Dense Non-rigid Structure-from-Motion Made Easy
A Spatial-Temporal Smoothness based Solution

Yuchao Dai [1,2], Huizhong Deng[1] and Mingyi He[2]
[1] Research School of Engineering, Australian National University
[2] School of Electronics and Information, Northwestern Polytechnical University.

Introduction

Motivation:

- Great progress in sparse non-rigid 3D reconstruction while dense reconstruction is still challenging;
- Existing methods to dense reconstruction use complex optimization (SDP and non-convex optimization);

Our contributions:

- A unified framework to **dense non-rigid 3D reconstruction** exploiting both spatial and temporal smoothness;
- The cost function has been robustified to deal with real world noise and outliers;
- Our method achieves competitive performance with state-of-the-art dense NRSfM methods.
- The implementation involves solving a series of least squares problems, thus **making dense NRSfM easy**.

Formulation and Solution

Problem Statement: Dense NRSfM takes a 2D video obtained by a monocular camera as input, with image frames $\mathtt{I}_1 \cdots \mathtt{I}_F$ each containing P pixels. Stacking all the feature tracks for all frames gives:

$$\mathtt{W} = \mathtt{RS}, \tag{1}$$

Dense NRSfM aims at simultaneously recovering camera motion $\mathtt{R}$ and non-rigid shape $\mathtt{S}$ from $\mathtt{W}$. The problem is inherently under-determined. Therefore, extra constraints are needed to regularize the problem.

Temporal Smoothness Revisited: We revisit the temporal smoothness and would like to argue that this simple strategy could be pretty efficient in achieving comparable performance with complex convex optimization or ADMM based methods.
Non-rigid shape recovery is formulated as

$$\min_{\mathtt{S}} \frac{1}{2}\|\mathtt{W}-\mathtt{RS}\|_F^2 + \frac{1}{2}\lambda\|\mathtt{HS}\|_F^2. \tag{2}$$

We could apply different smooth operators $\mathtt{H}$ to characterize various kinds of smoothness in temporal direction.
The resultant optimization problem in Eq.-(2) admits an analytical (closed-form) solution,

$$\mathtt{S}_{\text{smooth}} = (\mathtt{R}^T\mathtt{R} + \lambda\mathtt{H}^T\mathtt{H})^{\dagger}\mathtt{R}^T\mathtt{W}. \tag{3}$$

Formulation and Solution

Spatial Smoothness Simplified:

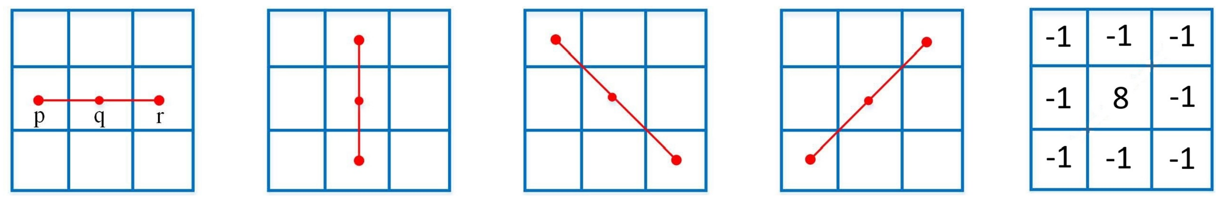


Figure 1: Our Laplacian filter (far right): 8-direction, sum of the 4 basic Laplacian filters.

The temporal smoothness constrains the dense non-rigid reconstruction from the temporal dimension, the smoothness of 3D trajectory. However, it could not regularize the 3D shape at each frame. We propose a simple filtering mechanism, namely Laplacian filter, which enforces spatial smoothness locally in the 3D shape space. The filtering output is defined as $\mathtt{Avec}(\mathtt{S})$.

Optimization Robustified: Noise and outliers are inevitable in real world measurements. Dense NRSfM methods must handle them robustly. Most of the existing methods apply L_2 on the data term, thus could not handle noise and outliers well. We propose to replace the L_2 norm with L_1 norm, thus increasing the robustness of the data term $\|\mathtt{W}-\mathtt{RS}\|_1$.

