Penn
UNIVERSITY of PENNSYLVANIA

Video Segmentation by Tracing Discontinuities in a Trajectory Embedding

Katerina Fragkiadaki, Geng Zhang and Jianbo Shi

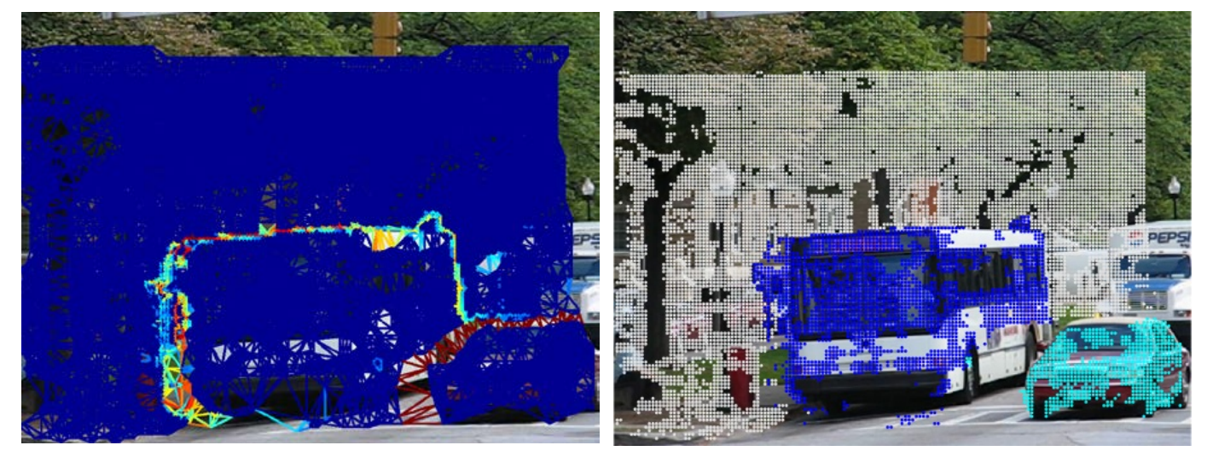
Video segmentation

Problem

Spectral clustering of point trajectories has shown good performance for video segmentation. Model selection, namely **selecting the number of clusters** and effectively discretizing the continuous spectral embedding, remains a challenging problem. Traditional algorithms that cluster trajectories in the spectral embedding often over over-segment or under-segment the objects. On the right we see that over-fragmentation of the background happens before the right segmentation pops out. So simply choosing the right number of clusters K is not enough!

Our contribution

We propose an embedding discontinuity detector that acts in the spectral trajectory embedding and measures motion discontinuities between spatially adjacent trajectories. Detected discontinuities are robust against varying the number of eigenvectors. As such, the proposed discontinuity-aware discretization recovers from artificial fragmentations by merging accordingly clusters with low discontinuities between them.



Embedding Discontinuity Detector

Trajectory spectral embedding.

Given a set of point trajectories, we compute pairwise affinities **A** reflecting motion similarities. Let X_l be the indicator of lth trajectory cluster , K the number of clusters and **D** the diagonal degree matrix of **A**, $\mathbf{D}_{i,i}=\sum_j \mathbf{A}_{ij}$. We maximize intra-cluster normalized affinities:

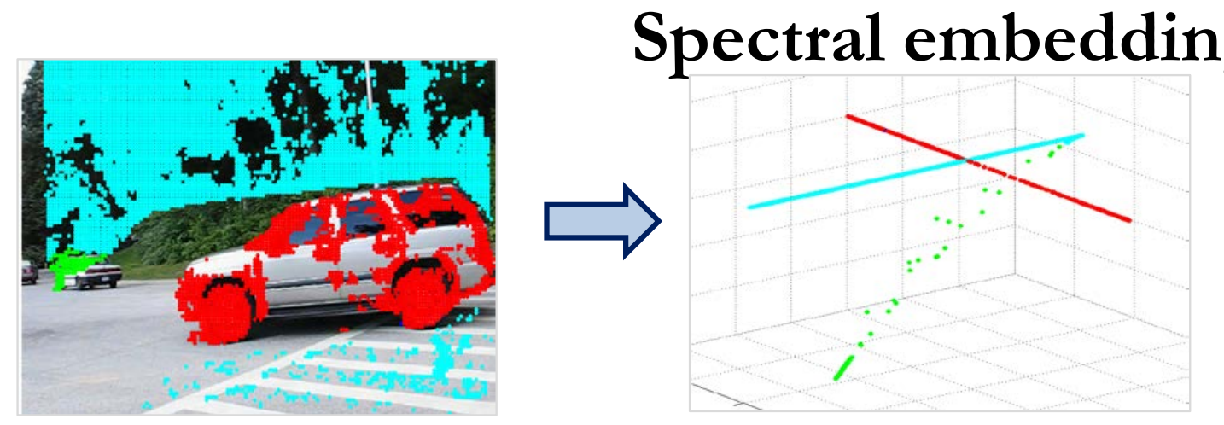
$$\max.\ \epsilon(X)=\frac{1}{K}\sum_{l=1}^{K}\frac{X_l^T\mathbf{A}X_l}{X_l^T\mathbf{D}X_l}$$
$$\text{subject to } X\in\{0,1\}^{N\times K}$$
$$X1_K=1_N$$

