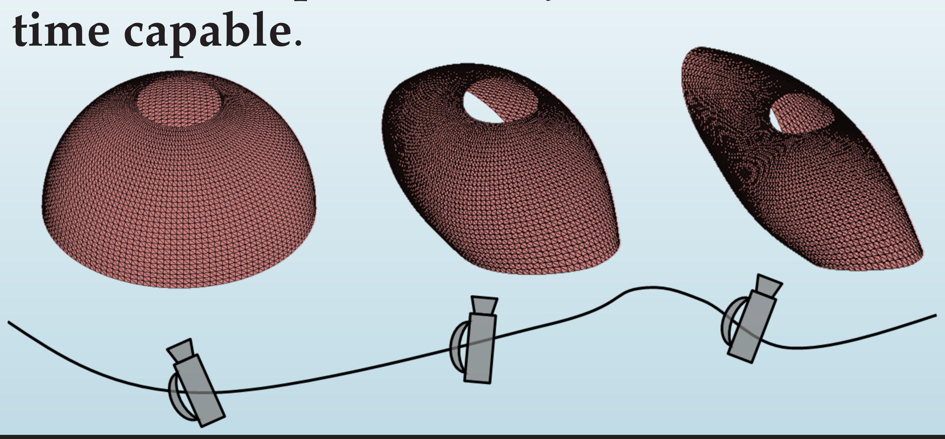
Online Dense Non-Rigid 3D Shape and Camera Motion Recovery

Antonio Agudo[1], J. M. M. Montiel[1], Lourdes Agapito[2] and Begoña Calvo[1]

[1]I3A-Universidad de Zaragoza, [2]University College London

Non-Rigid SfM

- 3D reconstruction of non-rigid objects from 2D temporal tracks in a monocular image sequence.
- So far most approaches are *batch*.
- **Our Goal:** A sequential NRS*f*M method that is **real-time capable**.



Our Contribution

- A *coarse to fine* approach to efficiently estimate the shape basis based on finite element modal analysis that allows to deal with **dense shapes**.
- An **online solution** to NRS*f*M that estimates camera pose and deformable shape on a per-frame basis.

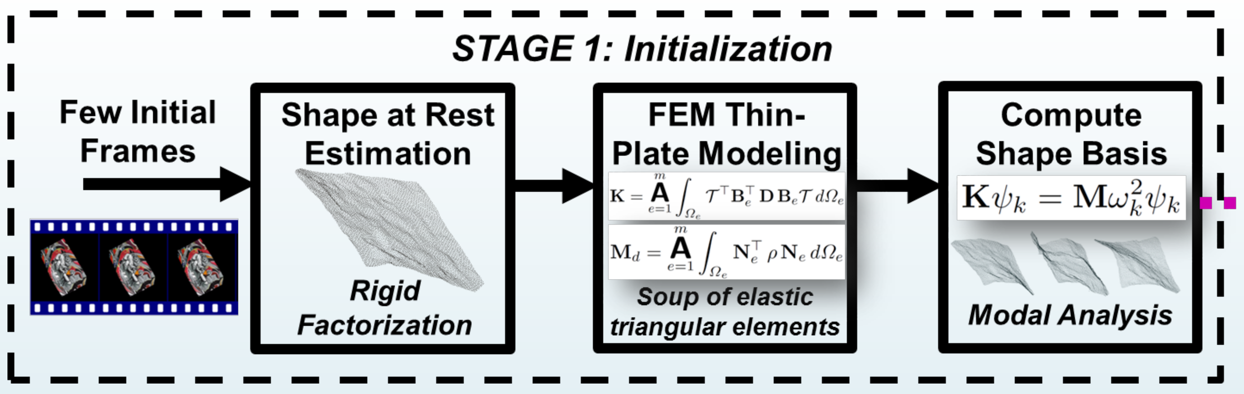
Our Approach

- **Stage 1:** Computation of the *shape basis* using a 3D shape at rest estimation. A coarse to fine modal analysis for dense 3D shape estimation.
- **Stage 2:** *Online Expectation Maximization* over a sliding temporal window of frames to optimize non-rigid shape and camera pose *as the data arrives*.
- Suitable to code a wide variety of deformations: from *inextensible* to highly **extensible surfaces**.

