elastic model $\mathbf{C}$ of the object.

---

1. An orthographic projection has five degrees of freedom, namely the three parameters describing the rotation matrix, plus two of the translation. Note that the translation is estimated up to depth.

Since $\mathbf{C}$ remains constant along the sequence, it introduces a fixed number of unknowns independently of the number of frames T. Specifically, $\mathbf{C}$ is a $3N \times 3N$ symmetric matrix, for which we only need to estimate the upper triangular part, i.e., $3N(3N+1)/2$ elements. Additionally, we still need to estimate the $5T$ pose parameters, $3NQ$ components for the low-rank force space (assuming we consider a force basis with Q components), and QT unknowns for the force coefficients.

In many real deformations, there may be a number of points in the object which can be considered rigid (w.r.t. the local coordinate system of the object). This knowledge can then be exploited in the form of boundary conditions, and constrain the deformation and the elastic model by reducing the total number of parameters for estimation. For instance, if there exist B of these rigid points the elements of the elastic model to estimate are those of a $3(N-B) \times 3(N-B)$ symmetric matrix, i.e., $3(N-B)\big(3(N-B)+1\big)/2$ elements. Similarly, the number of free components of the low-rank force space becomes $3(N-B)Q$.

In Table 1 we summarize the total number of unknowns as a function of the parameters N (number of points), T (number of frames) and Q (dimensionality of the low-rank space) and for the full-space problem and the four low-rank versions (shape, trajectory, shape-trajectory and force). In Table 2 we give the number of unknowns for the specific combinations of N, Q and T we will use in the experimental section. Observe that for long sequences (T large), the number of unknowns of the shape and force subspaces becomes similar, while our force-based model provides much richer information about the elastic object properties. Our method does not need a large number of points to recover the elastic model, and since we do not use connectivity constraints it can work even with very irregular meshes under the presence of holes (for instance, see results for the ASL sequences). On the negative side, the compliance matrix quadratically depends on the number of points N, increasing the computational cost and preventing our approach to be directly applied for dense reconstruction. To solve this limitation, we could include a coarse-to-fine framework with interpolated functions to transfer the elastic model from sparse to dense objects. Exploring this will be part of our future work.

4.2 Probabilistic Low-Rank Force Model

To simultaneously learn shape, pose and elastic models from 2D point tracks as described in Eq. (11), we follow a Probabilistic PCA formulation [43], [49], [51]. Broadly, this consists of two main steps. We start by writing the observations $\mathbf{w}^t$ as a probabilistic distribution and then we estimate the parameters that maximize its likelihood using EM. We next describe the first of these steps.

In order to estimate the distribution over the projected points $\mathbf{w}^t$ we first assume the weight coefficients $\boldsymbol{\gamma}^t$ to be modeled by a zero-mean Gaussian distribution $\boldsymbol{\gamma}^t \sim \mathcal{N}(\mathbf{0}; \mathbf{I}_Q)$. These weights become latent variables that can be marginalized out and are never explicitly computed, and using Eq. (10), we can propagate their distribution to the time-varying shapes, yielding $\mathbf{s}^t \sim \mathcal{N}\left(\mathbf{s}_0; \tilde{\mathbf{C}}\tilde{\mathbf{F}}\tilde{\mathbf{F}}^\top \mathbf{C}^\top\right)$. By also assuming the noise over the shape observations $\mathbf{n}^t$ to follow a Gaussian distribution with variance σ^2, i.e., $\mathbf{n}^t \sim \mathcal{N}(\mathbf{0}; \sigma^2 \mathbf{I}_{2N})$, we can finally estimate that the projected points $\mathbf{w}^t$ are also Gaussian:

$$\mathbf{w}^t \sim \mathcal{N}\left(\mathbf{G}^t\mathbf{s}_0 + \mathbf{h}^t; \mathbf{G}^t\mathbf{C}\tilde{\mathbf{F}}(\mathbf{G}^t\mathbf{C}\tilde{\mathbf{F}})^\top + \sigma^2\mathbf{I}_{2N}\right). \quad (12)$$

We next explain how we perform Maximum Likelihood Estimation (MLE) on this latent variable problem using EM.

4.3 Expectation Maximization

For the purpose of estimating the MLE of the distribution in Eq. (12), we use an EM algorithm in a similar way as done in [3], [51]. We denote by $\boldsymbol{\Theta}^t \equiv \{\mathbf{R}^t, \mathbf{t}^t\}$ the set of model parameters to estimate per frame, $\boldsymbol{\Upsilon} \equiv \{\mathbf{C}, \tilde{\mathbf{F}}, \sigma^2\}$ the set of parameters to estimate along the sequence, $\boldsymbol{\gamma}^t$ the latent variables and $\mathbf{w}^t$ the observed data. Given the 2D trajectories of all points $\mathbf{w} = \{\mathbf{w}^1, \ldots, \mathbf{w}^T\}$, we seek to estimate all set of parameters $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}^1, \ldots, \boldsymbol{\Theta}^T, \boldsymbol{\Upsilon}\}$. The EM algorithm iteratively estimates the maximum likelihood alternating between E- and M-step.

4.3.1 E-Step

We initially estimate the posterior distribution over the latent variables given the current observations and model parameters. Assuming independent and identically distributed random samples, and applying the Bayes' rule and the Woodbury's matrix identity [57], it can be shown this distribution to be:

$$p(\boldsymbol{\gamma}^t | \mathbf{w}^t, \boldsymbol{\Theta}^t, \boldsymbol{\Upsilon}) \sim \mathcal{N}(\boldsymbol{\mu}_\gamma^t; \boldsymbol{\Sigma}_\gamma^t), \quad (13)$$

where:

$$\boldsymbol{\mu}_\gamma^t = \boldsymbol{\Lambda}^t(\mathbf{w}^t - \mathbf{G}^t\mathbf{s}_0 - \mathbf{h}^t); \ \boldsymbol{\Sigma}_\gamma^t = \mathbf{I}_Q - \boldsymbol{\Lambda}^t\mathbf{G}^t\mathbf{C}\tilde{\mathbf{F}}$$

$$\boldsymbol{\Lambda}^t = \tilde{\mathbf{F}}^\top \mathbf{C}\left(\mathbf{G}^t\right)^\top (\sigma^2\mathbf{I}_{2N} + \mathbf{G}^t\mathbf{C}\tilde{\mathbf{F}}(\mathbf{G}^t\mathbf{C}\tilde{\mathbf{F}})^\top)^{-1}.$$

4.3.2 M-Step

We then replace the latent variables by their expected values and update the model parameters by optimizing the negative log-likelihood function $\mathcal{A}(\boldsymbol{\Theta}, \mathbf{w})$ with

respect to the parameters $\boldsymbol{\Theta}^t$, for $t = \{1, \ldots, T\}$, and $\boldsymbol{\Upsilon}$ where:

$$\mathcal{A}(\boldsymbol{\Theta}, \mathbf{w}) = \mathbb{E}\left[-\sum_{t=1}^T \log p(\mathbf{w}^t | \boldsymbol{\Theta}^t, \boldsymbol{\Upsilon})\right] = NT\log(2\pi\sigma^2)$$

$$+ \frac{1}{2\sigma^2}\sum_{t=1}^T \mathbb{E}\left[\|\mathbf{w}^t - \mathbf{G}^t(\mathbf{s}_0 + \mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\gamma}^t) - \mathbf{h}^t\|_2^2\right]. \quad (14)$$

Note that this log-likelihood function is quadratic in all parameters we seek to estimate, and in contrast to [25], [44], [45], it does not need regularization weights. To update every parameter, we compute the corresponding partial derivative assuming the other parameters are fixed, set it to zero and solve it. We next provide the update rules we obtain.

