Dense Variational Reconstruction of Non-Rigid Surfaces from Monocular Video

Yufeng Jiang

Abstract

This paper offers the first variational approach to the problem of dense 3D reconstruction of non-rigid surfaces from a nomocular video sequence. To estimate dense low-rank smooth 3D shapes for every frame with the camera motion matrices compared to the dense 2D, authors formulate non-rigid structure from motion (NRSfM) as a global variational energy minimization problem.

NRSfM model the low-rank non-rigid shape using a fixed number of basic shapes and corresponding coefficients. This approach is different from the tradiitonal factorization. By using the new approach, authors can minimize the rank of the matrix of time-varying shapes directly via trace norm minimization. And they can use an edge preserving total-variation regularization term to obtain spatially smooth shapes for every frame by the constraint of the low-rank.

1. Introduction

The key problem in computer vision is recovering completely dense 3D models of a scene observed by a moving camera. In this image, they can estimate the 3D location obtained for every pixel. With dense approaches to $multi-view\ stereo$ (MVS) [3, 7] can acquire highly accurate models from a collation of fully calibrated images, rigid structure from motion (SfM) algorithms have made siginificant progress towards this goal.

The most SfM approaches produre impressive and detailed models of 3D objects from video sequences. However, they have a common drawback that they can only handle scenes with rigid objects. Considering from this aspect, the non-rigid structure from motion (NRSfM) focuses on the reconstruction of deformable objects from video. Recently, the ability of reconstructing strong realistic non-rigid motions [2, 6, 8] has advanced siginificantly. The feature of NRSfM methods is typically sparse, *i.e.* they can only reconstruct a small set of salient points. This results in very low resolution 3D models that cannot capture fine detail.

In their seminal work, Bergler *et al.* [1] first proposed the solution to non-rigid sturcture from motion (NRSfM). Their though was incorporate a statistical shape before evolving non-rigid shape into the factorization formulation. Previous attempts to dense NRSfM have come from: piecewise approaches that reconstruct local patches using simple local models, and template based approaches that require a 3D template. And they also require a post-processing step to stitch all the local reconstructions into a single smooth surface.

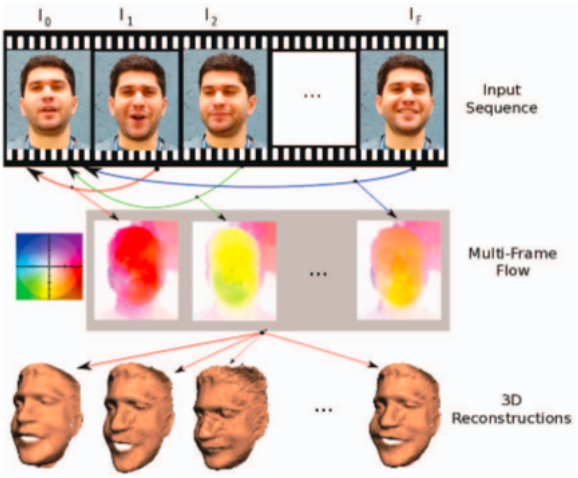

Figure 1. Our proposed pipeline for dense NRSfM. The first row shows the input image stream. Dense long-term 2D trajectories are first computed for every pixel in the reference frame using [5] and used as input to our dense NRSfM algorithm.

Their system Firstly, they acquire a video sequence with a single camera as input. And then, their approach can provide a complete pipeline for dense NRSfM integrating 2D image matching and 3D reconstruction in two steps which can be shown at Fig. 1 cited from [4].

2. Problem formulation

Consider an image sequence $I_1, \ldots, I_F$ of F frames with N pixels each where I_{ref} is chosen to be the reference frame. This pixel is the first frame at the most time. The input of their algorithm is a set of $dense$ 2D tracks that have been estimated in a pre-processing step. For every pixel in the reference image I_{ref}, each track encodes its image location in the subsequent F frames. Let $p = 1, \ldots, N$ be an index for the pixels and (x_{fp}, y_{yp}) the location of the p-

th point in the f-th frame, $f = 1, \ldots, F$. In the reference frame, this location coincides with the location of the p-th pixel on the image grid.

They adopt an orthographic camera model, where the 2×3 camera matrix $\mathbf{R}_f$ projects 3D points (X_{fp}, Y_{yp}, Z_{fp}) onto image frame f following the projection equation used in [4]:

$$\underbrace{\begin{bmatrix} x_{f1} & \ldots & x_{fN} \\ y_{f1} & \ldots & y_{fN} \end{bmatrix}}_{W_f} = \mathbf{R}_f \underbrace{\begin{bmatrix} X_{f1} & \ldots & X_{fN} \\ Y_{f1} & \ldots & Y_{fN} \\ Z_{f1} & \ldots & Z_{fN} \end{bmatrix}}_{S_f} \quad (1)$$

where $\mathbf{W}_f$ stores the 2D locaions of all N points in frame f and the $3 \times N$ matrix $\mathbf{S}_f$ represents the 3D shape observed in the frame f. Since the objects they are observing are non-rigid, the shape matrix $\mathbf{S}_f$ will be different for each frame. They hace eliminated the translation component from Eq. 1 by registering the image coordinates to the centroid in each frame f. They can formulate the projection used in [4]of the time varying shapes in all the frames as:

$$\mathbf{W} = \mathbf{RS} \quad (2)$$

where $\mathbf{W}$ is the input measurement matrix that contains the full 2D tracks, $\mathbf{S}$ is the non-rigid shape matrix and $\mathbf{R}$ is the motion matrix. These three matrix can be explained in [4]

$$\underbrace{\mathbf{W}}_{2F \times N} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_F \end{bmatrix}, \underbrace{\mathbf{R}}_{2F \times 3F} = \begin{bmatrix} \mathbf{R}_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \mathbf{R}_F \end{bmatrix},$$

$$\underbrace{\mathbf{S}}_{3F \times N} = \begin{bmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_F \end{bmatrix}. \quad (3)$$

References

[1] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *CVPR*, 2000. 1

[2] T. Collins and A. Bartoli. Locally affine and planar deformable surface reconstruction from video. In *VMV*, 2010. 1

[3] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE TPAMI*, 32(8):1362–1376, 2010. 1

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

[5] R. Garg, A. Roussos, and L. Agapito. A variational approach to video registration with subspace constraints. *IJCV*, 104(3):286–314, 2013. 1

[6] C. Russell, J. Fayad, and L. Agapito. Energy based multiple model fitting for non-rigid structure from motion. In *CVPR*, 2011. 1

[7] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, 2006. 1

[8] J. Taylor, A. D. Jepson, and K. N. Kutulakos. Non-rigid structure from locally-rigid motion. In *CVPR*, 2010. 1