which can be simplified by substituting $Z=X(X^TDX)^{-\frac{1}{2}}$

$$\max.\ \epsilon(Z)=\frac{1}{K}\mathrm{tr}(Z^T\mathbf{A}Z)$$
$$\text{subject to } Z^T\mathbf{D}Z=I_K$$

We obtain a near global-optimum **continuous** solution by the top K eigenvectors (and eigenvalues)$(\mathbf{V},\mathbf{\Lambda})$ of the normalized affinity matrix $\mathbf{D}^{-1}\mathbf{A}$.

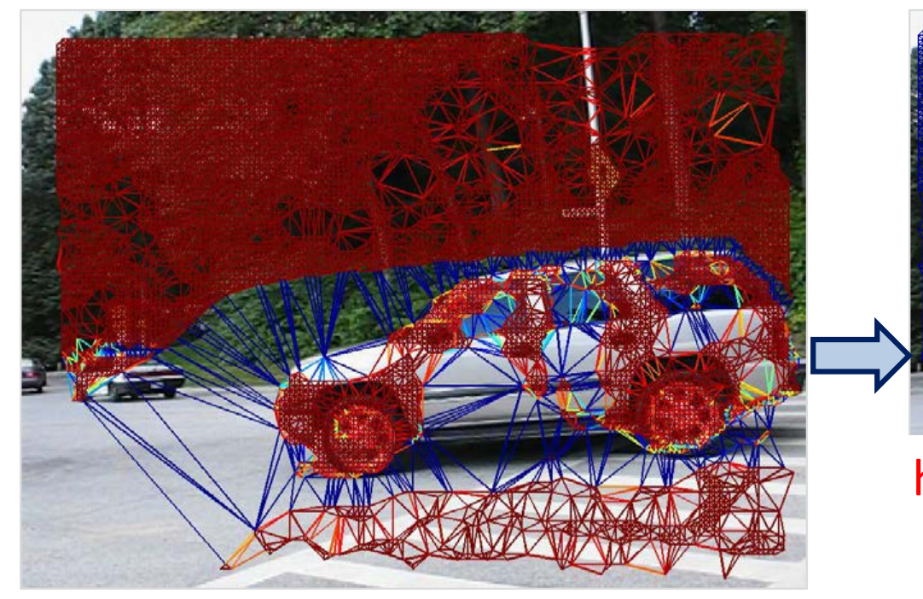
Spectral embedding



Notice that the optimum is not unique! Any rotation R of eigenvectors **V**, provides a global optimum with the same value. So the optimum is a subspace spanned by the top K eigenvectors of $\mathbf{D}^{-1}\mathbf{A}$ through orthonormal transformations:

$$\{Z^*R, RR^T=I_K, \mathbf{D}^{-1}\mathbf{A}Z^*=Z^*\Lambda^*\}$$

Motion affinities A



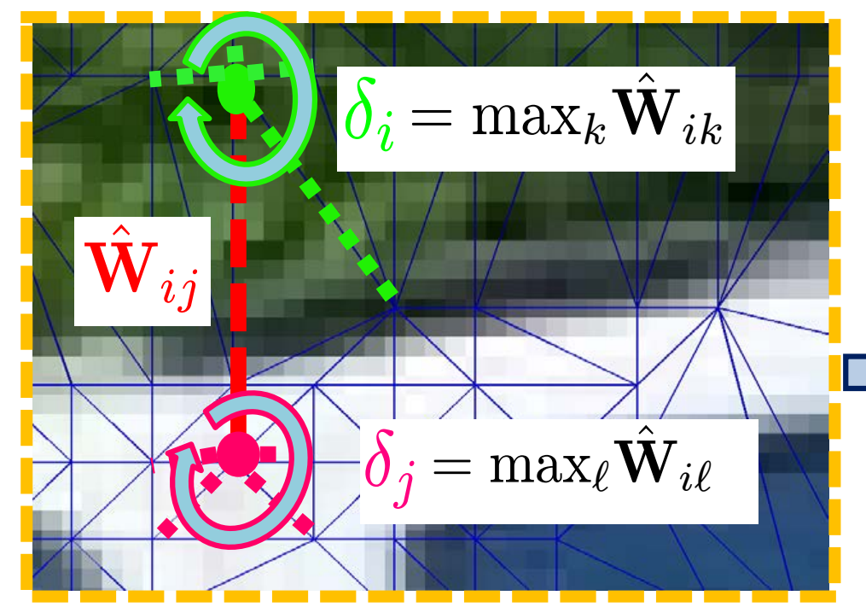
We model affinities on the edges of a Delaunay Graph **built from trajectory points in each frame to simultaneously represent affinity and spatial proximity in the image domain.**