Updating Elastic Model (C): To perform computations with the matrix $\mathbf{C}$ we need to rewrite it in vectorized form. Since $\mathbf{C}$ is symmetric, we only need to vectorize the upper triangular part of it. For this, we define the function $\text{vech}(\cdot)$, a generalization of the full-matrix vectorization operator $\text{vec}(\cdot)$. The two operators can be related through a duplication matrix $\mathbf{D}_r$, of size $r^2 \times \frac{r(r+1)}{2}$, where r is the size of the original matrix we are vectorizing [33]. The inverse mapping is computed by means of the pseudoinverse, that is, $\text{vech}(\mathbf{C}) = \mathbf{D}_r^\dagger \text{vec}(\mathbf{C})$. For $\mathbf{C}$, we have that $r = 3N$ and we can write:

$$\text{vec}(\mathbf{C}) = \mathbf{D}_r\text{vech}(\mathbf{C}). \quad (15)$$

For the general case when *no boundary conditions* are assumed, $\partial\mathcal{A}/\partial\text{vech}(\mathbf{C})$ can be written as:

$$\frac{1}{2\sigma^2}\sum_{t=1}^T \mathbb{E}\left[\mathbf{D}_r^\top(\tilde{\mathbf{F}}\boldsymbol{\gamma}^t \otimes \mathbf{I}_r)\mathbf{G}^{t\top}(\mathbf{w}^t - \mathbf{G}^t(\mathbf{s}_0 - \mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\gamma}^t) - \mathbf{h}^t)\right].$$

After equating this expression to zero, the update rule for $\text{vech}(\mathbf{C})$ can then be obtained in closed form as[2]:

$$\text{vech}(\mathbf{C}) \leftarrow \left(\sum_{t=1}^T \left((\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t)^\top \otimes (\mathbf{D}_r^\top(\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t \otimes \mathbf{I}_r)(\mathbf{G}^t)^\top \mathbf{G}^t)\right)\mathbf{D}_r\right)^{-1}$$

$$\cdot \sum_{t=1}^T \mathbf{D}_r^\top(\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t \otimes \mathbf{I}_r)(\mathbf{G}^t)^\top(\mathbf{w}^t - \mathbf{G}^t\mathbf{s}_0 - \mathbf{h}^t).$$

When considering the *boundary conditions* introduced by B anchored points, the corresponding rows and columns of these points in the compliance matrix become zero, except the elements in the diagonal which are set to one. If the compliance matrix is rearranged such that stationary points are collocated on the last $3B$ columns, instead of having to retrieve the full matrix $\mathbf{C}$ we will need to estimate a $3(N - B) \times 3(N - B)$ matrix $\mathbf{C}^*$. Equation (15) can then be rewritten as:

$$\text{vec}(\mathbf{C}) = \mathbf{b} + \mathbf{B}\mathbf{D}_q\text{vech}(\mathbf{C}^*), \quad (16)$$

where the duplication matrix $\mathbf{D}_q$ is now of size $q^2 \times q(q + 1)/2$ with $q = 3(N - B)$. $\mathbf{B}$ and $\mathbf{b}$ are a predefined $r^2 \times q^2$ matrix and r^2-dimensional vector used

2. We have applied the $\text{vec}(\cdot)$ operator and the rule $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A})\text{vec}(\mathbf{B})$, for arbitrary matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$.

to enforce, respectively, the out-diagonal zeros and the diagonal ones in $\mathbf{C}$.

In this case, we compute $\partial \mathcal{A}/\partial \mathrm{vech}(\mathbf{C}^*)$, which can be written as:

$$\frac{1}{2\sigma^2} \sum_{t=1}^{T} \mathbb{E}\left[(\mathbf{BD}_q)^\top (\tilde{\mathbf{F}}\gamma^t \otimes \mathbf{I}_r) \mathbf{G}^{t\top} (\mathbf{w}^t - \mathbf{G}^t(\mathbf{s}_0 - \mathbf{C}\tilde{\mathbf{F}}\gamma^t) - \mathbf{h}^t) \right].$$

Again, by setting equating this partial derivative and making use of Eq. (16), it can be shown that $\mathrm{vech}(\mathbf{C}^*)$ equals to:

$$\left(\sum_{t=1}^{T} \left((\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t)^\top \otimes ((\mathbf{BD}_q)^\top (\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t \otimes \mathbf{I}_r)(\mathbf{G}^t)^\top \mathbf{G}^t) \right) \mathbf{BD}_q \right)^{-1}$$

$$\times \left(\sum_{t=1}^{T} (\mathbf{BD}_q)^\top (\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t \otimes \mathbf{I}_r)(\mathbf{G}^t)^\top (\mathbf{w}^t - \mathbf{G}^t \mathbf{s}_0 - \mathbf{h}^t) \right.$$

$$\left. - \sum_{t=1}^{T} \left((\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t)^\top \otimes ((\mathbf{BD}_q)^\top (\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t \otimes \mathbf{I}_r)(\mathbf{G}^t)^\top \mathbf{G}^t) \right) \mathbf{b} \right).$$

Given $\mathrm{vech}(\mathbf{C}^*)$, it is then straightforward to build the symmetric matrix $\mathbf{C}^*$ and the full compliance matrix $\mathbf{C}$ with boundary conditions as:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{3B} \end{bmatrix}. \tag{17}$$

It is worth pointing out that our approach can indistinctly handle both locally stationary or deformable points using the same formulation. This is in contrast to other approaches such as [20], [35], which process them independently and, in particular, only the rigid points are used to estimate the camera motion.

Updating Low-Rank Force Space ($\tilde{\mathbf{F}}$): For computing $\tilde{\mathbf{F}}$ we need to first define the expectation $\phi_{\gamma\gamma}^t = \mathbb{E}[\gamma^t(\gamma^t)^\top] = \boldsymbol{\Sigma}_\gamma^t + \boldsymbol{\mu}_\gamma^t(\boldsymbol{\mu}_\gamma^t)^\top$. By considering again vectorized forms and computing the partial derivative of $\mathcal{A}$ w.r.t. $\mathrm{vec}(\tilde{\mathbf{F}})$, the update rule of the force space can be found to be:

$$\mathrm{vec}(\tilde{\mathbf{F}}) \leftarrow \left(\sum_{t=1}^{T} (\phi_{\gamma\gamma}^t)^\top \otimes (\mathbf{G}^t \mathbf{CJ})^\top \mathbf{G}^t \mathbf{CJ} \right)^{-1}$$

$$\cdot \mathrm{vec}\left(\sum_{t=1}^{T} (\mathbf{G}^t \mathbf{CJ})^\top (\mathbf{w}^t - \mathbf{G}^t \mathbf{s}_0 - \mathbf{h}^t)(\boldsymbol{\mu}_\gamma^t)^\top \right), \tag{18}$$

where for B boundary conditions we have that:

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_{3(N-B)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{3N \times 3N}. \tag{19}$$

The force basis can then be easily obtained by $\tilde{\mathbf{F}} = \mathbf{J} \cdot \mathrm{mat}(\mathrm{vec}(\tilde{\mathbf{F}}))$, where $\mathrm{mat}(\cdot)$ rearranges a vector into a symmetric-square matrix. Note that when no boundary conditions are assumed, $\mathbf{J} \equiv \mathbf{I}_{3N}$.

Observe that when no boundary conditions are considered, the compliance matrix needs to only be symmetric. Yet, as discussed in Sec. 4.1, when boundary conditions are considered, the matrix needs also to be positive-definite. We enforce this condition using the methodology proposed in [29], which, given an input symmetric matrix $\mathbf{C}$, iteratively performs eigen-decomposition

operations to compute a correction matrix $\mathbf{D}_{\mathrm{pd}}$ such that $\mathbf{D}_{\mathrm{pd}}\mathbf{C}$ is the positively-defined matrix closest to $\mathbf{C}$. We then use this matrix to update the force matrix $\tilde{\mathbf{F}}$, keeping the reprojection error in Eq. (14).

