Dense Variational Reconstruction of Non-Rigid Surfaces from Monocular Video

Yufeng Jiang

1. Related work

Most methods are required to add the extra pariors on shape or camera matrices to solve some problems, like temporal smoothness [7], near rigidity [7], smooth time trajectory basis [2], basis priors [3]. However, recently it has proved that the orthonormality constraints on the camera matrix were sufficient except the low-rank shape prior [1]. And there have been some practice methods [4, 6].

Most NRSfM methods add low-rank constraint by parameterizing the non-rigid shapes that using a pre-defined number basis shape and a time-varying coefficients. But, recently, Dai *et al.* [4] research that the non-rigid shape leads to the extra basis ambiguities on basis and coefficients. Instead, they put the low-rank shape constraints to the time-varying shape matrix directly by tracing norm minimization to be a tightest possible relaxation of rank minimization.

Dense 2D correspondences: The Dense 2D correspondences need to be established in the image sequence. In this paper, authors use the recent research of robust and dense variational multi-frame motion estimation. Although most 2D motions motion about the image sequence research the estimation of frame-to-frame optical flow fields. But there is a new method usiing Lagrangian recently. This method can handle the estimation of long-term trajectories that associate each world point with its entire 2D image trajectory over an image sequence.

Dense 3D reconstruction: Authors propose a new variational energy optimazation method to solve the alternation between the camera matrices and the non-rigid shape for every frame in the sequence at the basic of giving dense correspondences as input. Their energy combines: (i) a geometric data term that minimizes image reprojection error, (ii) a trace norm term that minimizes the rank of the time-evolving shape matrix and (iii) an edge-preseving spatial regularization term that provides smooth 3D shapes.

2. Dense reconstruction with trace norm and spatial smoothness prior

To solve the dense NRSfM problem as defined in the previous section, authors propose to minimize an energy of the following form, jointly with respect to the motion matrix $\mathbf{R}$ and the shape matrix $\mathbf{S}$, they are used in [5]:

$$E(\mathbf{R},\mathbf{S}) = \lambda E_{data}(\mathbf{R},\mathbf{S}) + E_{reg}(\mathbf{S}) + \tau E_{trace}(\mathbf{S}) \quad (1)$$

where E_{data} is a data attachment term, E_{trace} favours a low-rank shape matrix, and E_{reg} is a term for the spatial regularization of the trajectories in $\mathbf{S}$. The positive constants λ and τ are weights that control the balance between these terms. They now describe each of these terms in detail.

The **first term** (E_{data}) is a quadratic penalty of the image reprojection error [5]

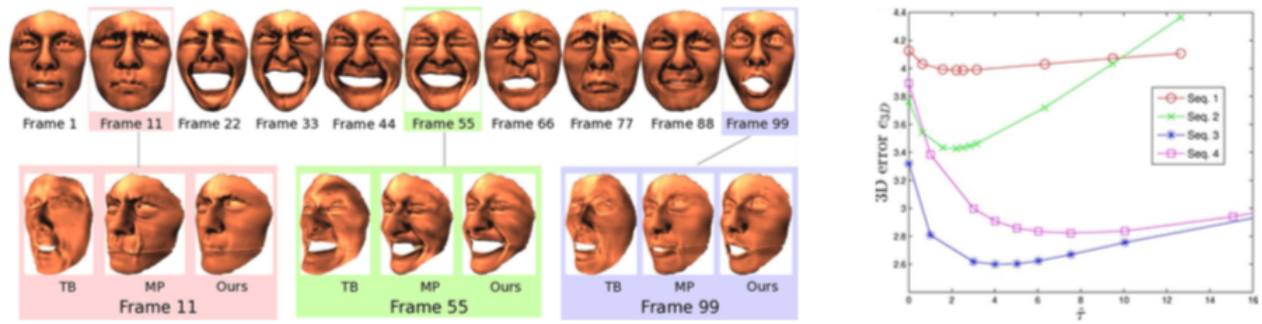
$$E_{data} = \frac{1}{2}||\mathbf{W} - \mathbf{RS}||^2_{\mathcal{F}} \quad (2)$$

where $||\cdot||_{\mathcal{F}}$ denotes the Frobenius norm of a matrix. This term penalizes deviations of the image measurements.

The **second term** (E_{reg}) enforces edge-preserving spatial regulatrization of the dense 3D trajectories that constitute the columns of $\mathbf{S}$. To formulate this term, let i be an index (i = 1,2,3) that selects the X, Y or Z coordinate of a 3D point. S^i_j will then be the i-th row of the 3D shape $\mathbf{S_f}$. Since the 3D points that they reconstruct are a associated with projected pixel locations on the reference image I_{ref}, each element of S^i_j is associated with a specific pixel of I_{ref}. By arranging these elements in the image grid of I_{ref}, they consider S^i_j as a discrete 2D iamge of the same size as I_{ref}. They now denote the 2D gradient of this image at pixel p by $\nabla S^i_f(p)$. Following [3], they define this discrete gradient using forward differences in both horizontal and vertical directions. They define E_{reg} as the summation of discretized Total Variation regularizers $TV\{\cdot\}$ [5]:

$$E_{reg} = \sum_{f=1}^{F}\sum_{i=1}^{3} TV\{S^i_f\} = \sum_{f=1}^{F}\sum_{i=1}^{3}\sum_{p=1}^{N} ||\nabla S^i_f(p)|| \quad (3)$$

The **third term** (E_{trace}) penalises the number of independent shapes needed to represent the deformable scene. This is based on the realistic assumption that the shapes that a deforming object undergoes over time lie on a low-dimensionaal linear subspace. Instead of using



(a) Ground truth 3D shapes (top row) and dense 3D reconstructions for selected frames (bottom row) in Sequence 4 using TB [2], MP [20] and our approach.See supplementary material for videos.

(b) Normalized RMS 3D error with varying trace norm strength for synthetic experiments.

Figure 1. Results on synthetic sequences.

some a priori dimension for this subspace to enforce a hard rank constraint, similarly to [4], they penalize the rank of the $F \times 3N$ matrix $P(\mathbf{S})$. This is implemented using the $trace\ norm || \cdot ||$, which is the tightest convex relation of the rank of a matrix and is given by the sum of its singular values Λ_j [5]:

$$E_{trace} = ||P(\mathbf{S})|| = \sum_{j=1}^{min(F,3N)} \Lambda_j \tag{4}$$

3. Synthetic face sequences

In this section authors evaluate the performance of their method quantitatively on sequences generated using dense ground truth 3D data of a deforming face. They generate four different sequences that differ in the number of frames and the range and smoothness of the camera rotations and deformations, just as shown at Fig. 1. By projecting the 3D data onto an iamge using an orthographic camera, they derived dense 2D tracks, which they feeded as input to the NRSfM estimations.

References

[1] I. Akhter, Y. Sheikh, and S. Khan. In defense of orthonormality constraints for nonrigid structure from motion. In *CVPR*, 2009. 1

[2] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from motion. *IEEE TPAMI*, 33(7):1442–1456, 2011. 1

[3] 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

[4] Y. Dai, H. Li, and M. He. A simple prior-free method for nonrigid structure-from-motion factorization. In *CVPR*, 2012. 1, 2

[5] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *CVPR*, 2013. 1, 2

[6] 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. 1

[7] L. Torresani, A. Hertzmann, and C. Bregler. Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE TPAMI*, 30(5):878–892, 2008. 1