To deal with the convex L_1 norm efficiently, we propose to use iterative reweighted least square (IRLS), where we solve for a least square problem in each iteration.

Spatial-Temporal smoothness constraint: By enforcing the spatial-temporal smoothness constraint and applying the robust L_1 norm on data term, we reach:

$$\min_{\mathtt{S}} \|\mathtt{W}-\mathtt{RS}\|_1 + \lambda_1\|\mathtt{HS}\|_F^2 + \lambda_2\|\mathtt{Avec}(\mathtt{S})\|_F^2, \tag{4}$$

Under IRLS formulation, we solve the following least square problem in each iteration:

$$\min_{\mathtt{S}^{it}} \|\mathtt{E}(\mathtt{W}-\mathtt{RS})\|_F^2 + \lambda_1\|\mathtt{HS}\|_F^2 + \lambda_2\|\mathtt{Avec}(\mathtt{S})\|_F^2. \tag{5}$$

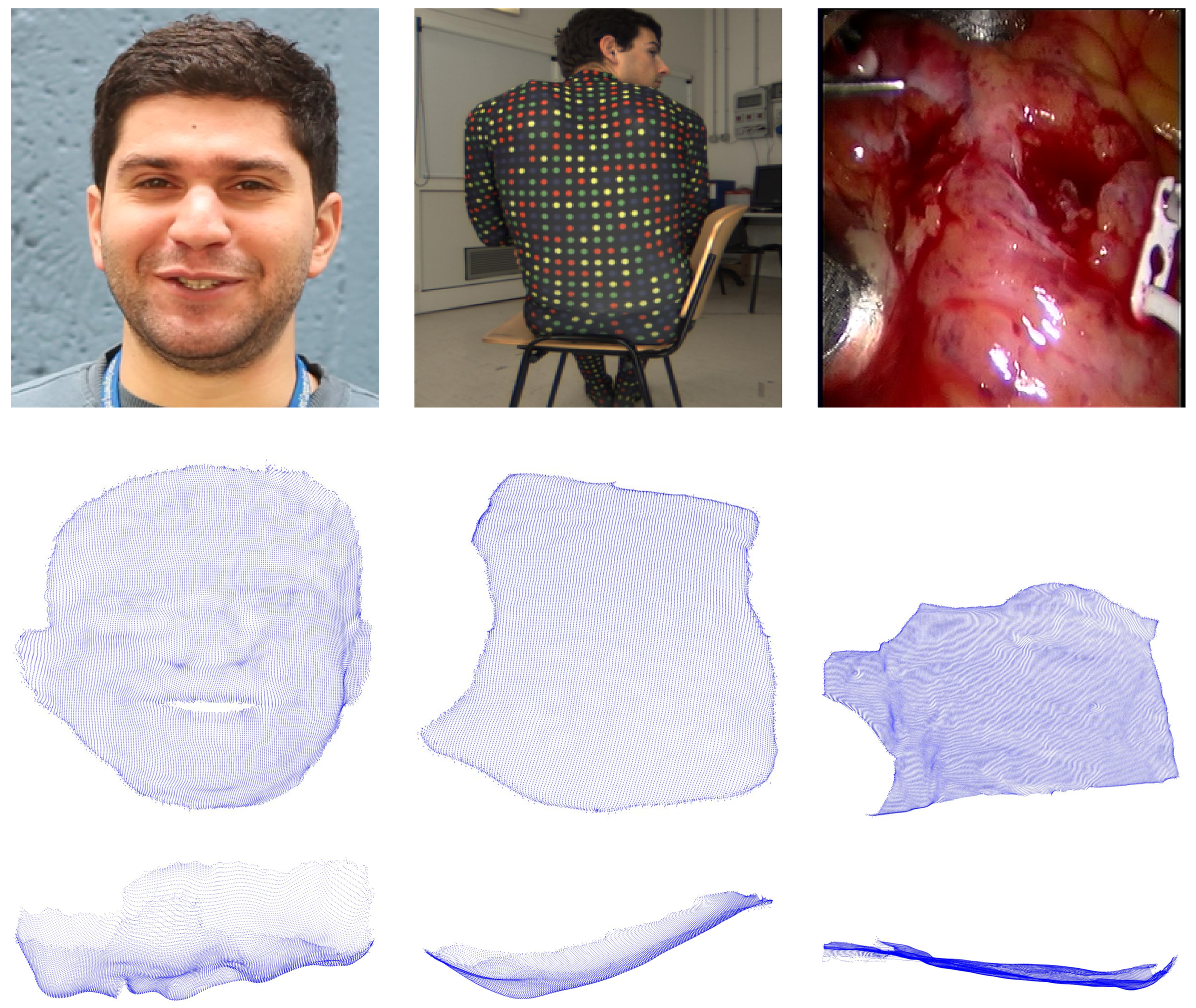
Instead, we propose to solve the least square problem with gradient descent, where the gradient is derived as:

