Learning Shape, Motion and Elastic Models in Force Space

Antonio Agudo[1,2] and Francesc Moreno-Noguer[1]
[1]Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Barcelona, Spain
[2]Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, Spain

Problem Statement

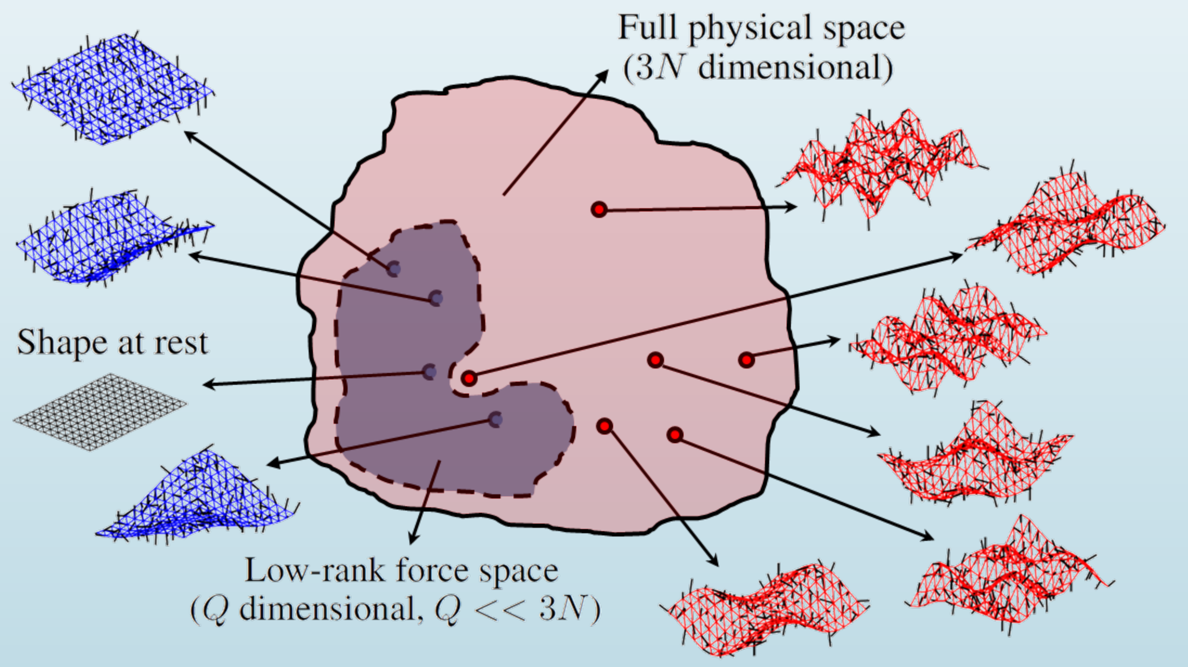
- Given: 2D temporal point tracks acquired with a monocular camera.
- We want: A NRS*f*M solution that estimates time-varying shape, camera motion and the **full elastic model** of the observed object.
- Existing approaches only recover small parts of the full physical model.

Approach

- Real-world objects are deformed by a set of acting forces that lie in a force linear subspace.
- Model parameters are *learned* using Expectation Maximization (EM) with partial M-steps.
- Approach suitable to model a wide variety of objects and deformations, even under *partial observations*.

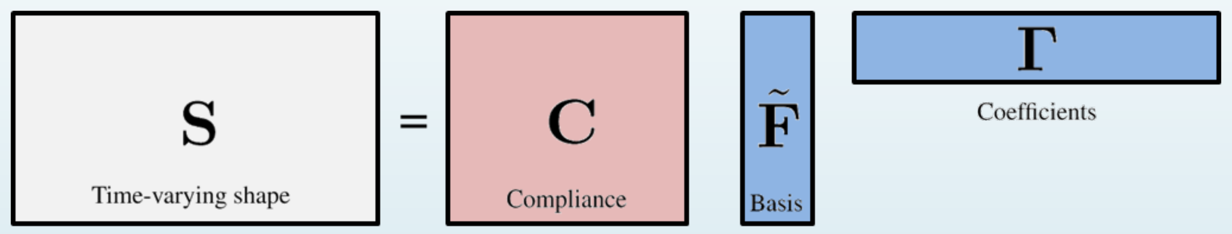
Contribution

- Interpretation of the NRS*f*M problem in terms of a **low-rank force** space.
- Learning the full elastic model *without requiring* any prior knowledge.
- A shape-trajectory-force duality that gives *physical interpretation* to low-rank methods.