Updating the Camera Pose ($\mathbf{R}^t, \mathbf{t}^t$): The camera rotation $\mathbf{R}^t$ is updated by enforcing orthonormality constraints. To this end, we define $\mathbf{R}^t = \boldsymbol{\Pi}\mathbf{Q}^t$, where $\boldsymbol{\Pi}$ is the 2×3 orthographic camera matrix and $\mathbf{Q}^t$ is the full camera rotation. We then follow the iterative strategy proposed in [3], where $\partial\mathcal{A}(\mathbf{Q}^t)/\partial\mathbf{Q}^t = 0$ is optimized while constraining $\mathbf{Q}^t$ to lie in the smooth manifold defined by the orthogonal group $SO(3)$:

$$\underset{\mathbf{Q}^t \in SO(3)}{\arg\min} \sum_{i=1}^{N} \mathbb{E}\left[\|\mathbf{w}_i^t - \boldsymbol{\Pi}\mathbf{Q}^t(\mathbf{s}_{0,i} + (\mathbf{C}\tilde{\mathbf{F}}\gamma^t)_i) - \mathbf{t}^t\|_{\mathcal{F}}^2 \right],$$

where $\mathbf{w}^t = [(\mathbf{w}_1^t)^\top, \ldots, (\mathbf{w}_N^t)^\top]^\top$, $\mathbf{w}_i$ are 2D coordinates, $\mathbf{s}_0 = [\mathbf{s}_{0,1}^\top, \ldots, \mathbf{s}_{0,N}^\top]^\top$, $\mathbf{s}_{0,i}$ are 3D coordinates, and $(\mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t)_i$ is the i-th 3D point of the $3N$ vector $\mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t$. $\|\cdot\|_{\mathcal{F}}$ denotes the Frobenius norm.

Regarding the translation vector $\mathbf{t}^t$ it is straightforward to show that it can be updated as:

$$\mathbf{t}^t \leftarrow \frac{1}{N} \sum_{i=1}^{N} (\mathbf{w}_i^t - \mathbf{R}^t(\mathbf{s}_{0,i} + (\mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t)_i)). \tag{20}$$

Updating Noise Variance (σ^2): By setting $\partial\mathcal{A}(\sigma^2)/\partial\sigma^2 = 0$ we finally update the noise variance in force space as:

$$\sigma^2 \leftarrow \frac{1}{2NT} \sum_{t=1}^{T} \left(\mathrm{tr}\left((\mathbf{G}^t \mathbf{C}\tilde{\mathbf{F}})^\top \mathbf{G}^t \mathbf{C}\tilde{\mathbf{F}}\phi_{\gamma\gamma}^t \right) \right. \tag{21}$$

$$\left. + \|\mathbf{w}^t - \mathbf{G}^t \mathbf{s}_0 - \mathbf{h}^t\|^2 - 2\left(\mathbf{w}^t - \mathbf{G}^t \mathbf{s}_0 - \mathbf{h}^t\right)^\top \mathbf{G}^t \mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\mu}_\gamma^t \right).$$

4.4 Resolving the Elastic Model Scale Ambiguity

When solving for $\mathbf{C}$ and $\tilde{\mathbf{F}}$ we need to enforce $\mathbf{C}$ to be symmetric. Additionally, if boundary conditions are assumed ($B > 0$), $\mathbf{C}$ has to be strictly positive-definite, as discussed in the previous subsection. Therefore, we could consider any symmetric and invertible matrix $\mathbf{A}$ such that $\mathbf{C}\tilde{\mathbf{F}} = \mathbf{CAA}^{-1}\tilde{\mathbf{F}}$, similar to the factor matrix we compute to guarantee the positiveness of the compliance matrix. A new compliance matrix $\mathbf{CA}$ would still be symmetric (and potentially positive-definite) and would yield the same solution for the shape reconstruction in Eq. (10) and reprojection in Eq. (11). That is, the values of $\mathbf{C}$ and $\tilde{\mathbf{F}}$ are retrieved up to a scale factor matrix. A similar ambiguity is produced between $\tilde{\mathbf{F}}$ and γ^t.

Nevertheless, the up to scale compliance matrix $\mathbf{C}$, besides yielding a correct solution to the NRSfM problem, it is also sufficient to model the full physical space. We can use $\mathbf{C}$ to generate, up to scale, any deformation $\mathbf{u}$ applying a given force vector $\mathbf{f}$. And vice-versa, we can obtain an scaled force field to produce a specific displacement. This kind of physical relations, are of course not possible with previous low-rank shape and trajectory approaches. What is not possible with the

compliance matrix we retrieve, though, is to directly estimate the ground truth values of the inherent physical parameters (e.g., Poisson's ratio or Young's modulus) that form the true stiffness matrix. For this to be possible a calibration step should be done to estimate the actual scale factor matrix, in the same line as [30] did for very specific force sensors.

4.5 Dealing with Missing Data

Unlike other methods such as [9], [16], [18], our approach can easily incorporate an strategy to handle incomplete measurements produced by occlusions or outliers. To achieve this, during the M-step of EM algorithm, we just need to estimate the expected log-likelihood of the 2D location $\hat{\mathbf{w}}_i^t$ of the missing points. Since we are using a global model, we can infer their value, despite not being available. In particular we set the 2D position of non-observed points to:

$$\hat{\mathbf{w}}_i^t \leftarrow \mathbf{R}^t(\mathbf{s}_{0,i} + (\mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\mu}_{\boldsymbol{\gamma}}^t)_i) + \mathbf{t}^t. \tag{22}$$

4.6 Initialization

The optimization of Eq. (14) is a highly non-linear problem involving a large number of parameters. For this, it is important not to initialize them completely at random. In particular, we initialize the rigid motion parameters $\{\mathbf{R}^t, \mathbf{t}^t\}$ and $\mathbf{s}_0$ considering the scene does not deform, and we apply rigid factorization [36]. This type of initialization is a standard practice in shape-based NRSfM techniques [1], [8], [11], [16], [20], [22], [25], [40], [39], [52]. The reader may wonder that we could have used other non-rigid formulations like [31] to initialize our approach. However, it turns out that most of the existing methods do not include the shape at rest term in their models, an thus, are not directly applicable. Regarding the compliance matrix $\mathbf{C}$, we do not use any physical prior, and initially set it to the identity matrix (this is equivalent to consider that all points are rigid). The force basis $\tilde{\mathbf{F}}$ matrix is initialized through a coarse-to-fine approach, in which a noise-free version of Eq. (11), where all parameters except $\tilde{\mathbf{F}}$ are given, is first solved for one force-mode, then for two modes, and so on until estimating the Q initial modes. Once all these parameters are set, the starting value of σ^2 is directly computed from Eq. (21). Finally, when dealing with missing data we assume that both the camera motion and 3D shape deformation are smooth over time, and obtain an initial estimation of the missing tracks $\hat{\mathbf{w}}_i^t$ by imposing smooth trajectories, as done in [28].

5 EXPERIMENTAL EVALUATION

We now present our experimental results for different types of sequences including articulated and non-rigid motion (see videos in the supplemental material). We provide both qualitative and quantitative results, where we compare our approach against state-of-the-art

methods, using several Mocap datasets with 3D ground truth. For these datasets we report the 3D reconstruction error, computed as $e_{3\mathcal{D}} = \frac{1}{T}\sum_{t=1}^{T}\frac{\|\mathbf{s}^t-\mathbf{s}_{GT}^t\|_{\mathcal{F}}}{\|\mathbf{s}_{GT}^t\|_{\mathcal{F}}}$, where $\mathbf{s}^t$ is the estimated 3D reconstruction and $\mathbf{s}_{GT}^t$ is the corresponding 3D ground truth. $e_{3\mathcal{D}}$ is computed after aligning the estimated 3D shape with the 3D ground truth using Procrustes analysis over all T frames.

5.1 Motion Capture Data

The standard way to compare NRSfM approaches is through a number of datasets with ground truth, acquired using Mocap systems. We consider the following ones: the face deformation sequences *Jacky* and *Face*, from [52] and [40], respectively; *Walking* for articulated motion from [52], and a sparse version of *Flag* waving in the wind [56].

We compare our approach, denoted EM-PFS (for Expectation-Maximization on Probabilistic Force Space) against nine other methods, which use low-rank models on both shape and trajectory spaces, or inextensibility constraints. Among the shape space methods we consider: EM-PPCA [52], EM-LDS [52], the Metric Projections (MP) [40], the block matrix approach for SPM [18] and EM-PND [31]. Regarding the trajectory-based ones, we evaluate the DCT-based 3D point trajectory (PTA) [9]. As shape-trajectory methods we consider Column Space Fitting (CSF2) [28] and the Kernel Shape Trajectory Approach (KSTA) [27]. We also consider the Inextensibility Fusion Movies (IFM) technique [55], which exploits isometry constraints. The parameters of these methods were set in accordance with their original papers.