Embedding affinities $\hat{\mathbf{W}}$

Embedding affinities:

$$\hat{\mathbf{W}}=\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

Density normalization

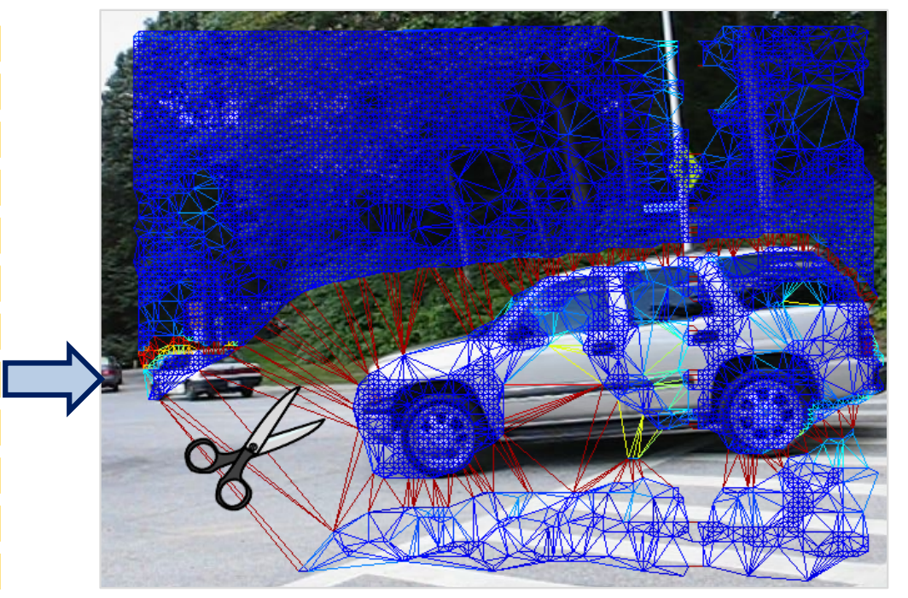


Trajectory **density**:

$$\delta_i=\max_{j\in\mathcal{N}_{\mathrm{Del}}^i}\hat{\mathbf{W}}_{ij}$$

$\mathcal{N}_{\mathrm{Del}}^i$: The Delaunay neighbors of trajectory i.

Embedding discontinuities

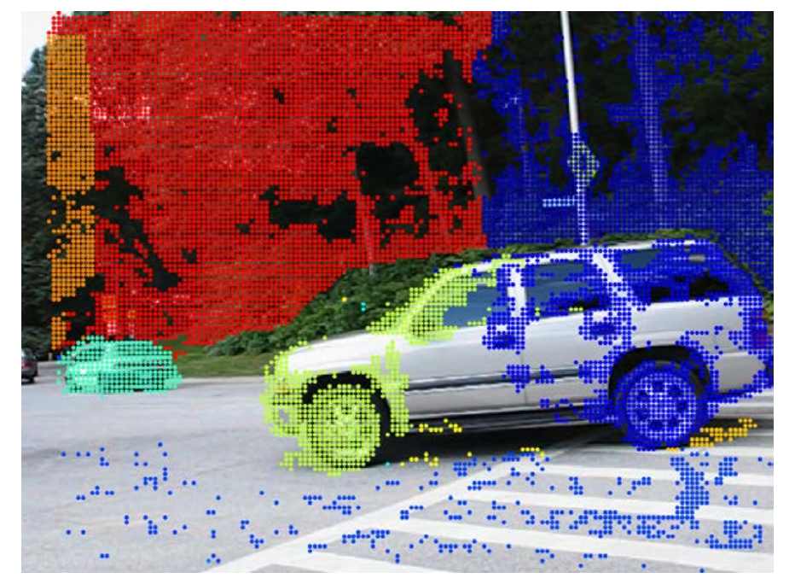


Trajectory embedding **discontinuity**:

$$\mathbf{d}_{ij}=1-\hat{\mathbf{W}}_{ij}/\max(\delta_i,\delta_j)$$

They are strong on Delaunay edges spanning cross-object boundaries.