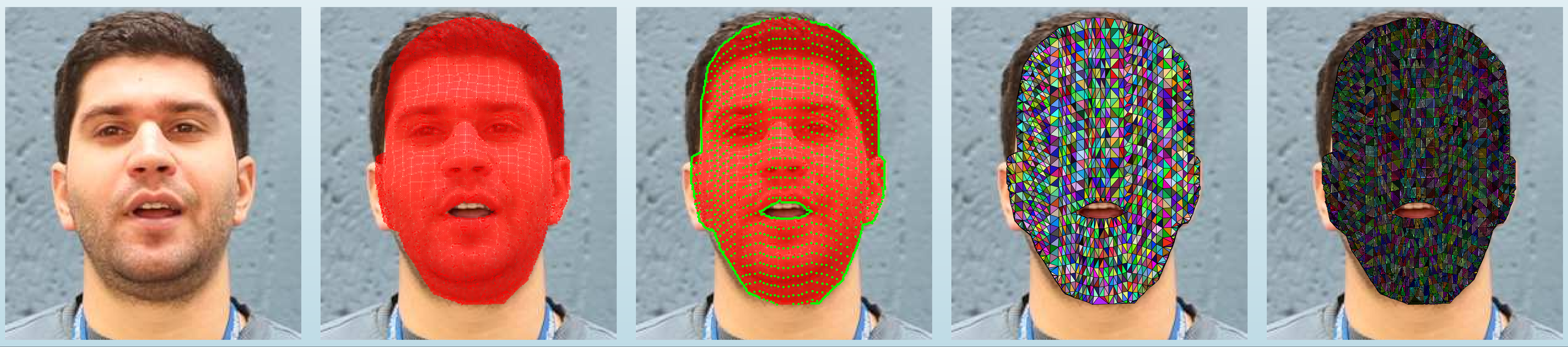
Conclusions

- Our coarse to fine approach to modal analysis allows to *extend* our method to the case of *dense (per pixel) reconstructions*.
- A modal shape basis with Gaussian priors is sufficient to model non-rigid shapes without additional temporal smoothness priors: *no tuning regularization weights*.

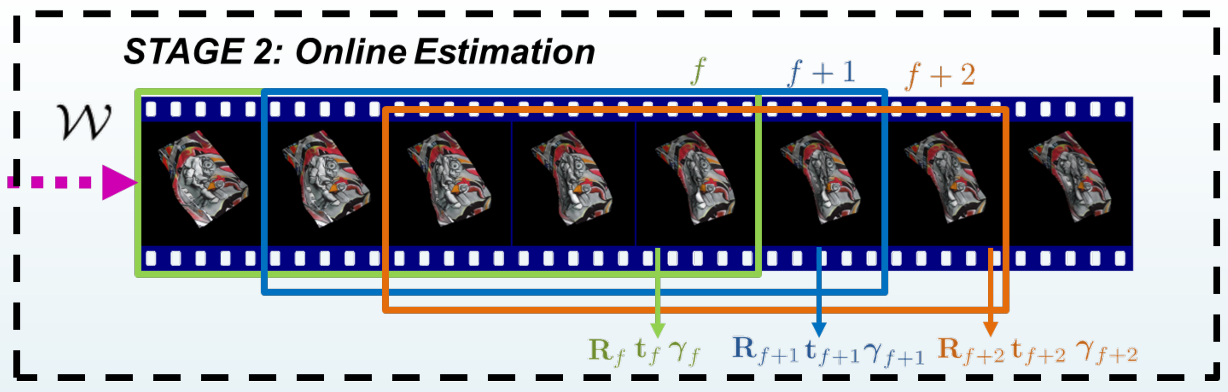
Stage 1: Coarse to Fine Approach to Modal Shape Basis Computation



- Non-rigid mode shapes are ordered by *frequency spectrum*: bending and stretching deformations.
- Bending is affordable even in dense shapes.
- Computing *stretching modes* may become prohibitive (cost and memory) in dense shapes. We propose to increase the density of some sparse modes to a down-sampled shape basis.



Stage 2: Online Sliding Window Expectation Maximization



- Orthographic camera model:

$$\mathbf{w}_f = \begin{bmatrix} u_{f1} & v_{f1} & \dots & u_{fp} & v_{fp} \end{bmatrix}^\top = \mathbf{G}_f \mathbf{S}_f + \mathbf{T}_f + \mathbf{N}_f$$

- Non-rigid 3D displacement per frame is modeled by means of a **probabilistic linear subspace** with $\gamma_f \sim \mathcal{N}(\mathbf{0};\mathbf{I}_r)$ latent variables.