Low-Rank Force Space

- A 3D displacement is defined as $\mathbf{S} = \mathbf{CF}$, with $\mathbf{F}$ the acting forces and $\mathbf{C}$ the compliance matrix (elasticity). $\mathbf{F} \approx \tilde{\mathbf{F}}\mathbf{\Gamma}$ is low-rank.



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

Force-Shape-Trajectory Duality

- The time-varying shape $\mathbf{S}$ can also be modeled using low-rank shape or trajectory spaces:
  - Shape: $\mathbf{S} = \tilde{\mathbf{S}}\mathbf{\Psi}$, with $\tilde{\mathbf{S}}$ the basis.
  - Trajectory: $\mathbf{S} = \mathbf{\Phi}\tilde{\mathbf{T}}$, with $\tilde{\mathbf{T}}$ the basis.
- From direct comparison with the force model: for shape $\tilde{\mathbf{S}} = \mathbf{C}\tilde{\mathbf{F}}$, and for trajectory $\tilde{\mathbf{T}} = \mathbf{\Gamma}$.
- Shape and trajectory bases are seen as **physical priors** (and not just *statistical*).

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

Learning Elastic Model, Shape and Camera Motion

- Projection model under orthography:

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

- Our problem is to recover the elastic model $\mathbf{C}$, camera motion $(\mathbf{G}^t, \mathbf{h}^t)$ and dynamic shape $\mathbf{s}^t = \mathbf{s}_0 + \mathbf{C}\tilde{\mathbf{F}}\boldsymbol{\gamma}^t$ from observations $\mathbf{w}^t$.
- Learning $\mathbf{C}$ is challenging due to its large number of parameters. We just estimate the upper triangular part using vectorization rules.
- Weight coefficients are modeled as Gaussian distributions $\boldsymbol{\gamma}^t \sim \mathcal{N}(\mathbf{0}; \mathbf{I}_Q)$.
- This problem is equivalent to recover the distribution over observations:

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

- We use an EM algorithm with partial M-steps to retrieve every model parameter.