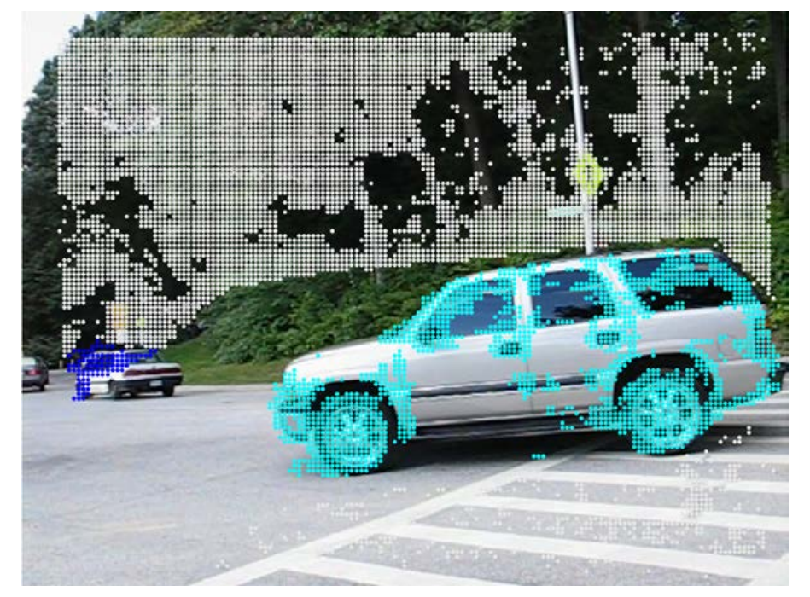
Eigenvector rotation



Discretizing with non maxima suppression the eigenvectors that have been rotated to be close to the continuous optimum:

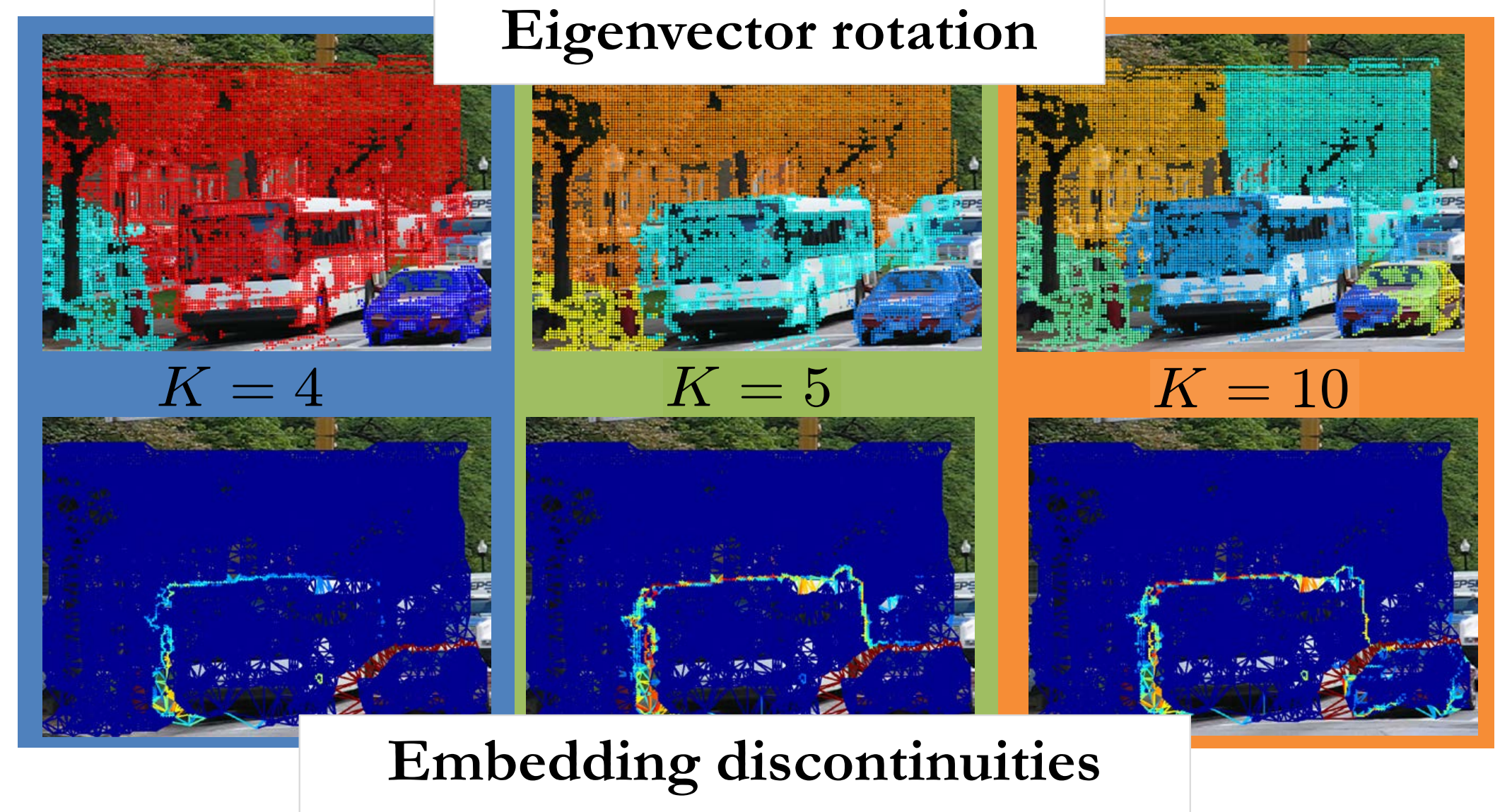
$$\mathbf{V}_R=\mathbf{V}R^*$$

Discontinuity-aware discretization



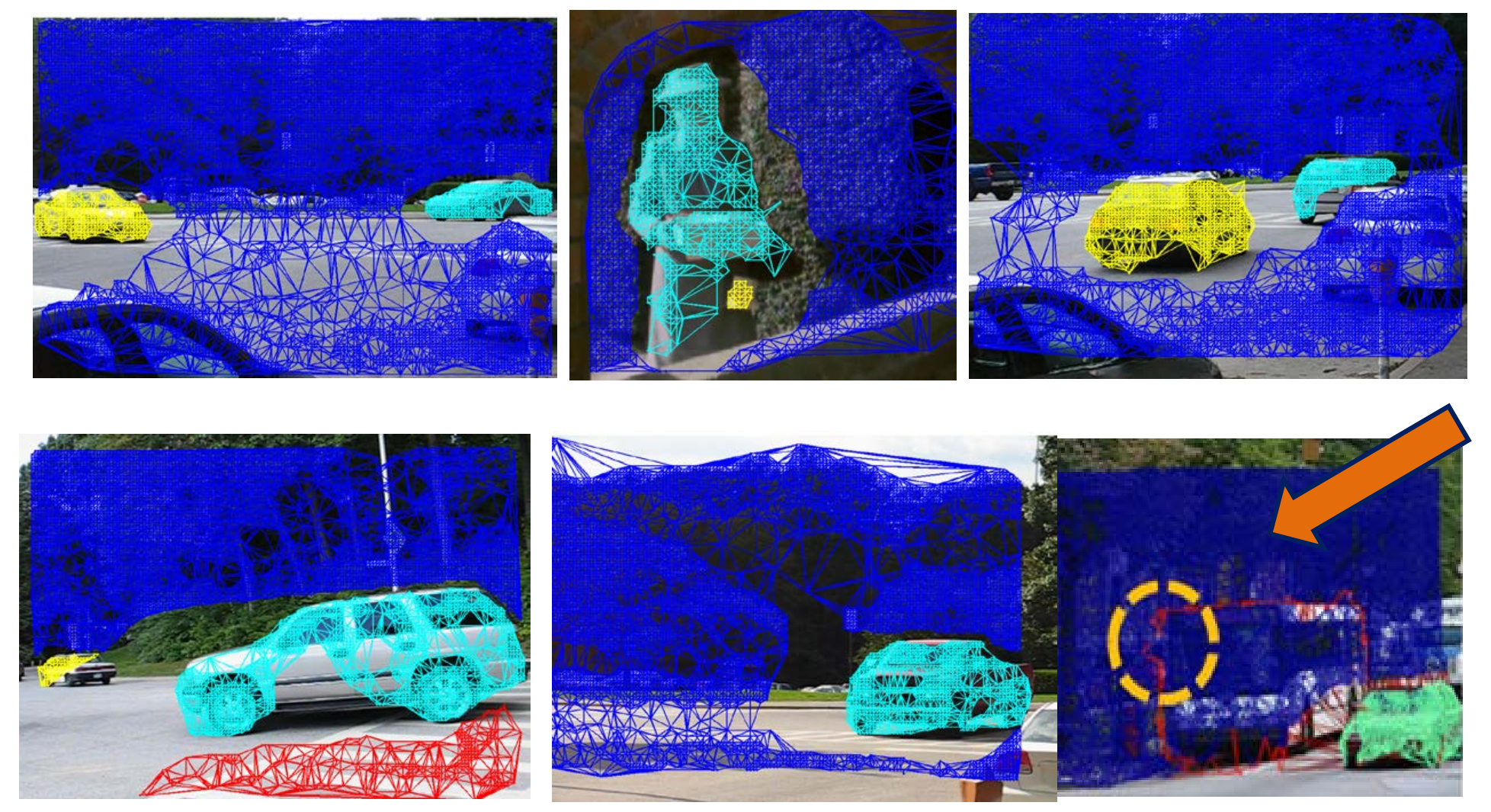
We **merge clusters whose inter-cluster boundaries have low discontinuities**. We recover from over-fragmentations while being robust to local embedding instabilities.

