Discrete Laplace Operator Estimation for Dynamic 3D Reconstruction

Xiangyu Xu, Enrique Dunn
Stevens Institute of Technology

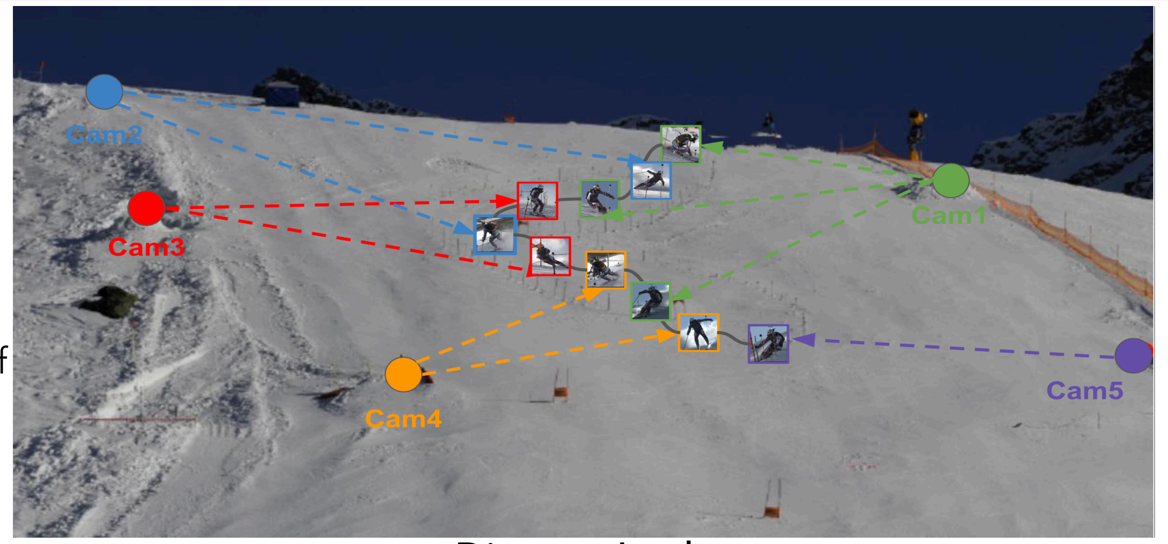
STEVENS INSTITUTE of TECHNOLOGY — THE INNOVATION UNIVERSITY — 1870