In our EM-PFS model we will consider two modalities: First, we will enforce $\mathbf{C} = \mathbf{I}$. By doing this, our deformation model reduces to the same considered in EM-PPCA [52]. Yet, there exist still differences between the two approaches regarding the methodology used to estimate the camera motion. While the constraint $\mathbf{C} = \mathbf{I}$ can also provide reasonable results in many cases, the elastic information we recover for this case is very limited, mainly in terms of correlations between nodal points. For continuous materials, or objects with sparse connections, there always exist certain degree of connection between nodal points. And second, we will consider our full model in which the compliance matrix $\mathbf{C}$ is estimated. In both cases, the only parameter that needs to be manually set is the number Q of modes of the low-rank force space. There is no other parameter nor regularization weight that needs to be tuned. In Fig 4 we have evaluated the sensitivity of the reconstruction results (when $\mathbf{C}$ is fully estimated) to the choice of the parameter Q in the four Mocap sequences. Note that in two of the sequences there is almost no influence while in the other two, there are specific dimensions which are noisier. Interestingly, note that increasing the rank of the subspace does not guarantee reducing the error. This is because the basis shapes are simultaneously learned with the shape, and hence, larger rank values

Space: Met. Seq.	Shape					Trajectory	Shape-Trajectory		Isometric	Force (EM-PFS)	
	EM-PPCA [52]	EM-LDS [52]	MP [40]	SPM [18]	EM-PND [31]	PTA [9]	CSF2 [28]	KSTA [27]	IFM [55]	$\mathbf{C} = \mathbf{I}$	$\mathbf{C}$
Noise-less Observations											
Jacky [52]	1.80(5)	2.79(2)	2.74(5)	1.82(7)	1.41	2.69(3)	1.93(5)	2.12(4)	4.04	1.86(7)	1.80(7)
Face [40]	7.30(9)	6.67(2)	3.77(7)	2.67(9)	25.79	5.79(2)	6.34(5)	6.14(8)	5.96	3.86(5)	2.85(5)
Flag	4.22(12)	6.34(3)	10.72(3)	7.84(5)	4.11	8.12(6)	7.96(2)	7.74(2)	2.33	5.43(12)	5.29(12)
Walking [52]	11.11(10)	27.29(2)	17.51(3)	8.02(6)	3.90	23.60(2)	6.39(5)	6.36(5)	28.83	9.41(11)	8.54(11)
Average error:	6.11	10.77	8.69	5.09	8.80	10.05	5.66	5.59	10.29	5.14	**4.62**
Noisy Observations											
Jacky [52]	2.21(5)	2.80(2)	7.70(5)	3.42(7)	2.14	3.04(3)	2.78(5)	2.48(4)	8.56	2.88(7)	2.79(7)
Face [40]	7.79(9)	5.85(2)	4.82(7)	5.17(9)	29.27	5.84(2)	6.66(5)	10.88(8)	6.86	4.31(5)	3.30(5)
Flag	5.03(12)	6.37(3)	11.30(3)	9.03(5)	4.73	8.84(6)	8.56(2)	9.65(2)	5.83	5.89(12)	5.74(12)
Walking [52]	10.93(10)	32.40(2)	18.89(3)	11.39(6)	5.00	22.18(2)	6.87(5)	6.81(5)	30.38	13.22(11)	10.50(11)
Average error:	6.49	11.86	10.67	7.25	10.28	9.97	6.22	7.45	12.91	6.57	**5.58**
30% Random Missing Observations											
Jacky [52]	2.07(5)	2.76(2)	4.00(5)	–	1.51	–	2.39(5)	2.41(4)	¬	2.75(7)	2.71(7)
Face [40]	8.58(9)	49.77(2)	8.20(7)	–	26.12	–	7.55(5)	10.36(8)	¬	3.98(5)	3.96(5)
Flag	4.74(12)	6.72(3)	11.48(3)	–	4.27	–	9.16(2)	9.01	¬	6.05(12)	5.92(12)
Walking [52]	30.58(10)	17.57(2)	22.79(3)	–	3.93	–	7.75(5)	7.03(5)	¬	15.52(11)	9.05(11)
Average error:	11.49	19.21	11.62		8.96		6.71	7.20		7.07	**5.41**

TABLE 3

Reconstruction error of all methods for noise-free, noisy and missing observations. We report $e_{3\mathcal{D}}[\%]$ for shape basis methods EM-PPCA [52], EM-LDS [52], MP [40], SPM [18] and EM-PND [31]; for the trajectory basis method PTA [9]; for shape-trajectory basis methods CSF2 [28] and KSTA [27]; for the inextensibility-based method IFM [55]; and for our force basis approach denoted EM-PFS. For our method, we propose two modalities: 1) when the elastic model is fixed $\mathbf{C} = \mathbf{I}$ and 2) when it is fully estimated. We have chosen the basis rank (in parenthesis) that gave the lowest $e_{3\mathcal{D}}$ error for the noise-free case. The symbol "$-$" indicates the algorithm cannot handle missing entries, and "¬", that the code to handle missing tracks is not available.

introduce additional unknown parameters to estimate. We experimentally observe that $Q = 5$ yields a good trade-off between accuracy and computational cost. For the real applications described later we use this value.

We then compare all methods in three situations: 1) noise-free observations, 2) when the 2D point tracks were artificially corrupted by zero-mean Gaussian noise with standard deviation $\sigma_{noise} = 0.01\rho$, with ρ being the maximum distance of an image point to the centroid of all the points, and 3) randomly removing 30% of the observed tracks. The mean 3D reconstruction errors are summarized in Table 3. Observe that our approach consistently performs either the best or among the best in all sequences for all cases, and on average is the one with smaller error. In particular note that we slightly outperform SPM [18] and KSTA [27], which are acknowledged to be at the top of the state-of-the-art in low-rank based models. And most importantly, we do not only solve for the NRSfM problem, but we additionally provide an estimation of the full elastic model of the object. Note also that while IFM [55] provides accurate reconstructions for isometric deformations (e.g., flag sequence), these type constraints do not seem adequate under noisy observations.

It is also worth noting that our approach seems to perform better for continuous surfaces rather than for articulated shapes ("walking" sequence). This was indeed expected as the underlying theory ruling the generation of the compliance matrix $\mathbf{C}$ is based on continuum mechanics. In any event, the errors obtained in the "walking" sequence are still within very reasonable bounds, indicating that the estimated compliance matrix is a good approximation of the body joint correlations.
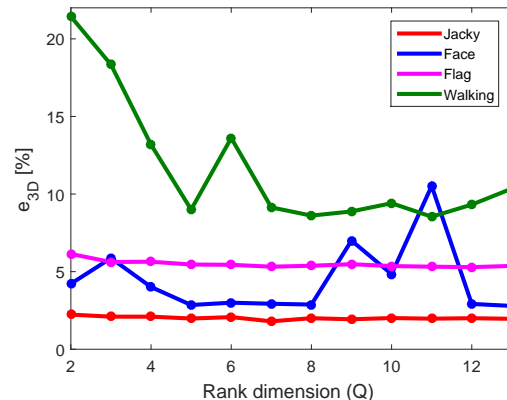

Fig. 4. **3D Reconstruction error as a function of the rank $\mathbf{Q}$ of the force subspace**. Results on the four Mocap sequences. For the *Jacky* and *Flag* sequences the rank $\mathbf{Q}$ has no influence on the reconstruction errors. The other two sequences (*Face* and *Walking* are more sensitive, but the error always remains within reasonable bounds.

5.2 Real Videos

We have also evaluated our approach on several real sequences, which despite not having ground truth, allow a qualitative evaluation in different real-world scenarios and under the presence of structured occlusions, where other approaches like [9], [18] are prone to fail. Since the results on Mocap sequences suggest that our method does not strongly depend on the number of force vectors Q, we chose a rank of 5 for all real experiments.