$$g(\mathtt{S}) = 2\mathtt{R}^T\mathtt{E}^T\mathtt{ERS} - 2\mathtt{R}^T\mathtt{W} + 2\lambda_1\mathtt{H}^T\mathtt{HS} + 2\lambda_2\text{ivec}((\mathtt{A}^T\mathtt{A})\text{vec}(\mathtt{S})), \tag{6}$$

ivec denotes the inverse operator of vectorization, which transforms a vector to matrix with proper dimension.

Experiments

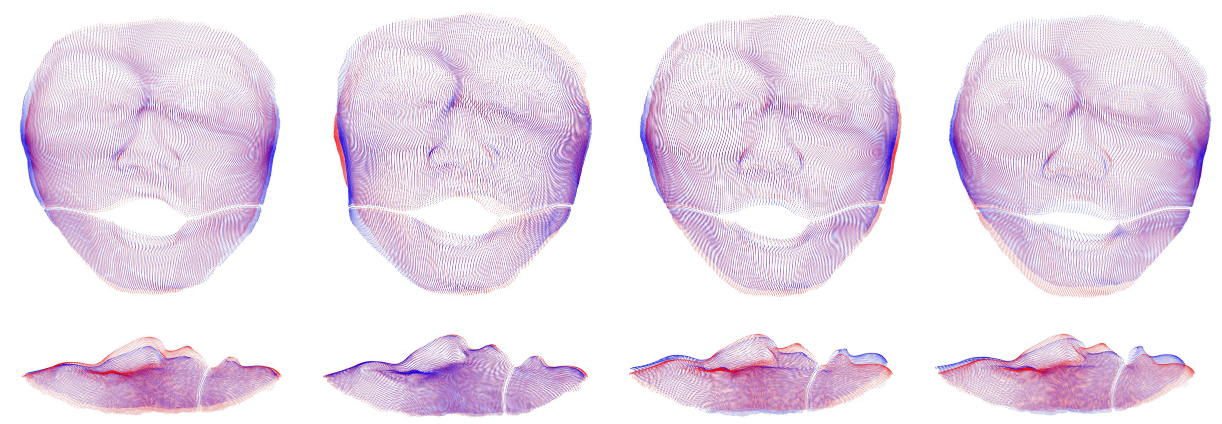
Real Image Results



(a) Face (b) Back (c) Heart

Figure 2: Dense non-rigid 3D reconstruction results on the real image sequences (Face, Back and Heart). Top row: input 2D images. Middle row: front views of the respective sequences. Bottom row: side views.