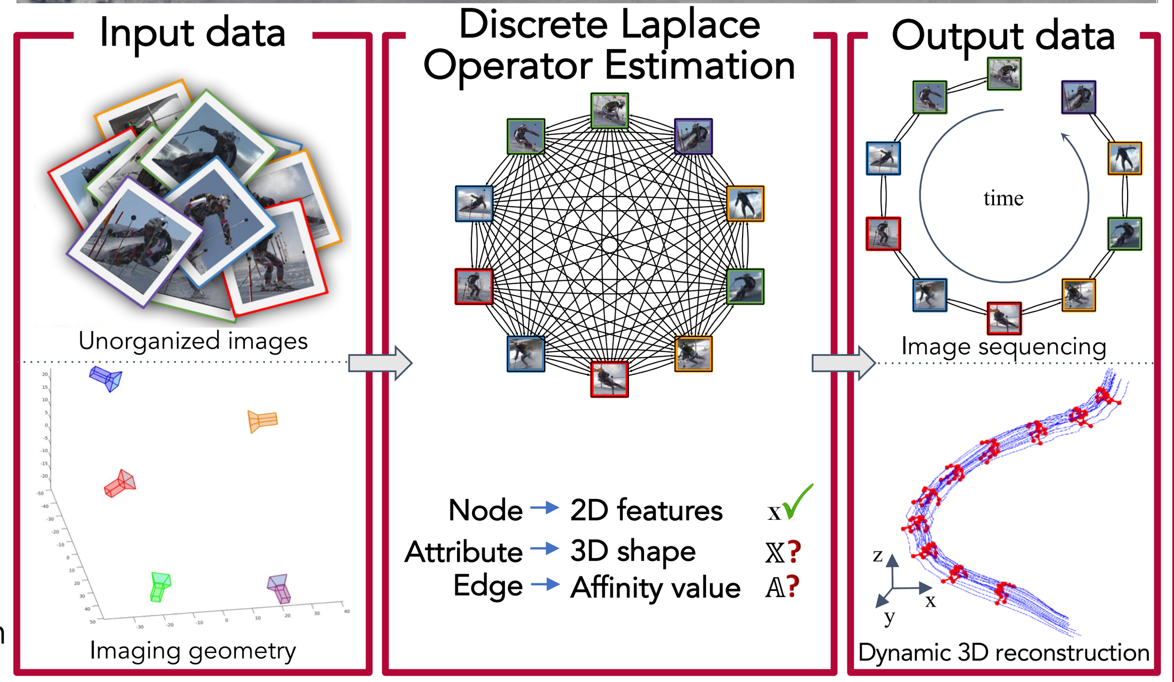
ICCV 2019 Seoul, Korea

Problem Statement

Goal

We address the problem of reconstructing the temporally evolving 3D geometry of set of points given a set unsynchronized 2D observations with unknown ordering and arbitrary temporal distribution. Our problem, which straddles both trajectory triangulation and image sequencing, naturally arises in the context of uncoordinated distributed capture of an event (e.g. crowd-sourced images or video)





What is known?

- 2D feature tracks
- Camera parameters and poses
- Local image sequencing in video capture

What's the output?

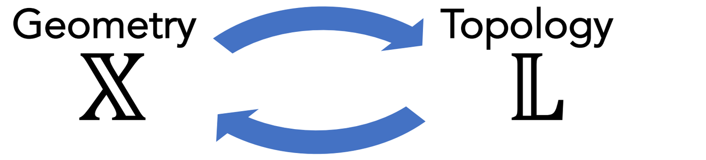
- Global image sequencing
- Dynamic 3D structure

What's the challenges?

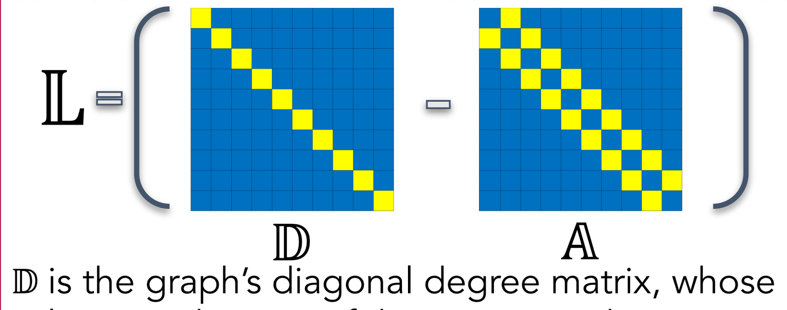
- Non-rigid object motion
- No object model
- Unsynchronized image capture
- Unknown global image sequencing
- Arbitrary temporal sampling density and distribution

Motivation

Jointly estimate 3D geometry & graph's discrete Laplace operator



The graph's Laplacian defines the topology in terms of the affinities between our 3D estimates



$\mathbb{D}$ is the graph's diagonal degree matrix, whose values are the sum of the corresponding row in $\mathbb{A}$

Laplacian decomposition

$$\mathbb{L} = \mathbb{D}(\mathbb{I} - \mathbb{W})$$
$$tr(\mathbb{D}) = 1, \mathbb{D} \geq 0$$
$$\mathbb{W}\mathbf{1}_{N\times 1} = \mathbf{1}_{N\times 1}, \mathbb{W} \geq \mathbf{0}$$

$\mathbb{W}$ denotes the relative affinities in a **local neighborhood**, and $\mathbb{D}$ denotes the **density and flatness** of the neighborhood

- Enforce **sparsity** of $\mathbb{W}$ through least squares minimization by constraining each row of $\mathbb{W}$ sum to 1 and positive.
- Each row of of $\mathbb{W}$ is independent and able **to optimize in parallel** using an efficient Active-set method (Chen, 2014).
- Result in solving a **tri-convex optimization** problem over $\mathbb{W}$, $\mathbb{D}$, and $\mathbb{X}$.