First, we processed the beating heart sequence, of 79 frames and acquired during bypass surgery. We use the outlier-free point tracks of [26], computed using optical flow. Figure 5 shows the 3D reconstruction we obtain, where one of the main challenges is that the movement of the camera is very small. This especially penalizes
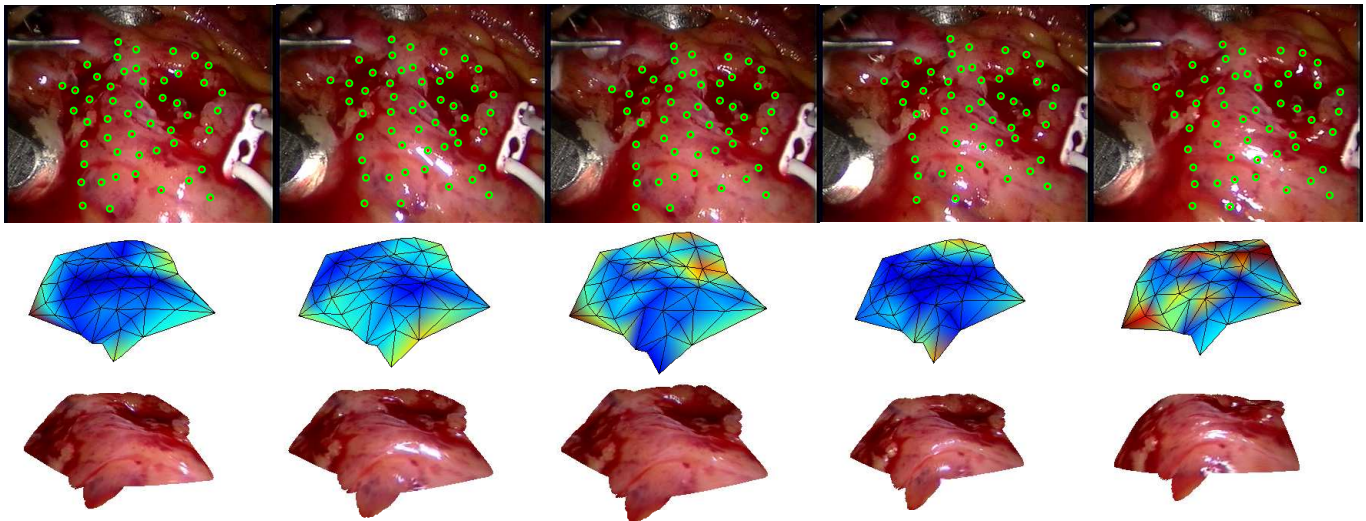
Fig. 5. **Beating heart sequence. Top:** 2D tracking data and reconstructed 3D shape reprojected onto several images with green circles and red dots, respectively. **Middle:** Reconstructed 3D shape, color code such that reddish areas indicate larger displacements. **Bottom:** Reconstructed 3D shape, using the original texture. Best viewed in color.
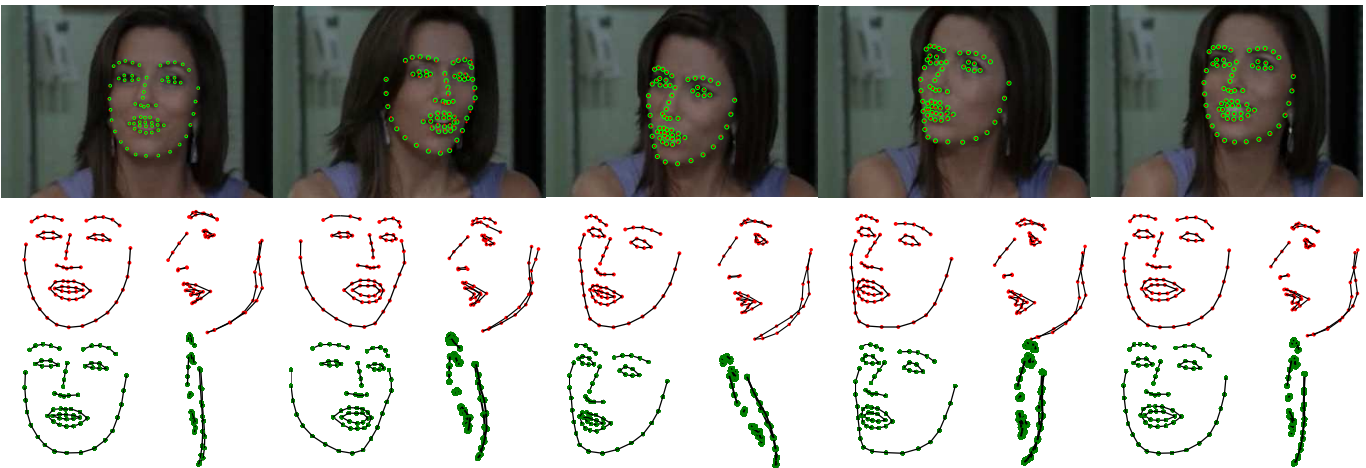

Fig. 6. **Actress sequence. Top:** 2D tracking data (green circles) and reprojection (red dots) of the reconstructed 3D shape. **Middle:** Camera and side-views of the reconstructed shapes obtained by our approach. **Bottom:** Same views using EM-PND [31].

trajectory-based methods. The color-coded reconstructions, representing the amount of deformation, show that we can recover the rhythmic deformations of the heart, while learning its elastic model.

We also processed the actress sequence, with 102 frames where a woman is talking and moving her head. The point tracks were provided by [11]. Figure 6 shows the 3D reconstruction, appropriately rotated according to the estimated pose. We also show the results of the EM-PND [31], known to be very accurate except for situations like this sequence, in which the camera rotation is small.

Fig. 7 shows the 3D reconstruction of the back of a person. Point tracks are obtained from [44]. Again, one of the difficulties of this sequence is to deal with small camera motions, which our approach handles without much difficulty.

Finally, we have also processed two ASL sequences of an American Sign Language (ASL), consisting of a person moving the head while talking and hand gesturing.

