ONLINE DENSE NON-RIGID 3D SHAPE AND CAMERA MOTION RECOVERY

ANTONIO AGUDO1, J. M. M. MONTIEL1, LOURDES AGAPITO2 AND BEGOÑA CALVO1
13BA-UNIVERSIDAD DE ZARAGOZA, 24UNIVERSITY 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 NRFM 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 NRFM 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.

Experimental Results

DENSE STRETCHING RIBBON SEQUENCE

(q/p) = 78/2,273 points

Frame #20

Frame #60

Frame #109

DENSE FACE REAL SEQUENCE

(q/p) = 1,442/28,332 points

DENSE FLAG MOCAP SEQUENCE

(q/p) = 594/9,622 points

Algorithm | Sparse Flag | Dense Flag |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7.10(38)</td>
<td>0.38/62.32</td>
<td>13.48(38)</td>
</tr>
<tr>
<td>7.22(10)</td>
<td>19.30/1.96</td>
<td>2.70(10)</td>
</tr>
<tr>
<td>3.28(10)</td>
<td>19.30/1.53</td>
<td>3.41(10)</td>
</tr>
<tr>
<td>2.81(40)</td>
<td>19.50/2.28</td>
<td>3.08(25)</td>
</tr>
<tr>
<td>44.62/82.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

1 SBA [Paladini et al. ECCV’10], 2 BA-FEM [Agudo et al. CVPR’14]
3 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-HUMAN3D, and by a scholarship FPU12/04886.

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

Stage 2: Online Sliding Window Expectation Maximization

We propose an online EM-based to solve maximum likelihood as the data arrives. The distribution to be estimated is $w_f \sim N \left( G \hat{f}_f + T_f + N_f \right)$. In E-step, we compute posterior distribution over latent variables $G_f$ within a temporal sliding window of $W$ frames:

$$p(\gamma_f | \tilde{w}_f, \theta) \sim \prod_{i=f-W+1}^{f} N(\beta_i (\tilde{w}_i - G_i S_i - T_i), \Sigma_i)$$

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

$$\arg\min_{\theta} \mathbb{E} \left[ - \sum_{i=f-W+1}^{f} \log p(w_i | \theta_i) \right] \approx \arg\min_{G_i, \Sigma_i} \frac{1}{2\pi} \sum_{i=f-W+1}^{f} \mathbb{E} \left[ \|w_i - G_i (S + \gamma_i) - T_i\|^2 \right] + pW \log(2\pi)$$

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.