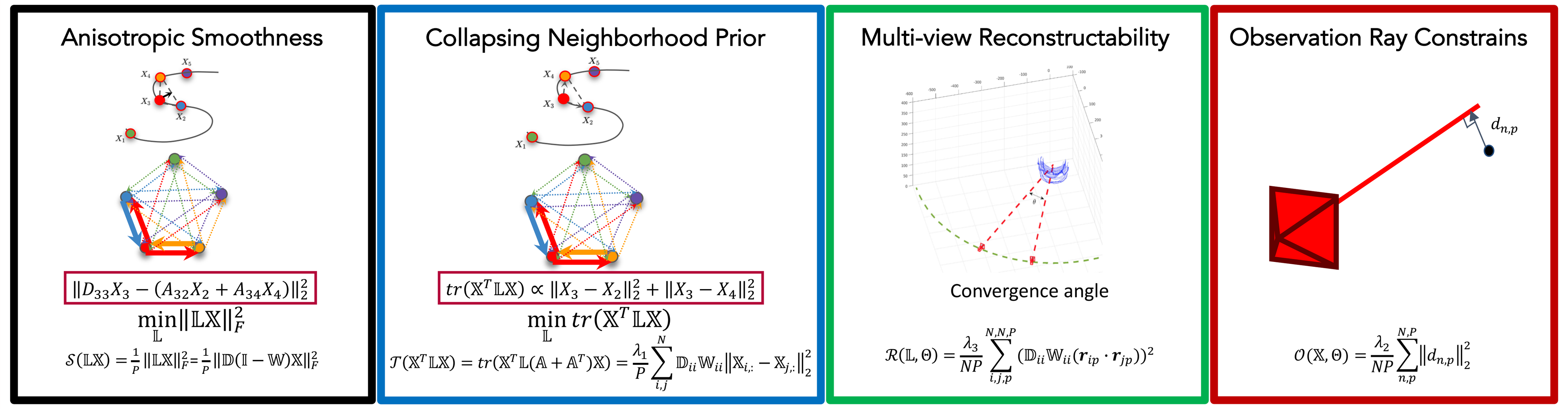
Contribution

- A graph-theoretic formulation of the dynamic reconstruction problem, where 2D observations are mapped to nodes, 3D geometry are node attributes, and spatiotemporal affinities correspond to graph edges.
- The definition and enforcement of spatio-temporal priors, (e.g. anisotropic smoothness, topological compactness/sparsity, and multi-view reconstructability) in terms of the discrete Laplace operator.
- Integration of available per-stream (e.g. intra-video) sequencing info into global ordering priors enforced in terms of the Laplacian spectral signature.