The goal is to reconstruct the face which, in some frames is partially occluded by one or two hands, or by the face self-rotation. The ASL1 sequence consists of 115 frames and 77 feature points, with a 17.4% of missing data. The ASL2 sequence consists of 114 frames and also 77 feature points, with a 11.5% of missing data [28]. For these sequences, we have evaluated the case when considering boundary conditions. For this purpose, we chose $B = 14$ points on the contour of the face to be rigid. These points are displayed as squares in Fig. 8. The reconstruction results are shown on the second and fourth rows of the figure. Note that even when occlusions appear, our model provides a correct estimation for the occluded shape. While this reconstruction is very similar to that obtained by CSF2 [28], SPM [18], a method that showed great performance in the Mocap data experiments of Table 3, is not able to handle missing data. For this specific example, the reconstruction obtained when no boundary conditions are used were virtually the same (we do not plot these results). However, the use of the boundary
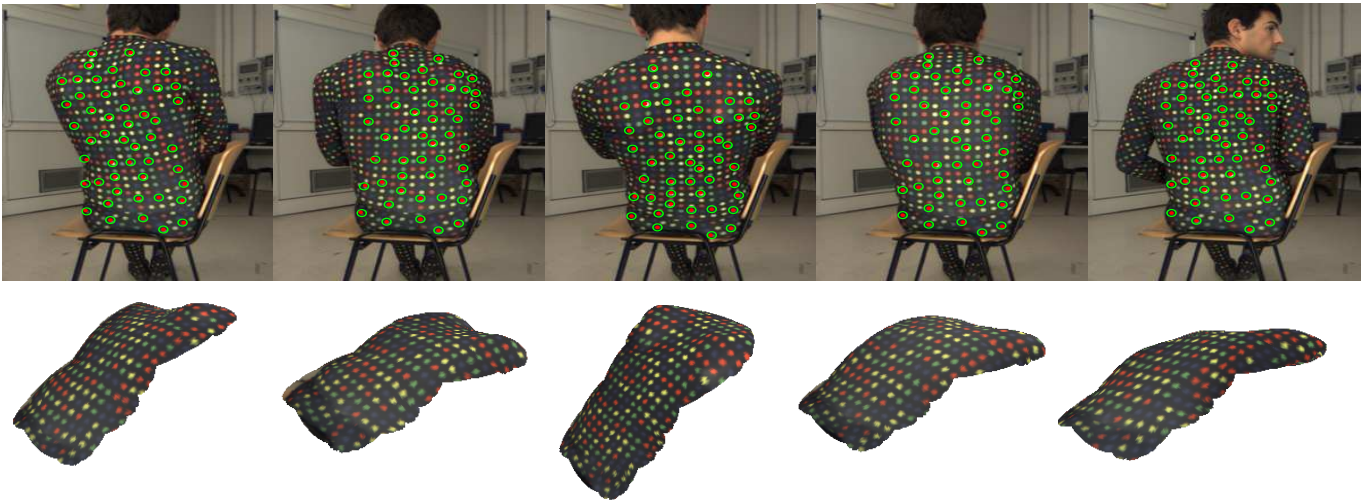
Fig. 7. **Back sequence. Top:** 2D tracking data and reconstructed 3D shape reprojected into several images with green circles and red dots, respectively. **Bottom:** Side view of the reconstructed shape.
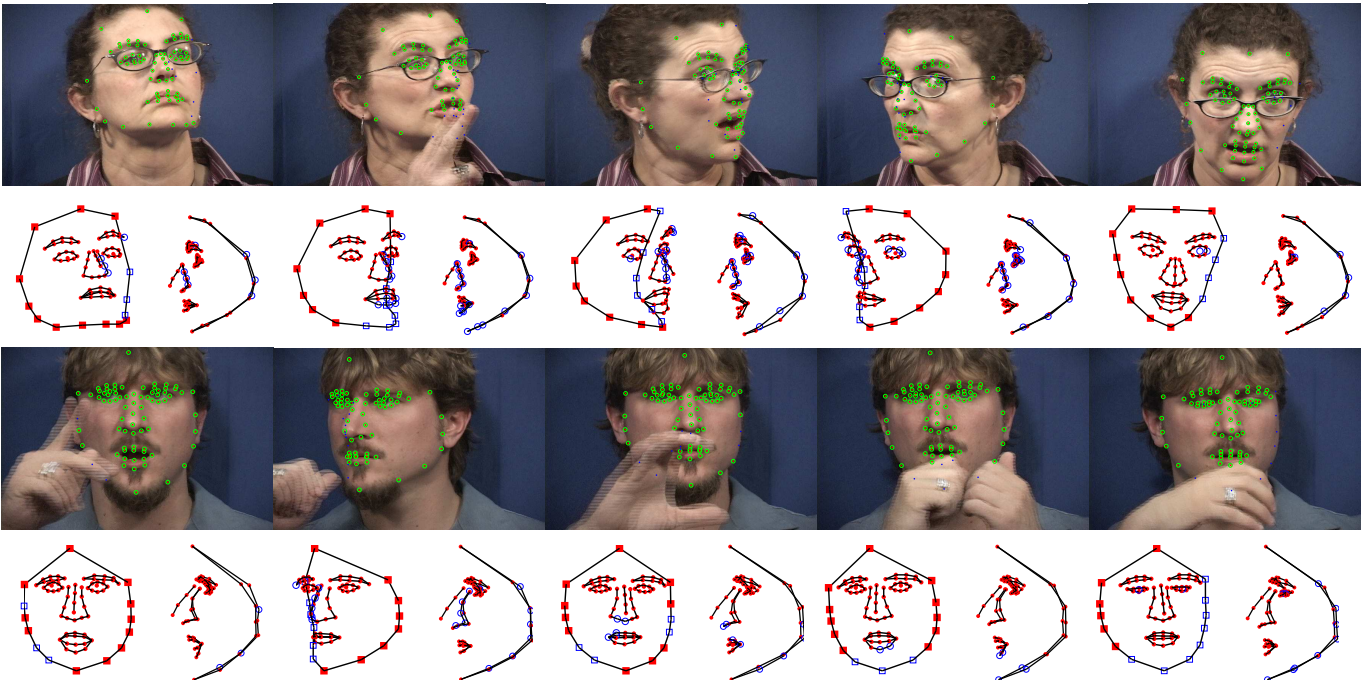

Fig. 8. **ASL1 and ASL2 sequences.** The same information is shown for the two experiments. **Top:** 2D tracking data (green circles) and reconstructed 3D shape (red dots) reprojected onto several images. Blue circles correspond to reconstructed missing points. **Bottom:** Camera frame and side-views of the reconstructed 3D shape when considering boundary conditions with $B = 14$ rigid points. These points are represented by squares. The 3D reconstruction without assuming these priors is very similar, but computationally more expensive. Best viewed in color.

conditions highly reduces the number of free parameters to estimate, and consequently the computation time. For the ASL1 sequence, the computation time was reduced from 588 sec. to 374 sec. and from 583 sec to 369 for the ASL2 sequence (results obtained in a commodity laptop Intel core i7@2.67 GHz).

5.3 Elastic Model Estimation

An interesting contribution of our approach is that besides estimating the shape and camera trajectory, we provide an estimation of the elastic model $\mathbf{C}$ of the object, and a low-rank force subspace $\tilde{\mathbf{F}}$ (with the corresponding force coefficients $\boldsymbol{\Gamma}$). Additionally, as discussed

in Section 3.3, once we have estimated these parameters, we can directly compute the equivalence between the force, shape and trajectory spaces (the shape-trajectory model is a combination of these results). Concretely, the low-rank shape space has been shown to be $\tilde{\mathbf{S}} = \mathbf{C}\tilde{\mathbf{F}}$, and the low-rank trajectory space $\tilde{\mathbf{T}} = \boldsymbol{\Gamma}$.

In Fig. 9 we plot these equivalences for the example of the actress sequence introduced previously. We separately consider the cases with and without boundary conditions. On top we plot the first five force modes, as vectors overlaying the shape at rest, when there are no boundary conditions. Observe that the larger magnitudes of the modes concentrate around the mouth, which
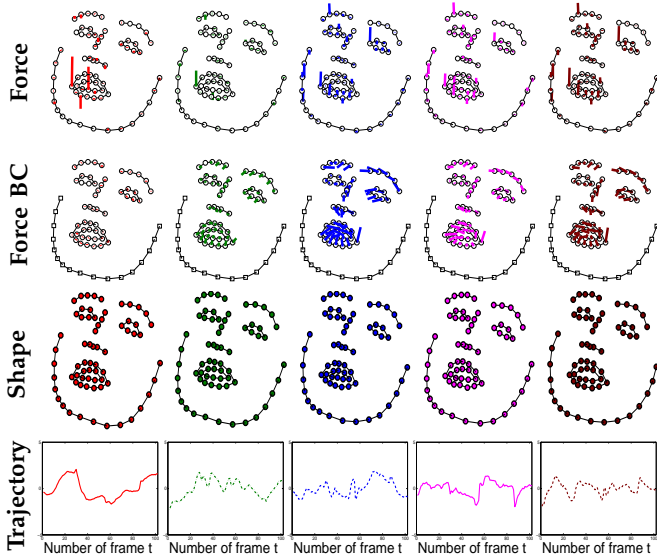
Fig. 9. Comparison of low-rank spaces for the Actress sequence. Equivalence between the force, shape and trajectory spaces, for rank $Q = 5$ without boundary conditions and with $B = 17$ fixed points (represented by squares). **First and second row:** Modes in the force space without/with boundary conditions, respectively. **Third row:** Modes in the shape space. **Fourth row:** Modes in the trajectory space.

is the part of the face undergoing larger deformations. In the second row of Fig. 9 we plot the same force vectors when $B = 17$ points on the contour of the face are forced to be rigid. Note that the distribution of forces within the face has changed w.r.t. to the previous case. In particular, the magnitude of the forces for the non-rigid points is much larger, i.e., when anchoring points on the boundary, it is then necessary to apply more force on the rest of points to achieve similar deformation patterns.