We propose an **online EM-based** to solve maximum likelihood as the data arrives. The distribution to be estimated is $\mathbf{w}_f \sim \mathcal{N}\left(\mathbf{G}_f\bar{\mathbf{S}} + \mathbf{T}_f; \mathbf{G}_f\mathcal{S}\mathcal{S}^\top\mathbf{G}_f^\top + \sigma^2\mathbf{I}\right)$. In *E-step:*, we compute posterior distribution over latent variables $\gamma_{\hat{\mathcal{W}}}$ within a temporal sliding window of $\mathcal{W}$ frames:

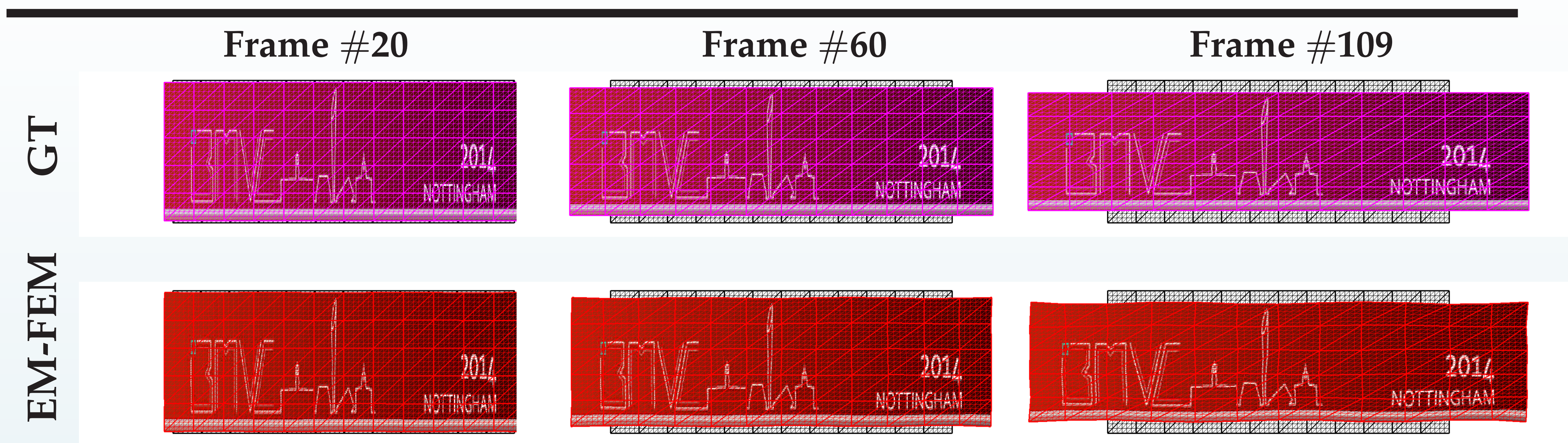
$$p\left(\gamma_{\hat{\mathcal{W}}}|\mathbf{w}_{\hat{\mathcal{W}}},\Theta_{\hat{\mathcal{W}}}\right) \sim \prod_{i=f-\mathcal{W}+1}^{f} \mathcal{N}\left(\beta_i\left(\mathbf{w}_i - \mathbf{G}_i\bar{\mathbf{S}} - \mathbf{T}_i\right); \mathbf{I}_r - \beta_i\mathbf{G}_i\mathcal{S}\right), \quad \beta_i = \mathcal{S}^\top\mathbf{G}_f^\top\left(\mathbf{G}_f\mathcal{S}\mathcal{S}^\top\mathbf{G}_f^\top + \sigma^2\mathbf{I}_r\right)^{-1}$$

In *M-step:*, we optimize expected value of log-likelihood function w.r.t model parameters Θ_i. M-steps are necessary to individually update each parameter. To update rotation matrices, we use a Riemannian manifold:

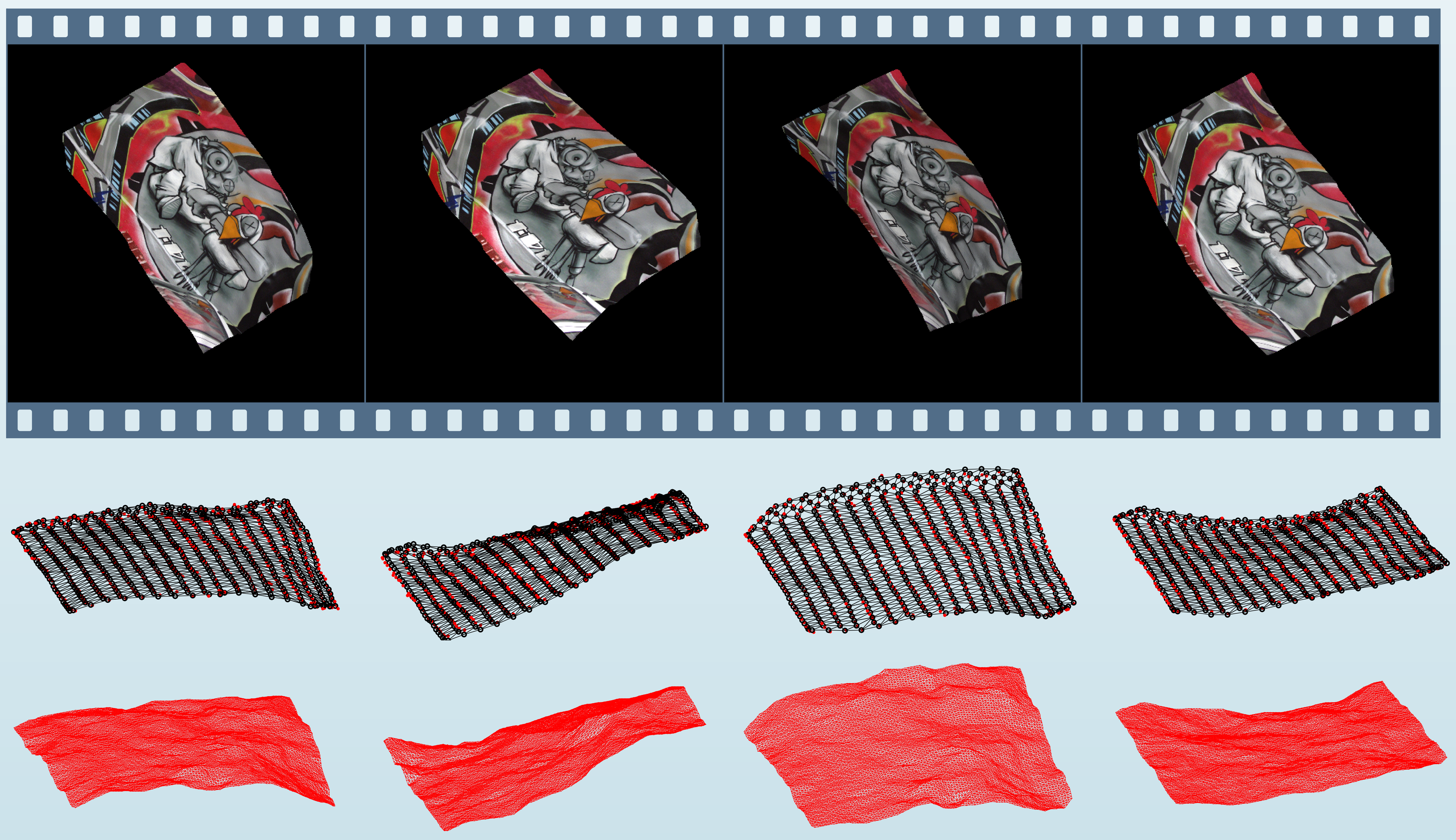
$$\arg\min_{\Theta_i} \mathbb{E}\left[-\sum_{i=f-\mathcal{W}+1}^{f} \log p\left(\mathbf{w}_i|\Theta_i\right)\right] = \arg\min_{\mathbf{G}_i,\mathbf{T}_i,\sigma^2} \frac{1}{2\sigma^2}\sum_{i=f-\mathcal{W}+1}^{f} \mathbb{E}\left[\left\|\mathbf{w}_i - \mathbf{G}_i\left(\bar{\mathbf{S}} + \mathcal{S}\gamma_i\right) - \mathbf{T}_i\right\|_2^2\right] + p\mathcal{W}\log\left(2\pi\sigma^2\right)$$

Experimental Results

Dense Stretching Ribbon Sequence (q/p) = 78/2,273 points



Dense Flag Mocap Sequence (q/p) = 594/9,622 points



Algorithm	Sparse Flag 594 points		Dense Flag 9,622 points	
	$e_{3\mathcal{D}}$(%)	*in* / *op* (sec)‡	$e_{3\mathcal{D}}$(%)	*in* / *op* (sec)‡
SBA†	7.10(38)	0.58/82.32	13.48(38)	25.67/895
BA-FEM†	3.72(10)	19.50/1.96	3.50(10)	300/75
	3.49(40)	19.50/24.83	3.29(25)	300/186
EM-FEM	3.28(10)	19.50/1.53	3.41(10)	44.62/62
	2.81(40)	19.50/2.28	3.08(25)	44.62/68

Dense Face Real Sequence (q/p) = 1,442/28,332 points

For all experiments (q/p) means number of points in sparse and dense mesh respectively.

†SBA [Paladini *et al.* ECCV'10], †BA-FEM [Agudo *et al.* CVPR'14].

‡*in*: initialization time (stage 1), *op*: online optimization time per frame (stage 2). Shape basis rank in brackets.

Acknowledgements: This work was partly funded by the MINECO projects DIP2012-32168 and DPI2011-27939-C02-01; by the ERC Starting Grant agreement 204871-HUMANIS; and by a scholarship FPU12/04886.