Our approach for unorganized images

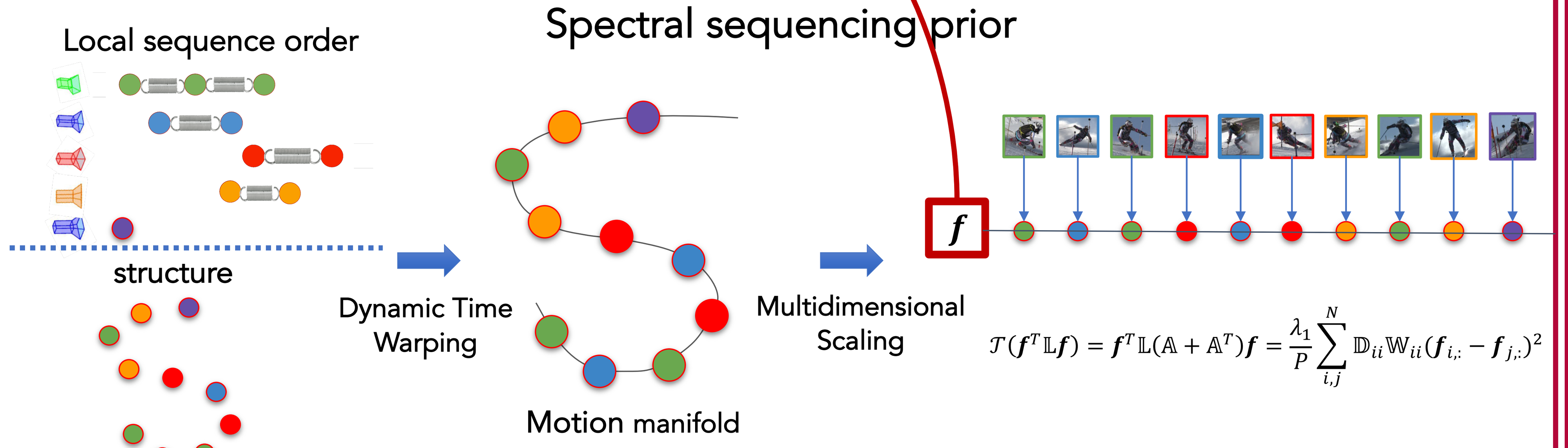
$$\min_{\mathbb{X},\mathbb{W},\mathbb{D}} \mathcal{S}(\mathbb{L}\mathbb{X}) + \mathcal{T}(\mathbb{X}^T\mathbb{L}\mathbb{X}) + \mathcal{R}(\mathbb{L},\Theta) + \mathcal{O}(\mathbb{X},\Theta)$$
$$subject\ to\ \ \mathbb{L} = \mathbb{D}(\mathbb{I}-\mathbb{W}), tr(\mathbb{D}) = 1, \mathbb{D}\geq 0, \mathbb{W}\mathbf{1}_{N\times 1}, \mathbb{W}\geq 0$$

($\Theta = \{\{x_{n,p}\},\{K_n\},\{M_n\}\}$ denotes the aggregation of all input 2D observations and their camera parameters)



Our approach for unsynchronized videos

$$\min_{\mathbb{X},\mathbb{W},\mathbb{D}} \mathcal{S}(\mathbb{L}\mathbb{X}) + \mathcal{T}(\boldsymbol{f}^T\mathbb{L}\boldsymbol{f}) + \mathcal{R}(\mathbb{L},\Theta) + \mathcal{O}(\mathbb{X},\Theta)$$
$$subject\ to\ \ \mathbb{L} = \mathbb{D}(\mathbb{I}-\mathbb{W}), tr(\mathbb{D}) = 1, \mathbb{D}\geq 0, \mathbb{W}\mathbf{1}_{N\times 1}, \mathbb{W}\geq 0$$



Reconstructability analysis

Assumption

1) $\mathbb{L}$ is fixed
2) $\mathbb{L}$ encodes ground truth temporal adjacency
3) Noise free 2D observations

Then, cost function becomes

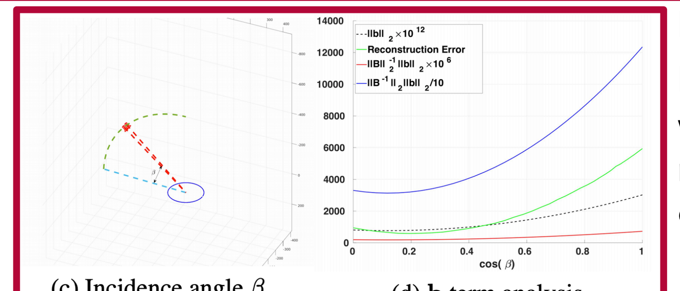
$$\min_{l} \mathcal{S}(\mathbb{L}\mathbb{X}) + \mathcal{T}(\mathbb{X}^T\mathbb{L}\mathbb{X})$$

where, $\mathbb{X}_{n,:} = \mathbb{X}^*_{n,:} + l_n \boldsymbol{r}_n$.and solved by setting the derivative over l to zero, yielding to

