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.

Experiments
Given camera rotations $R$ we minimize nuclear norm regularized reconstruction error:

$$\min_{W, R} \frac{1}{2} \|W - R \cdot S\|^2_F + \frac{1}{2} \|R\|^2_F + \lambda \|S\|_n,$$

subject to $R_{t+1} \cdot G_{t+1} \cdot \tilde{R}_{t+1} = 0, \quad t = 1 \ldots F$.

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

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:

$$x_{ij} = R_{ij} X_i + t_i,$$

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

$$W = R \cdot S,$$

where $W$ are incomplete trajectory matrix, $S$ are complete trajectory matrix, $R$ are camera rotations.

Formulation
We solve it in three steps.

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

$$\min_{W, U_{2F \times 3K}, V_{P \times 3K}} \frac{1}{2} \|W - W - \tilde{W}\|^2_F + \frac{1}{2} (\|U\|^2_F + \|V\|^2_F),$$

subject to $W = UV^T$.

We minimize it using the method of Augmented Lagrange Multipliers.

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

Dense reconstructions in VSB100 and Moseg video segmentation benchmarks.