

Dense Variational Reconstruction of Non-Rigid Surfaces from Monocular Video

Yufeng Jiang

1. Optimisation of the proposed energy

In this section, authors solve the minimization of the proposed energy, that can be written as follows [4]:

$$\min_{\mathbf{S}, \mathbf{R}} \frac{\lambda}{2} \|\mathbf{W} - \mathbf{R}\mathbf{S}\|_{\mathcal{F}}^2 + \sum_{f,i,p} \|\nabla S_f^i(p)\| + \tau \|P(\mathbf{S})\|_* \quad (1)$$

Note that this energy is biconvex because of the bilinear term in E_{data} . They alternate between the estimation of the motion matrix \mathbf{R} and the shape matrix \mathbf{S} leaving the other fixed to minimize it.

1.1. Motion matrix estimation

The first alternation step involves the refinement of the motion \mathbf{R} by minimising Eq. 1 at the basis of the known shape matrix \mathbf{S} . Authors parameterize the camera matrices \mathbf{R}_f using quaternions, which guarantee orthonormality. Only the term E_{data} depends on \mathbf{R} and they minimise it using Levenberg-Marquardt.

1.2. Shape estimation

The second alternation step assumes that the motion matrix \mathbf{R} is known and minimises Eq. 1 to the shape matrix \mathbf{S} . It is non-trivial to minimise it using standard gradient descent methods. To facilitate such minimisation they use proximal splitting techniques [3] to decouple the trace norm and TV regularisation parts of the energy. They introduce an auxiliary variable $\bar{\mathbf{S}}$ by alternatin between the following two minimizations [4]:

$$\min_{\mathbf{S}} \frac{1}{2\theta} \|\mathbf{S} - \bar{\mathbf{S}}\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \|\mathbf{W} - \mathbf{R}\mathbf{S}\|_{\mathcal{F}}^2 + \sum_{f,i,p} \|\nabla S_f^i(p)\| \quad (2)$$

$$\min_{\bar{\mathbf{S}}} \frac{1}{2\theta} \|\mathbf{S} - \bar{\mathbf{S}}\|_{\mathcal{F}}^2 + \tau \|P(\bar{\mathbf{S}})\|_* \quad (3)$$

where θ is a quadratic relaxation parameter that is relatively small so that the optimal \mathbf{S} and $\bar{\mathbf{S}}$ are close. They can efficiently solve the problems Eq. 2 and Eq. 3 using convex optimization techniques.

Solving problem The energy in Eq. 2 is convex but because of the TV regularisation term it is non-differentiable. However, using the Legendre-Fenchel transform, they can dualise the regularization term in Eq. 2 and rewrite the corresponding minimisation in its form as [4]:

$$\min_{\mathbf{S}} \max_{\mathbf{q}} \left\{ \frac{1}{2\theta} \|\mathbf{S} - \bar{\mathbf{S}}\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \|\mathbf{W} - \mathbf{R}\mathbf{S}\|_{\mathcal{F}}^2 + \sum_{f,i,p} \{S_{fi}(p) \nabla^* \mathbf{q}_f^i(p) - \delta(\mathbf{q}_f^i(p))\} \right\} \quad (4)$$

where \mathbf{q} is the dual variable that contains the 2-vectors $\mathbf{q}_f^i(p)$, for each frame f , coordinate i and pixel p . Also, ∇^* is the adjoint of the discrete gradient operator ∇ and can be expressed as $\nabla^* = -div(\cdot)$, where div is the divergence operator, after discretisation using backward differences [2]. Finally, $\delta(\cdot)$ is the indicator function of the unit ball [4]:

$$\delta(\mathbf{s}) = \begin{cases} 0 & \text{if } \|\mathbf{s}\| \leq 1 \\ \infty & \text{if } \|\mathbf{s}\| > 1 \end{cases} \quad (5)$$

Following duality principles [2], they solve the saddle point problem Eq. 4 by deriving a primal-dual algorithm.

2. Experimental evaluation

Tab. 1 shows the results for each method. Authors define the normalised per frame RMS error for the reconstructed 3D shape \mathbf{S}_f with respect to the corresponding ground truth shape \mathbf{S}_f^{GT} as: $e_{3D} = \frac{\|\mathbf{S}_f - \mathbf{S}_f^{GT}\|_{\mathcal{F}}}{\|\mathbf{S}_f^{GT}\|_{\mathcal{F}}}$. They report the mean RMS error over all frames. They also provide the rank of the reconstructed result for each case¹

Sequence 1. In this 10 frame long sequence, each frame corresponds to a different facial expression. The results shown on the first row of Tab. 1 reveal that both MP and TB fail to reconstruct this sequence while their approach performs well.

¹Since the trace norm approximation of the rank results in some singular values being small but not exactly 0, for our method (case $\tau = 0$) we report the rank of the optimal 3D reconstructions retaining the singular values of $P(\mathbf{S})$ that explain 99% of the shape variance.



Figure 1. 3D reconstruction results of the real face sequence. (a) Input images. (b) Corresponding 3D shapes from original viewpoint of the camera while the face rotates and deforms. (c-d) Shape deformation as observed from three-quarter and profile views respectively (after taking out the rotational component of the face). (e) Rendered surfaces from a different viewpoint, using computed deformations and rotations with augmented texture placed on the reference image. See supplementary material for videos.

	TB [1]	MP [5]	Ours	Ours($\tau = 0$)
Seq.1	18.38%	19.44%	4.01%	4.13%
Seq.2	7.47%	4.87%	3.45%	3.76%
Seq.3	4.50%	5.13%	2.60%	3.32%
Seq.4	6.61%	5.81%	2.81%	3.89%

Table 1. Quantitative evaluation on 4 synthetic face sequences. We show average RMS 3D reconstruction errors for TB [1], MP [5] and our approach. In all cases, the rank of the reconstructed result is shown in brackets.

Sequence 2. This simple fact allows MP and TB to reduce their errors substantially, which indicates that large rotations help in NRSfM. The main drawback is that establishing 2D correspondences under this amount of rotation

would be unrealistic in a real world scenario due to occlusions.

Sequence 3. 99 frame long sequence generated by linearly interpolating between pairs of views to obtain smooth 3D deformations. Realistic rotations that contain some high frequencies are simulated by upscaling their algorithm on the real face sequence just as shown at Fig. 1.

Sequence 4. This setup is equivalent to the previous one with the exception that the rotations are projected onto a low frequency subspace to simulate the case of both smooth rotations and deformations.

References

- [1] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from mo-

- tion. *IEEE TPAMI*, 33(7):1442–1456, 2011. [2](#)
- [2] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *JMIV*, 40(1):120–145, 2011. [1](#)
- [3] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. In *CVPR*, 2012. [1](#)
- [4] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *CVPR*, 2013. [1](#)
- [5] M. Paladini, A. D. Bue, J. Xavier, L. Agapito, M. Stosic, and M. Dodig. Optimal metric projections for deformable and articulated structure-from-motion. *IJCV*, 96(2):252–276, 2012. [2](#)