Discontinuities are robust to number of eigenvectors


Eigenvector rotation

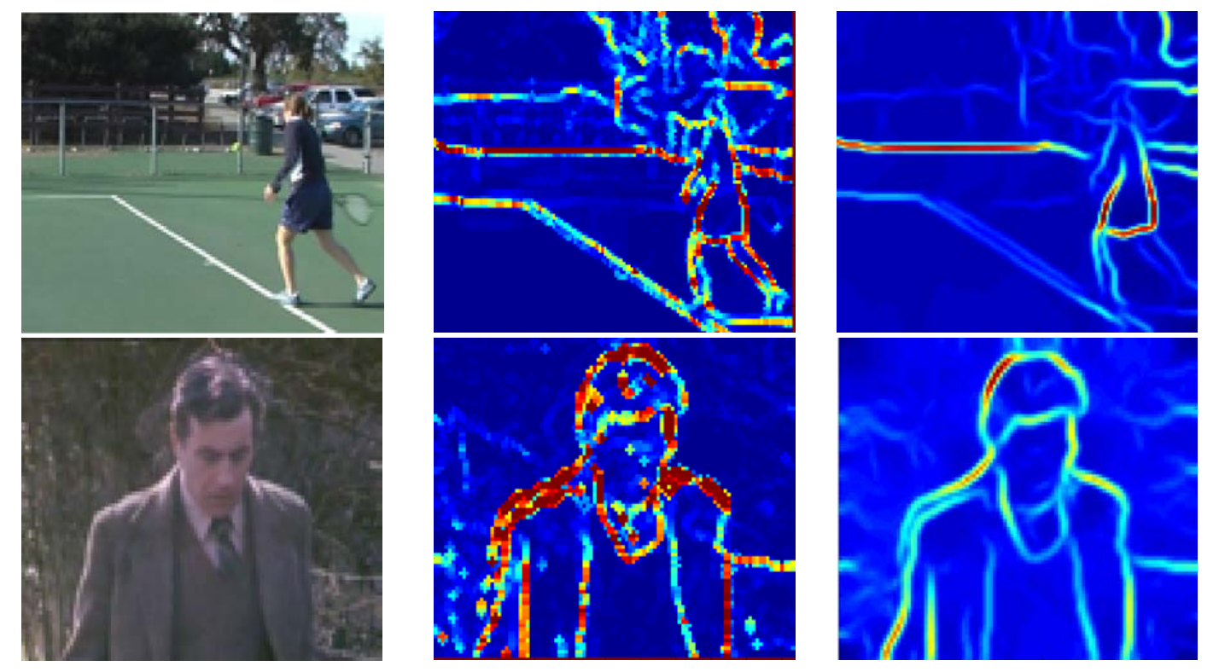
Embedding discontinuities

Why not simply threshold discontinuities?