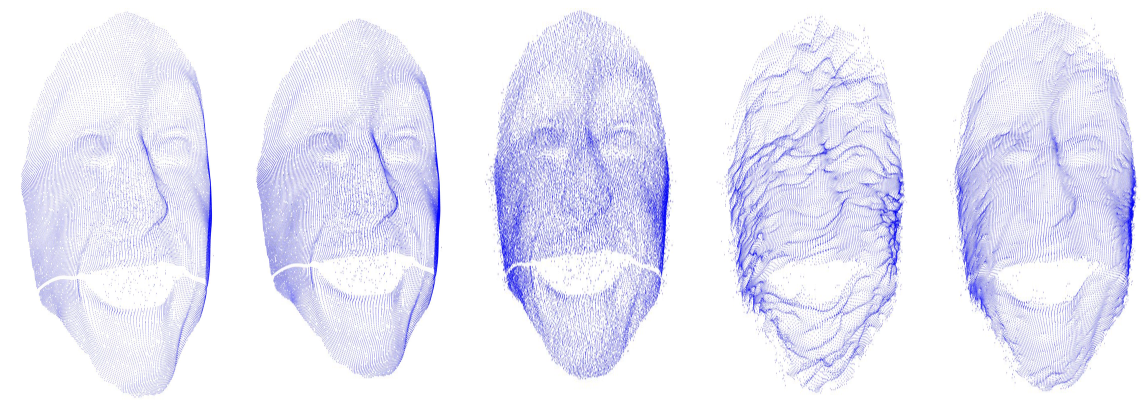
Synthetic Image Results



(a) Seq 1 (b) Seq 2 (c) Seq 3 (d) Seq 4

Figure 3: Dense non-rigid reconstruction results on synthetic face sequences. Red: ground truth; Blue: our results. Top row: front view. Bottom row: side view.

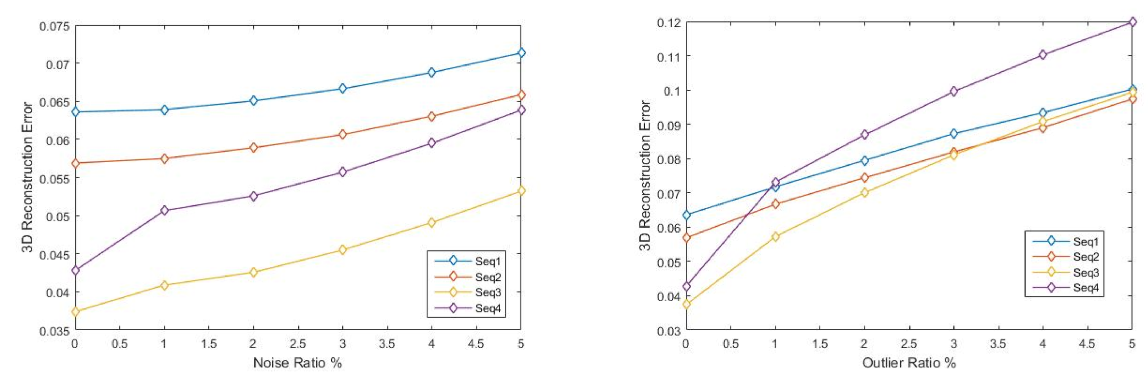
Performance Evolution



(a) Input W (b) Pseudo-inverse (0.5084) (c) Temporal (0.1835) (d) Spatial-temporal (0.0757) (e) Final (0.0700)

Figure 4: By enforcing the temporal smoothness, spatial smoothness and applying robust cost function, the dense 3D reconstruction has been gradually improved.

Robust to Noise and Outliers



(a) (b)

Figure 5: Performance evaluation under noise and outliers. (a) Experimental results (3D error) w.r.t noise levels. (b) 3D reconstruction error w.r.t outlier ratios.

Table 1: Quantitative evaluation on 4 synthetic face sequences. (Average RMS 3D reconstruction error.)

Dataset	PTA	MP	DV	Ours
Seq1	0.2431	0.2575	0.0531	0.0636
Seq2	0.0988	0.0644	0.0457	0.0569
Seq3	0.0596	0.0682	0.0346	0.0374
Seq4	0.0877	0.0772	0.0379	0.0428

References

- Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. IJCV, 2014.
- R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. CVPR, 2013.

Acknowledgments

This work was supported by Australian Research Council grants (DE140100180) and Natural Science Foundation of China (61420106007).