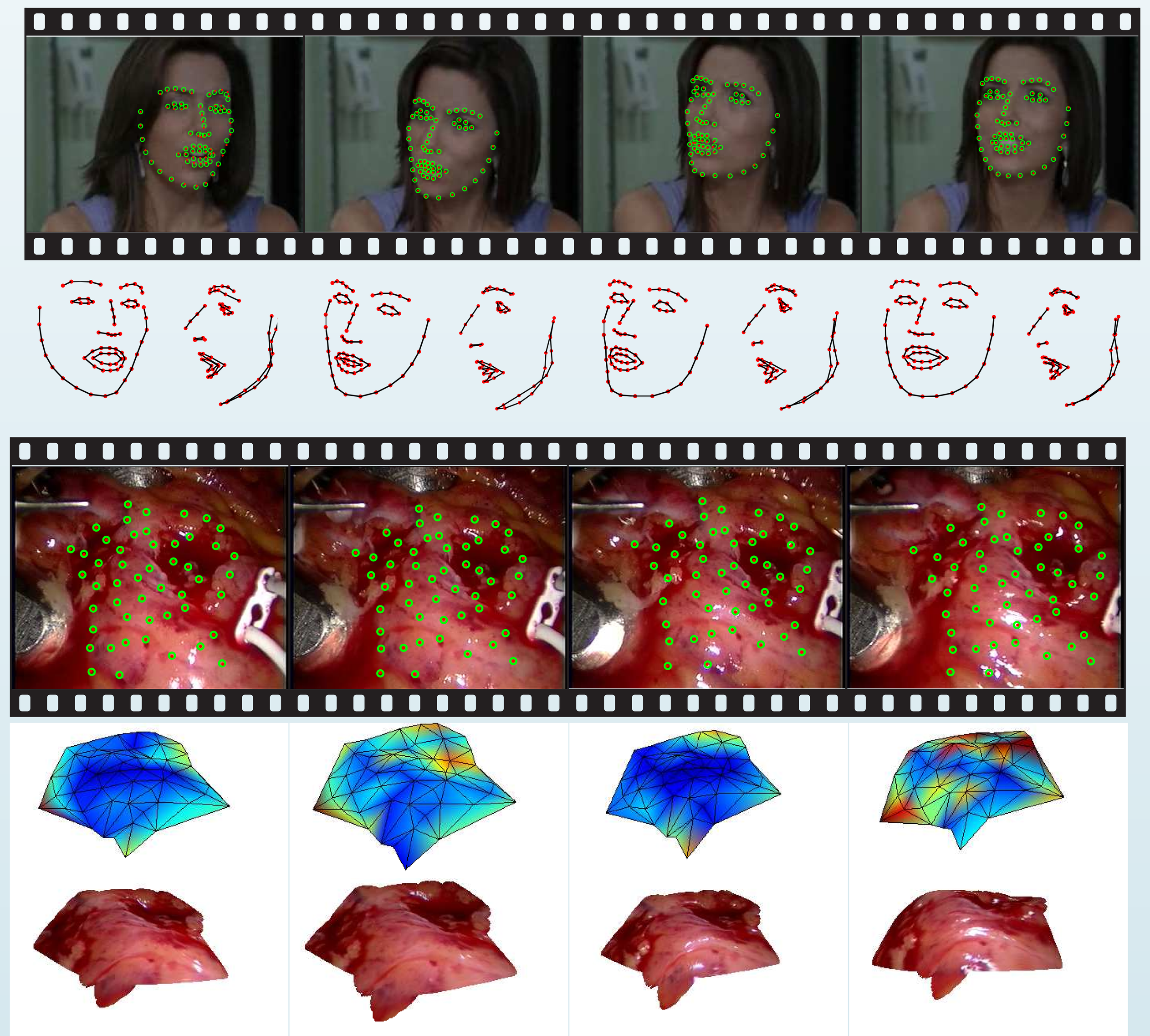
Experimental Evaluation

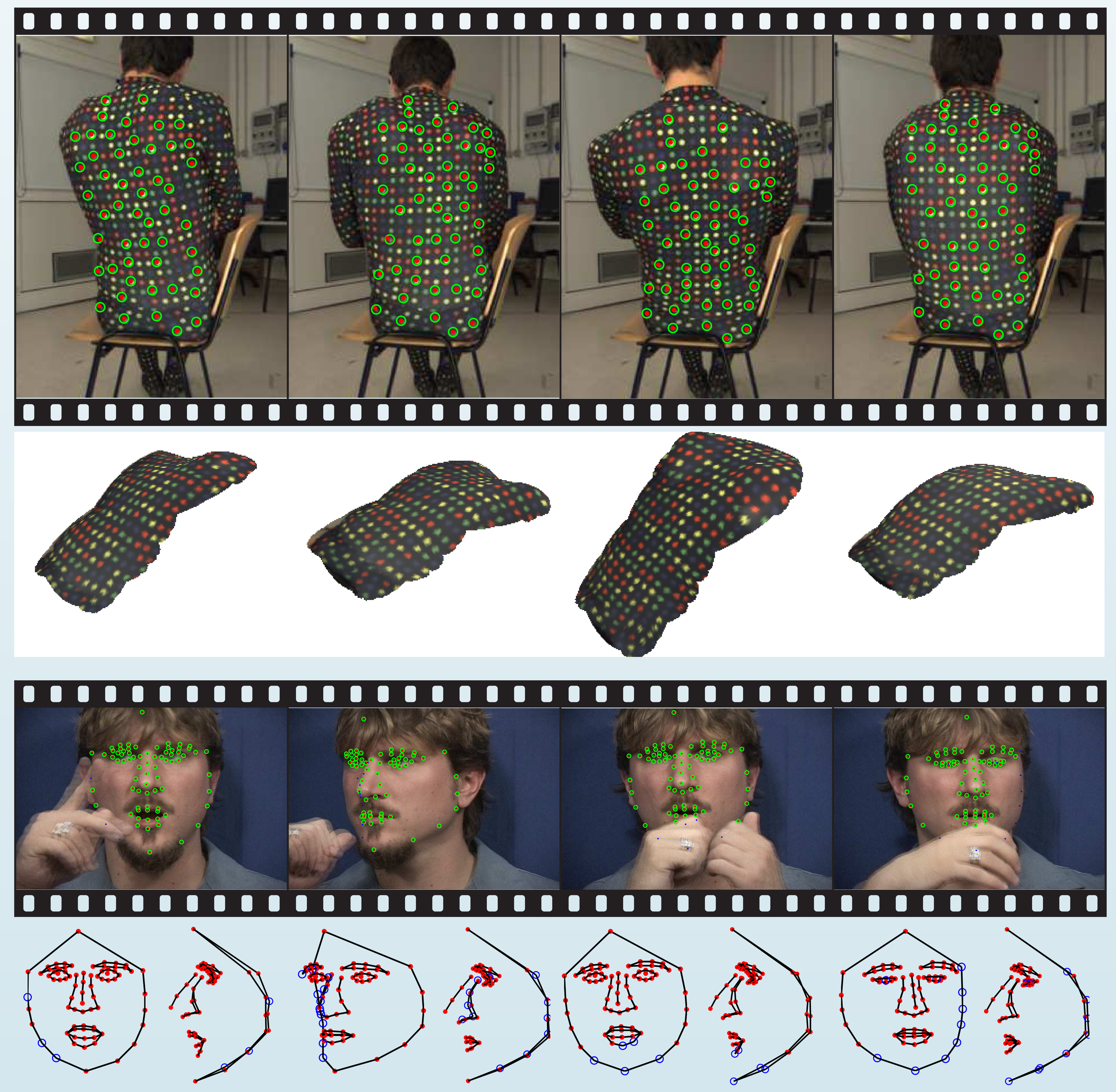
Motion Capture Sequences

Space: Met. / Seq.	Shape: EM-PPCA†	Shape: EM-LDS†	Shape: MP†	Shape: SPM†	Shape: EM-PND†	Trajectory: PTA†	Shape-Trajectory: CSF2†	Shape-Trajectory: KSTA†	Force: EM-PFS
Jacky	1.80(5)	2.79(2)	2.74(5)	1.82(7)	1.41	2.69(3)	1.93(5)	2.12(4)	1.80(7)
Face	7.30(9)	6.67(2)	3.77(7)	2.67(9)	25.79	5.79(2)	6.34(5)	6.14(8)	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)	5.29(12)
Walking	11.11(10)	27.29(2)	17.51(3)	8.02(6)	3.90	23.60(2)	6.39(5)	6.36(5)	8.54(11)
Average error:	6.11	10.77	8.69	5.09	8.80	10.05	5.66	5.59	**4.62**

† EM-PPCA [Torresani *et al.* PAMI'08], EM-PPCA [Torresani *et al.* PAMI'08], MP [Paladini *et al.* CVPR'09], SPM [Dai *et al.* CVPR'12], EM-PND [Lee *et al.* CVPR'13], PTA [Akhter *et al.* PAMI'11], CSF2 [Gotardo *et al.* CVPR'11] and KSTA [Gotardo *et al.* ICCV'11].