$$\mathbb{B}\boldsymbol{l} = \boldsymbol{b}$$
$$\mathbb{B} = (\mathbb{L}^T\mathbb{L} + \lambda_1\mathbb{L}(\mathbb{A}+\mathbb{A}^T)) \odot \begin{bmatrix} \boldsymbol{r}_1^T\boldsymbol{r}_1 & \cdots & \boldsymbol{r}_N^T\boldsymbol{r}_1 \\ \vdots & \ddots & \vdots \\ \boldsymbol{r}_1^T\boldsymbol{r}_N & \cdots & \boldsymbol{r}_N^T\boldsymbol{r}_N \end{bmatrix}$$
$$b_n = (\mathbb{L}_{:,n}^T\mathbb{L}\mathbb{X}^* + \lambda_1\mathbb{L}(\mathbb{A}+\mathbb{A}^T)_{n,:}\mathbb{X}^*)\boldsymbol{r}_n$$

Attain the lower and upper bound

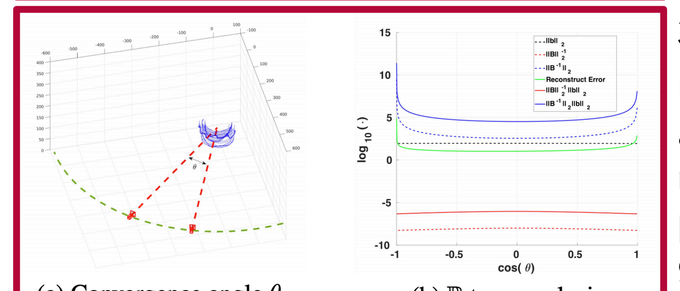
$$\|\mathbb{B}\|_2^{-1}\|\boldsymbol{b}\|_2 \leq \|\boldsymbol{l}\|_2 \leq \|\mathbb{B}^{-1}\|_2\|\boldsymbol{b}\|_2$$



(c) Incidence angle β (d) b term analysis

Imaging geometry convergence

Reconstruction error is reduced when neighboring viewing rays near orthogonality. Also, the error bounds get tighter.



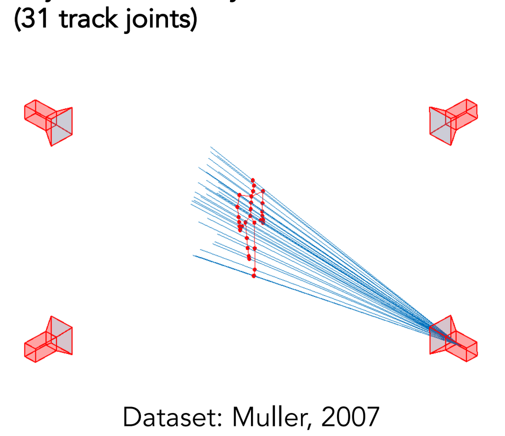
(a) Convergence angle θ (b) $\mathbb{B}$ term analysis

3D motion observability

More accurate reconstruction is attained for viewing directions near orthogonal to the motion plane. Also, the error bounds get tighter.

Experiments on synthetic imagery from motion capture

Project to four unsynchronized cameras (31 track joints)



Dataset: Muller, 2007

Baseline methods

- Trajectory basis (Park, 2015)
- High-pass filter (Valmadre, 2012)
- Triangulation w/sequence

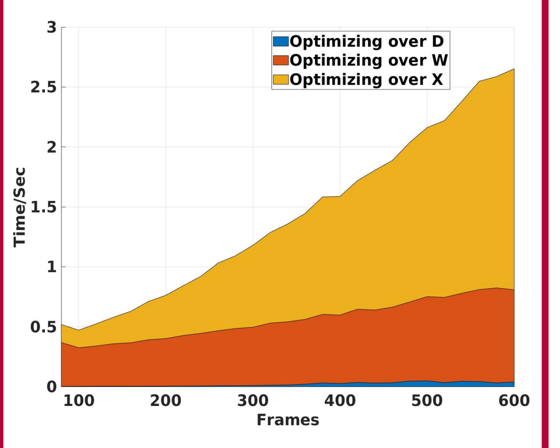
Require global sequencing

- Triangulation
- Self-expressive dictionary (Zheng, 2017)

