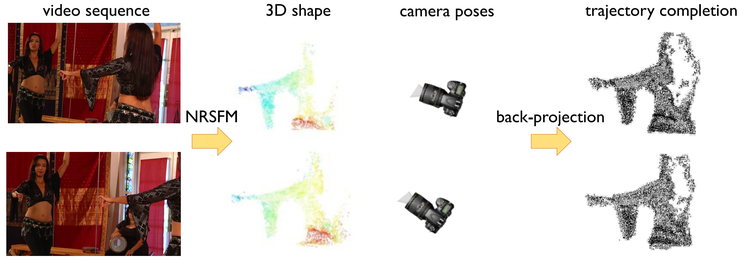
Grouping-Based Low-Rank Trajectory Completion and 3D Reconstruction

Katerina Fragkiadaki Marta Salas Pablo Arbelaez Jitendra Malik

Berkeley UNIVERSITY OF CALIFORNIA

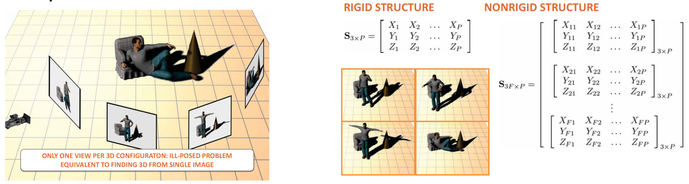
Overview

Given a monocular video, we segment the moving objects and reconstruct the 3D shape and camera viewpoints for each using low-rank shape priors. Back-projecting the 3D shapes under the recovered camera viewpoints provides *complete* point trajectories through object occlusions, rotations, image borders. Our contribution is applying NRSFM in realistic videos, using incomplete trajectories and real object segments, in contrast to most previous works, that operate in lab conditions, using full-length trajectories and pre-segmented objects.



Non-rigid Structure-from-Motion

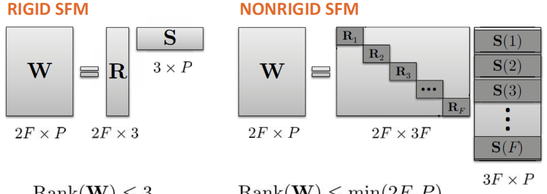
Given a video sequence of a non-rigid object, we want to extract 3D shape and camera poses.



Camera projection equation under a scaled orthographic camera:

$$\underset{2\times 1}{\mathbf{x}_{ij}} = \underset{2\times 3}{\mathbf{R}_i}\ \underset{3\times 1}{\mathbf{X}_j} + \underset{2\times 1}{\mathbf{t}_i}$$

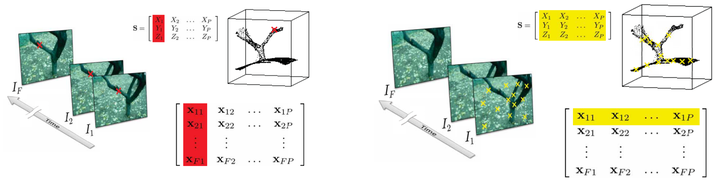
Eliminate the translation by fixing the world coordinate system at the object's centroid. Concatenate projection equations across all objects points and frames:



$\text{Rank}(\mathbf{W}) \leq 3$ $\quad$ $\text{Rank}(\mathbf{W}) \leq \min(2F, P)$

2D trajectory matrix

It contains the ``trackable'' image projections of the points of the object in all frames.



$\mathbf{W}$ is highly incomplete: only part of the object is visible at each frame. We obtain the trajectory matrix for each moving object using multi-scale trajectory spectral clustering. Trajectories terminate under low texturedness and under occlusions or self-occlusions of the object, or at image borders.

Formulation

NRSfM(K):

$$\min_{\mathbf{W},\mathbf{R},\mathbf{S}} \quad \|\mathbf{H}\odot(\mathbf{W}-\tilde{\mathbf{W}})\|_F^2 + \|\mathbf{W}-\mathbf{R}\cdot\mathbf{S}\|_F^2 + \mathbf{1}_{K>1}\cdot\mu\|\mathbf{S}^{\mathrm{v}}\|_*$$
$$\text{subject to} \quad \text{Rank}(\mathbf{W}) \leq 3K,\ \exists\alpha_t, \text{s.t. } \mathrm{R}^t(\mathrm{R}^t)^T = \alpha_t I_{2\times 2},\ t = 1\cdots F$$

$\tilde{\mathbf{W}}$: incomplete trajectory matrix
$\mathbf{H}$:indicator of present entries
$\|\cdot\|_*$:nuclear matrix norm

$$\mathbf{S}^{\mathrm{v}}_{F\times 3P} = \begin{bmatrix} X_1^1 & Y_1^1 & Z_1^1 & \cdots & X_P^1 & Y_P^1 & Z_P^1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ X_1^F & Y_1^F & Z_1^F & \cdots & X_P^F & Y_P^F & Z_P^F \end{bmatrix}$$

We solve it in three steps.