Drifting trajectories create instabilities in the embedding that confuse the discontinuities locally

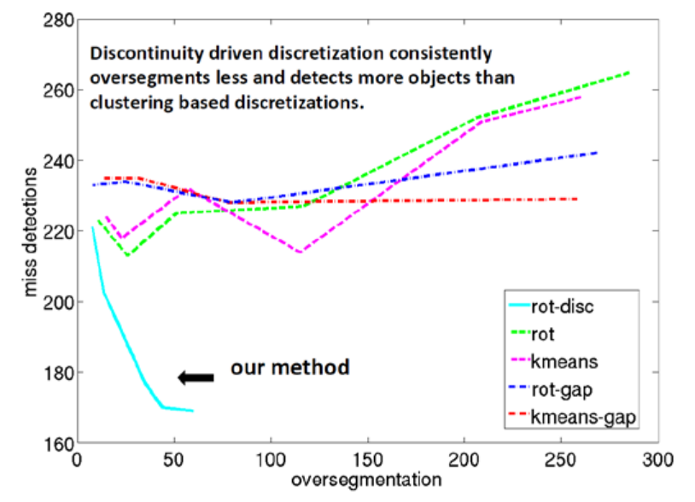
Pixel discontinuities / Spectral Pb



Results

Discretizing the trajectory embedding

Comparison of discontinuity-driven discretization against eigenvector rotation and k-means. The oversegmentation error for same miss detection error is much lower for our method that recovers from artificial fragmentations. The curve has been obtained by varying the number of eigenvectors.



Figment	density	clustering error	region clustering error	over-segmentation	recall	leakage	tracking time
our method	**7.05%**	7.90%	**18.47%**	**1.5**	**33.28%**	19.55%	**82.29%**
our method w/o FG	4.90%	17.49%	41.06%	3.21	19.19%	44.96%	48.49%
Fragkiadaki et al 2010	5.21%	**4.73%**	20.32%	1.57	31.07%	**16.52%**	75.13%

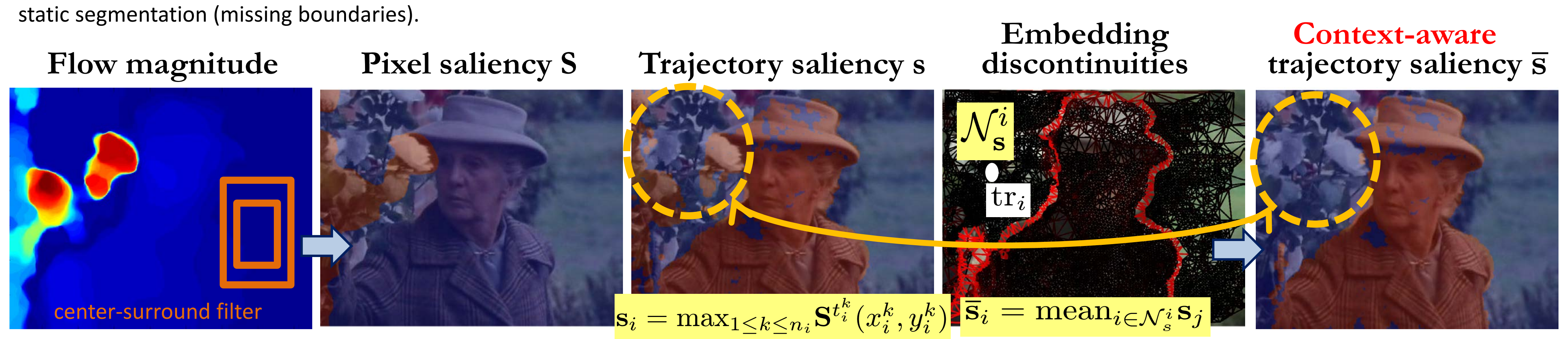
Moseg	density	clustering error	region clustering error	over-segmentation	extracted objects
our method (trajectory clustering)	3.07%	**2.29%**	20.93%	0.29	**29**
our method w/o FG (traject. clustering)	3.15%	2.55%	**20.63%**	0.48	28
our method (pixel segmentation)	**93.72%**	3.95%	26.14%	**0.25**	26
Fragkiadaki et al. 2010	3.22%	3.76%	22.06%	1.15	25
Brox et al. 2010	3.32%	3.43%	27.06%	0.4	26

Figment

Moseg

Context-aware trajectory motion saliency

Center-surround motion saliency is a popular way for obtaining figure-ground information in videos. It is robust to motion of virtual contours and to mistakes of static segmentation (missing boundaries).



Center-surround filters often **cannot select the right scale of objects**. They utilize no segmentation (grouping) information.