Real Video Sequences

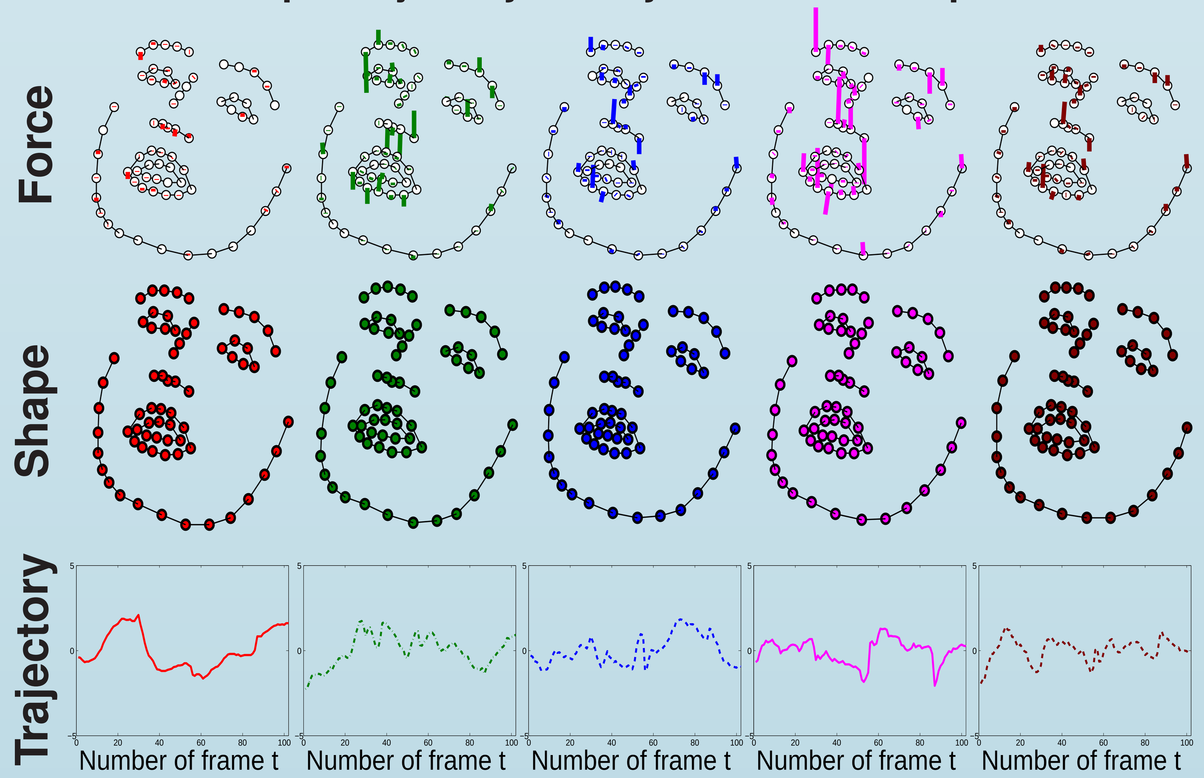




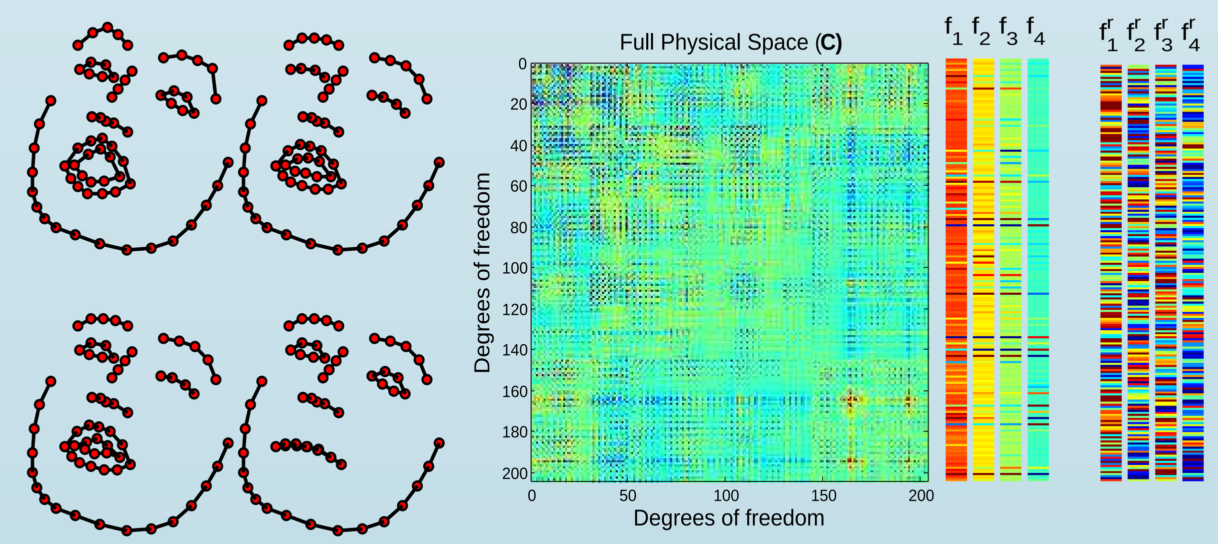
- For ASL sequence, blue points represent missing data.

Elastic Model Estimation

Force-Shape-Trajectory Duality for Actress Sequence



Full Elastic Model for Actress Sequence



- We can model shapes outside of the subspace, but in the full physical space.
- Our forces are mathematically and physically plausible. This is not the case when a random $\mathbf{C}$ mathematically plausible is used.