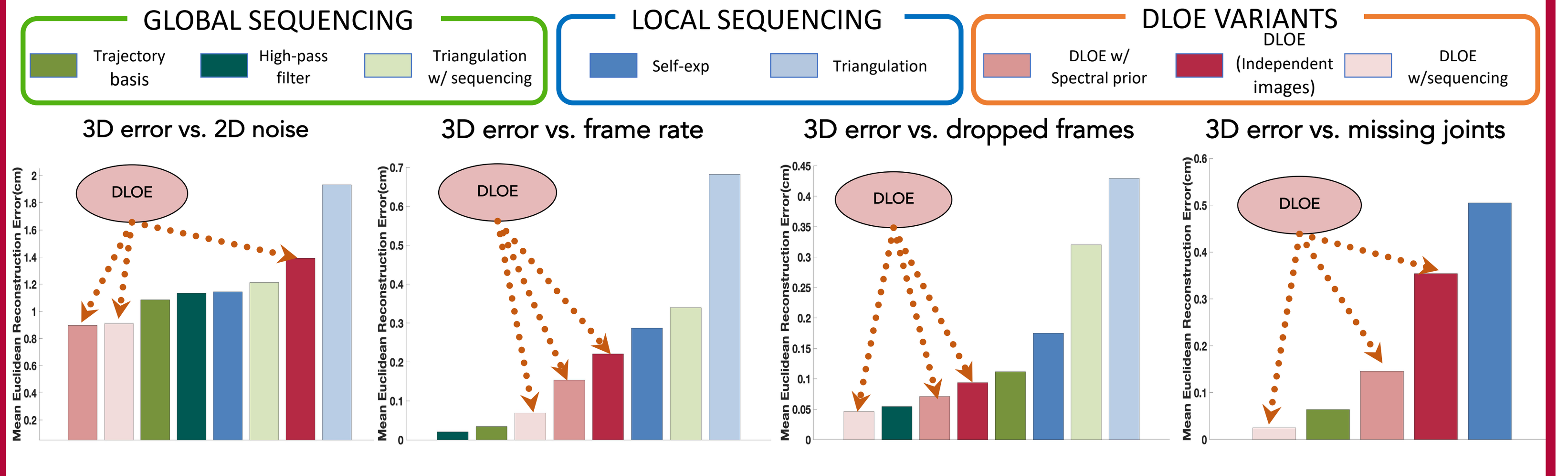
Require local sequencing

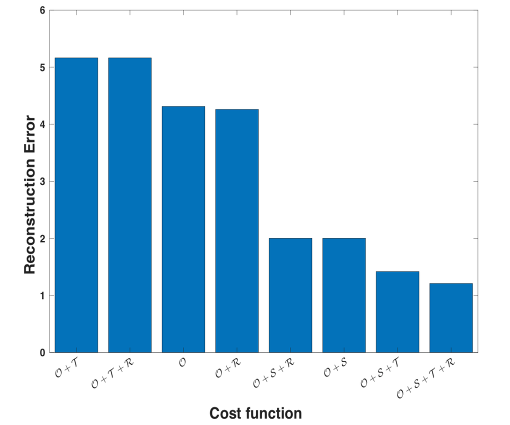
Discrete Laplace Operator Estimation (DLOE) variants

- DLOE w/ sequence ← Require global sequencing
- DLOE w/ spectral prior ← Require local sequencing
- DLOE (Independent images) ← No sequencing