Step I: Given incomplete trajectory matrix $\tilde{\mathbf{W}}$ we compute complete trajectory matrix via bilinear factorization, imposing rank bound 3K:

$$\min_{\mathbf{W},\mathbf{U}_{2F\times 3K},\mathbf{V}_{P\times 3K}} \quad \|\mathbf{H}\odot(\mathbf{W}-\tilde{\mathbf{W}})\|_F^2 + \frac{\lambda}{2}(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2)$$
$$\text{subject to} \quad \mathbf{W} = \mathbf{U}\mathbf{V}^T.$$

We minimize it using the method of Augmented Lagrange Multipliers.

Step II: Given complete trajectory matrix $\mathbf{W}$ we recover camera rotations using a Euclidean upgrade similar to Tomasi and Kanade 1991:

$$\min_{\mathbf{R},\mathbf{S}} \quad \|\mathbf{W}-\mathbf{R}\cdot\mathbf{S}\|_F^2$$
$$\text{subject to} \quad \exists\alpha_t, \text{ s.t. } \mathrm{R}^t(\mathrm{R}^t)^T = \alpha_t I_{2\times 2},\ t = 1\cdots F.$$

We compute SVD of W truncated at rank 3:

$$\mathbf{W} = \mathrm{U}\mathrm{D}\mathrm{V}^{\mathrm{T}} = (\mathrm{U}\mathrm{D}^{1/2})(\mathrm{D}^{1/2}\mathrm{V}^{\mathrm{T}}) = \hat{\mathbf{R}}\cdot\hat{\mathbf{S}}$$

and a corrective transform G so that orthonormality constraints hold:

$$\hat{\mathbf{R}}_{2t-1}\mathrm{G}\mathrm{G}^T\hat{\mathbf{R}}_{2t}^T = 0, \quad t = 1\cdots F$$
$$\hat{\mathbf{R}}_{2t-1}\mathrm{G}\mathrm{G}^T\hat{\mathbf{R}}_{2t-1}^T = \hat{\mathbf{R}}_{2t}\mathrm{G}\mathrm{G}^T\hat{\mathbf{R}}_{2t}^T, \quad t = 1\cdots F$$

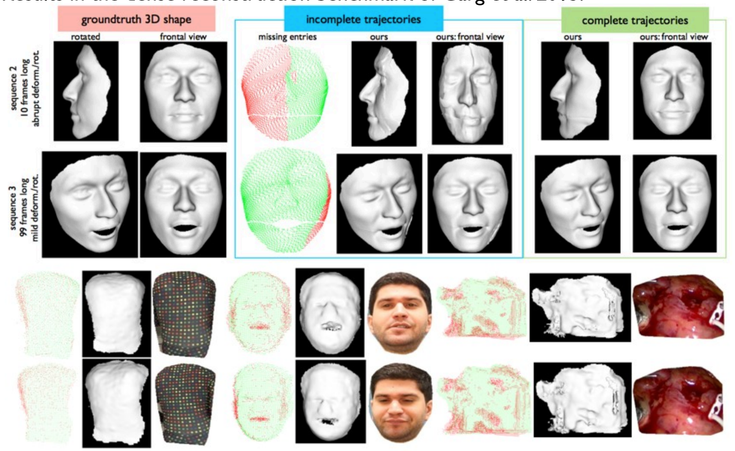
Step III: Given camera rotations $\mathbf{R}$ we minimize nuclear norm regularized reconstruction error:

$$\min_{\mathbf{S}} \quad \tfrac{1}{2}\|\mathbf{H}\odot(\tilde{\mathbf{W}}-\mathbf{R}\cdot\mathbf{S})\|_F^2 + \mu\|\mathbf{S}^{\mathrm{v}}\|_*$$
$$\text{subject to} \quad \mathbf{S}^{\mathrm{v}} = [P_X \quad P_Y \quad P_Z]\,(I_3\otimes\mathbf{S}).$$

We solve it adapting the accelerated proximal gradient method of Toh and Yun 2010.

Experiments

Results in the dense reconstruction benchmark of Garg et al. 2013.



Dense reconstructions in VSB100 and Moseg video segmentation benchmarks.