The two bottom rows of Fig. 9 depict the corresponding shape and trajectory modes for the case without boundary conditions (when introducing boundary conditions these modes hardly change, and we do not plot them). Regarding the shape basis we retrieve, although it is difficult to appreciate from non-overlapping images, note the subtle differences between the configuration of each mode, and again, particularly around the area of the mouth. The bottom-most plot, plots the five trajectory modes, with size equal to the sequence length. The theoretical modes used in the trajectory-based methods correspond to the sinusoidal functions of a DCT. Observe that the first mode estimated by the proposed approach, quite resembles such a function.

In Fig. 10 we demonstrate that the compliance matrix we estimate allows recovering the full physical space. Although in our formulation we do not explicitly enforce the rank of $\mathbf{C}$, it is always full rank. For instance the four face configurations we plot on the left are produced by applying specific forces $\mathbf{f}$ and computing the resulting deformations $\mathbf{u}$ via the relation $\mathbf{u} = \mathbf{C}\mathbf{f}$. Each face corresponds to the product of the compliance matrix $\mathbf{C}$, by one of the force vectors $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4$ depicted on the right

side of the matrix, plus the shape at rest. Observe that with this force model we can generate shape configurations (e.g., winking one or two eyes, mouth wide open) that would be hard or impossible to obtain using low-rank shape, trajectory and shape-trajectory spaces unless similar shapes are explicitly observed (in shape-based methods) or they use a very large number of modes (in trajectory-based methods). In contrast, using the physical space we propose, we can produce these shapes even when they have not been observed and directly from the elastic model we have learned. Additionally, note how the forces $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4$ necessary to produce these shape configurations are smooth (their color coded components do not abruptly change).

This would not happen if we had used a random symmetric compliance matrix $\mathbf{C}^{\mathrm{r}}$, as shown on the center of Fig. 10. This matrix would also allow minimizing Eq. (14), but the resulting forces $\mathbf{f}_1^{\mathrm{r}}, \mathbf{f}_2^{\mathrm{r}}, \mathbf{f}_3^{\mathrm{r}}, \mathbf{f}_4^{\mathrm{r}}$ would not be quite realistic. The representation of these forces (on the right side of the random compliance matrix) depicts sharp changes, indicating that a such a compliance matrix would not appropriately model the underlying physics of the object.

Finally, on the right-most of Fig. 10 we plot the corresponding compliance matrix $\mathbf{C}^{\mathrm{bc}}$ and resulting forces $\mathbf{f}_1^{\mathrm{bc}}, \mathbf{f}_2^{\mathrm{bc}}, \mathbf{f}_3^{\mathrm{bc}}, \mathbf{f}_4^{\mathrm{bc}}$ when boundary conditions are considered. Note that the smoothness pattern is similar as that in the boundary-less case, but in this case, the compliance matrix has only diagonal elements and null forces for the entries corresponding to the anchored points.

5.4 Transferring Elastic Models

Once the elastic model $\mathbf{C}$ is learned for one specific object, it can be used to encode the deformation of another object of the same family, represented by the same number of points. Using the force-space formulation we propose, the compliance matrix for this new object can be assumed to be known, reducing thus the problem to only having to estimate the force and pose parameters. We have evaluated this alternative on the ASL sequences, and in particular, we use the compliance matrix $\mathbf{C}$ estimated for ASL1, to solve the NRSfM problem for the ASL2 sequence. The 3D reconstruction results we obtain whether transferring or not the elastic model are practically identical. The only difference is on the estimated force bases, which in Fig. 11 we plot onto the mean shape. It can be observed that the forces for the transferred case have large magnitudes. This is because the deformations within ASL1 are smaller than in ASL2 (e.g., for ASL1 the eyes are never closed as in ASL2), and the forces need to compensate this larger deformation effect of ASL2. Finally, mention that an important advantage of using the learned elastic model is on the computation time: not having to estimate the parameters of $\mathbf{C}$ gives an speed up of $288\times$.
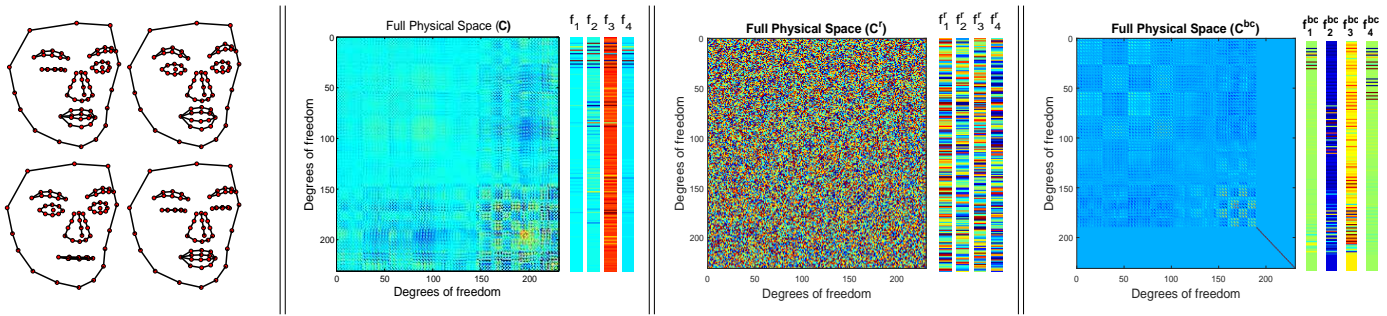
Fig. 10. **Estimating forces that produce a specific deformation for the ASL1 sequence.** Once the compliance matrix is learned, we can estimate the forces that define any shape in the full physical space. **First column:** Four target shapes. **Second column:** Compliance matrix $\mathbf{C}$ estimated without assuming boundary conditions. $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4$, are the forces necessary to deform the mean shape into the target shapes. **Third column:** Random symmetric matrix $\mathbf{C}^{\mathrm{r}}$, and the corresponding forces $\mathbf{f}_1^{\mathrm{r}}, \mathbf{f}_2^{\mathrm{r}}, \mathbf{f}_3^{\mathrm{r}}, \mathbf{f}_4^{\mathrm{r}}$ to produce the target shapes. **Fourth column:** Compliance matrix $\mathbf{C}^{\mathrm{bc}}$ estimated after enforcing $B = 14$ boundary-condition constraints, and the corresponding forces to produce the target shapes. The figure is best viewed in color.

6 CONCLUSION

In this paper we have formulated the NRSfM problem using a new low-rank force model. From only 2D point tracks in a monocular video, besides recovering shape and camera motion, we also estimate an elastic model of the object. This allows for rich physical interpretations of the dynamics in terms of force and displacement. Additionally, we have shown the connections of our force-model to the shape, trajectory and shape-trajectory based spaces used so far. The results demonstrate that the proposed technique is applicable to a wide variety of real-world deformations and materials, without requiring any prior knowledge about the physical or geometric object properties. We obtain state-of-the-art performance in reconstruction accuracy, while also providing an estimation of the object elastic model.

Further experiments show that we can realistically transfer the learned elastic model between objects of the same family, highly relieving the computational cost for shape estimation. We have also shown that once the elastic model of the object is learned, we can infer the forces that produce specific deformations. This is especially interesting in robotic manipulation tasks. However, in order to make this applicable to real systems, we need to resolve the scale ambiguity that still remains on the estimated forces. In the future we plan doing this by introducing certain constraints into the optimization, as well as the optimal connectivity between points. Additionally, in order to alleviate the computational cost of our approach we plan to research coarse-to-fine strategies to transfer elastic models from sparse to dense object configurations.
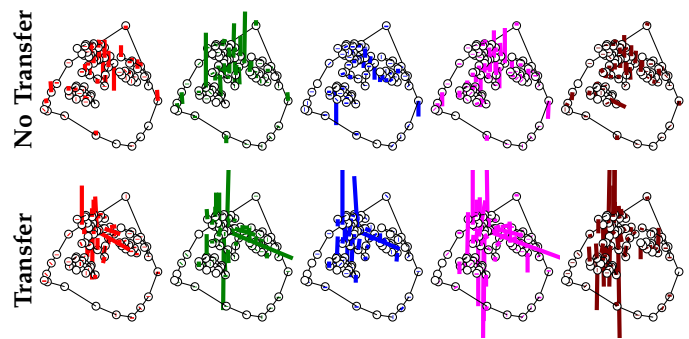
ACKNOWLEDGMENTS

Fig. 11. **Transferring models between ASL sequences.** Force basis learned for ASL2, either considering unknown compliance matrix (top) or using the compliance matrix estimated from ASL1 (bottom).