Single iteration running time



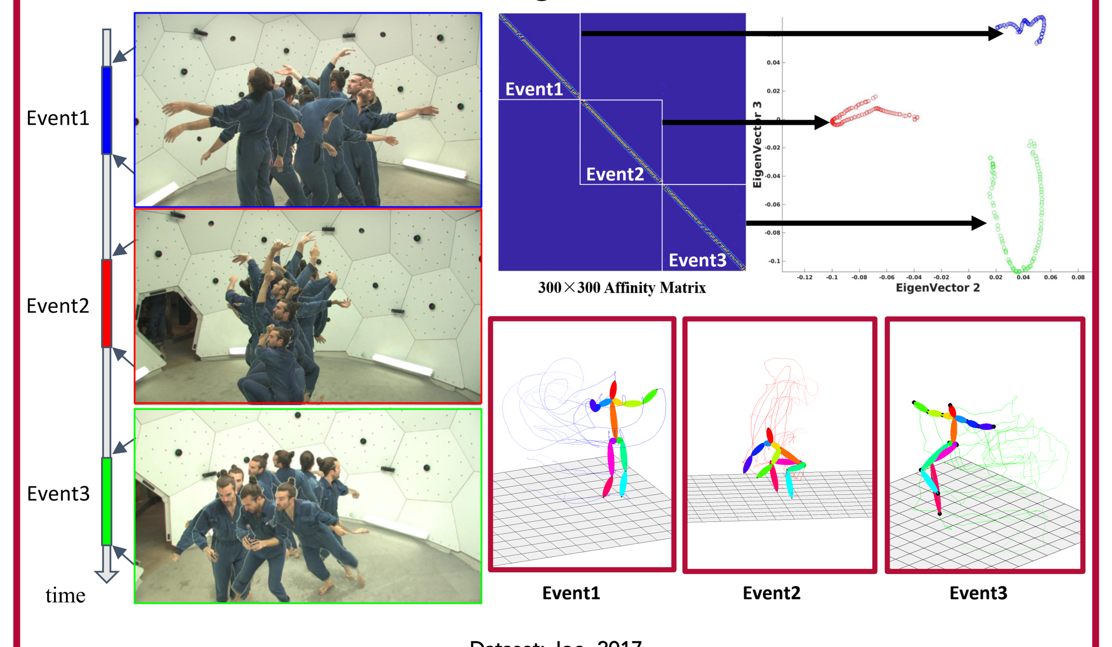


Ablation analysis

Experiments on multi-view imagery

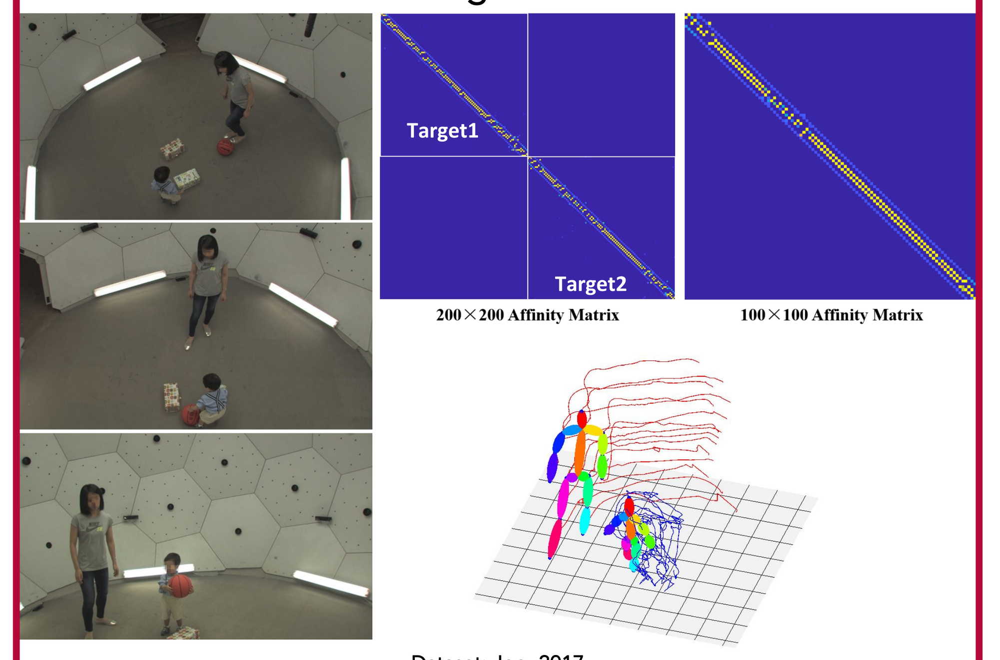
Dataset: Ballan, 2017 — Dataset: Park, 2010 — Dataset: Rhodin, 2018

Event segmentation



Dataset: Joo, 2017

Multi-Target Scenario



Dataset: Joo, 2017