REFERENCES

[1] A. Agudo, L. Agapito, B. Calvo, and J. M. M. Montiel. Good vibrations: A modal analysis approach for sequential non-rigid structure from motion. In *CVPR*, 2014.

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

[3] A. Agudo, J. M. M. Montiel, L. Agapito, and B. Calvo. Online dense non-rigid 3D shape and camera motion recovery. In *BMVC*, 2014.

[4] A. Agudo, J. M. M. Montiel, L. Agapito, and B. Calvo. Modal space: A physics-based model for sequential estimation of time-varying shape from monocular video. *JMIV*, 57(1):75–98, 2017.

[5] A. Agudo, J. M. M. Montiel, B. Calvo, and F. Moreno-Noguer. Mode-shape interpretation: Re-thinking modal space for recovering deformable shapes. In *WACV*, 2016.

[6] A. Agudo and F. Moreno-Noguer. Learning shape, motion and elastic models in force space. In *ICCV*, 2015.

[7] A. Agudo and F. Moreno-Noguer. Simultaneous pose and non-rigid shape with particle dynamics. In *CVPR*, 2015.

[8] A. Agudo, F. Moreno-Noguer, B. Calvo, and J. M. M. Montiel. Sequential non-rigid structure from motion using physical priors. *TPAMI*, 38(5):979–994, 2016.

[9] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Non-rigid structure from motion in trajectory space. In *NIPS*, 2008.

[10] J. Barbic and D. James. Real-time subspace integration for st. venant-kirchhoff deformable models. *TOG*, 24(3):982–990, 2005.

[11] 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.

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

[13] M. Becker and M. Teschner. Robust and efficient estimation of elasticity parameters using the linear finite element method. In *SV*, 2007.

[14] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *ACM SIGGRAPH*, 1999.

[15] K. L. Bouman, B. Xiao, P. Battaglia, and W. T. Freeman. Estimating the material properties of fabric from video. In *ICCV*, 2013.

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

[17] M. Brubaker, L. Sigal, and D. Fleet. Estimating contact dynamics. In *ICCV*, 2009.

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

[19] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman. Visual vibrometry: Estimating material properties from small motions in video. In *CVPR*, 2015.

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

[21] H. Eskandari, S. Salcudean, R. Rohling, and I. Bell. Real-time solution of the finite element inverse problem of viscoelasticity. *IP*, 27(8):1–16, 2011.

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

[23] J. Fayad, A. Del Bue, L. Agapito, and P. M. Q. Aguiar. Non-rigid structure from motion using quadratic deformation models. In *BMVC*, 2009.

[24] K. Fragkiadaki, M. Salas, P. Arbeláez, and J. Malik. Grouping-based low-rank trajectory completion and 3D reconstruction. In *NIPS*, 2014.

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

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

[27] P. F. U. Gotardo and A. M. Martinez. Kernel non-rigid structure from motion. In *ICCV*, 2011.

[28] P. F. U. Gotardo and A. M. Martinez. Non-rigid structure from motion with complementary rank-3 spaces. In *CVPR*, 2011.

[29] N. J. Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103:103–118, 1988.

[30] M. Hwangbo and T. Kanade. Factorization-based calibration method for MEMS inertial measurement unit. In *ICRA*, 2008.

[31] M. Lee, J. Cho, C. H. Choi, and S. Oh. Procrustean normal distribution for non-rigid structure from motion. In *CVPR*, 2013.

[32] M. Lee, C. H. Choi, and S. Oh. A procrustean markov process for non-rigid structure recovery. In *CVPR*, 2014.

[33] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons: Chichester/New York, 1988.

[34] A. Malti, A. Bartoli, and R. Hartley. A linear least-squares solution to elastic shape-from-template. In *CVPR*, 2015.

[35] A. Malti, R. Hartley, A. Bartoli, and J. H. Kim. Monocular template-based 3D reconstruction of extensible surfaces with local linear elasticity. In *CVPR*, 2013.

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

[37] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *TPAMI*, 15(6):580–591, 1993.

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

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

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

[41] H. S. Park, T. Shiratori, I. Matthews, and Y. Sheikh. 3D reconstruction of a moving point from a series of 2D projections. In *ECCV*, 2010.

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

[43] S. Roweis. EM algorithms for PCA and SPCA. In *NIPS*, 1998.

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

[45] M. Salzmann and R. Urtasun. Physically-based motion models for 3D tracking: A convex formulation. In *ICCV*, 2011.

[46] S. Sclaroff and A. P. Pentland. Physically-based combinations of views: Representing rigid and nonrigid motion. In *WMNRAO*, 1994.

[47] T. Simon, J. Valmadre, I. Matthews, and Y. Sheikh. Separable spatiotemporal priors for convex reconstruction of time-varying 3D point clouds. In *ECCV*, 2014.

[48] C. Syllebranque and S. Boivin. Estimation of mechanical parameters of deformable solids from videos. *The Visual Computer*, 4(11):963–972, 2008.

[49] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analysers. *NC*, 11(2):443–482, 1999.

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

[51] L. Torresani, A. Hertzmann, and C. Bregler. Learning non-rigid 3D shape from 2D motion. In *NIPS*, 2004.

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

[53] L. V. Tsap, D. B. Goldof, and S. Sarkar. Nonrigid motion analysis based on dynamic refinement of finite element models. *TPAMI*, 22(5):526–543, 2000.

[54] J. Valmadre and S. Lucey. General trajectory prior for non-rigid reconstruction. In *CVPR*, 2012.

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

[56] R. White, K. Crane, and D. Forsyth. Capturing and animating occluded cloth. In *ACM SIGGRAPH*, 2007.

[57] M. A. Woodbury. Inverting modified matrices. *Statistical Research Group, Memorandum Rept. 42*, 1950.

[58] S. Wuhrer, J. Lang, and C. Shu. Tracking complete deformable objects with finite elements. In *3DV*, 2012.

[59] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2D+3D active appearance models. In *CVPR*, 2004.

[60] Y. Zhu, T. J. Hall, and J. Jiang. A finite-element approach for young modulus reconstruction. *TMI*, 22(7):890–901, 2003.

[61] Y. Zhu, D. Huang, F. De La Torre, and S. Lucey. Complex non-rigid motion 3D reconstruction by union of subspaces. In *CVPR*, 2014.

Antonio Agudo received the M.Sc. degree in industrial engineering and electronics in 2010, M.Sc. degree in computer science in 2011, and the Ph.D. degree in computer vision and robotics in 2015, from University of Zaragoza. He was a visiting student at vision group of Queen Mary University of London in 2013 and with the vision and imaging science group of University College London in 2014. He was also a visiting fellow at Harvard University in 2015. He is a postdoctoral fellow at the computer vision department of the Institut de Robòtica i Informàtica Industrial (CSIC-UPC) in Barcelona. His research interests include non-rigid structure from motion, machine learning, and deformation analysis to medical and robotics applications.

Francesc Moreno-Noguer received the MSc degrees in industrial engineering and electronics from the Technical University of Catalonia (UPC) and the Universitat de Barcelona in 2001 and 2002, respectively, and the PhD degree from UPC in 2005. From 2006 to 2008, he was a postdoctoral fellow at the computer vision departments of Columbia University and the École Polytechnique Fédérale de Lausanne. In 2009, he joined the Institut de Robòtica i Informàtica Industrial in Barcelona as an associate researcher of the Spanish Scientific Research Council. His research interests include retrieving rigid and nonrigid shape, motion, and camera pose from single images and video sequences. He received UPC's Doctoral Dissertation Extraordinary Award for his work.