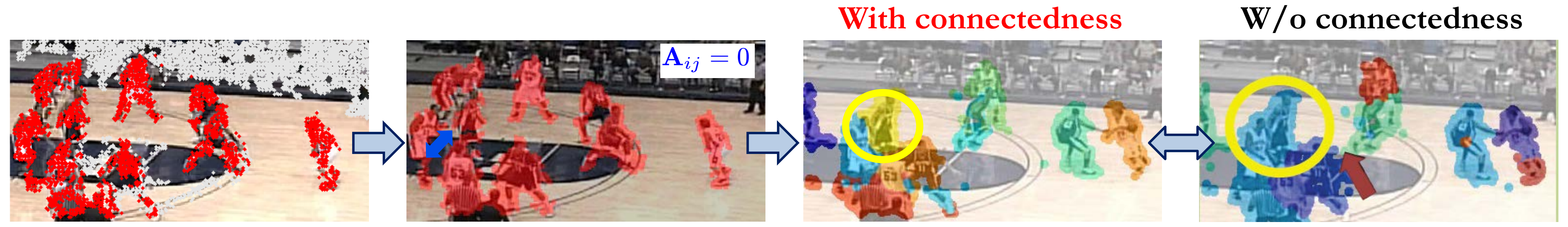
Trajectory saliency assigns objects as salient even if they don't move in the current frame.

Trajectory saliency is averaged across embedding neighborhoods.

Context-aware trajectory saliency is spatially smooth and recovers from mistakes of the center-surround saliency filter.

Object connectedness constraints

Motion alone is often insufficient for segmenting articulated bodies. Figure-ground information helps distinguishing object articulation versus object separation. Trajectories that belong to disconnected components of the foreground, violate object connectedness and should belong to separate objects.



We obtain a video figure-ground segmentation by thresholding context-aware trajectory saliencies .

We cancel affinities between trajectories that belong to different connected components of the foreground.

Figure-ground information allows correct segmentation of articulated bodies. Motion discontinuities exist both across objects as well as across articulated body parts. However, violation of object connectedness happens only between trajectories that belong to different objects.

Trajectory Clustering to Pixel segmentation

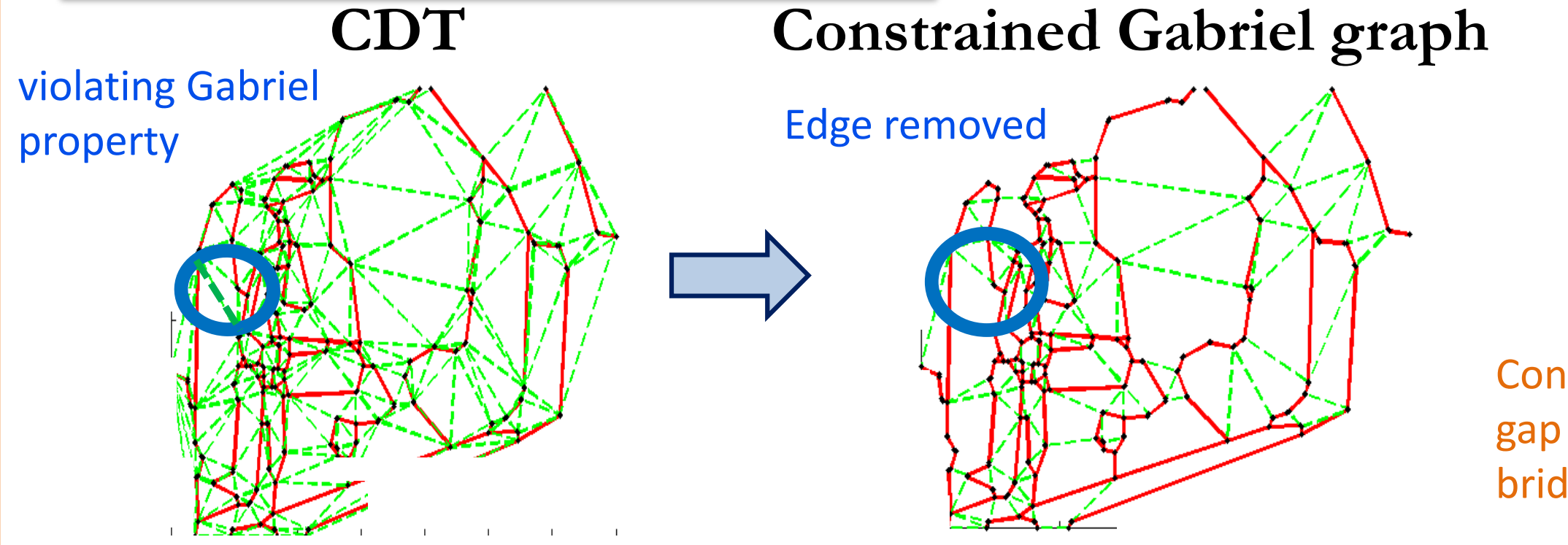
Problems

1. Optical flow trajectories **bleed** across segment boundaries.
2. **Untextured** image areas are sparsely populated with trajectories.
3. Faint static boundaries make segments / **superpixels leak**.

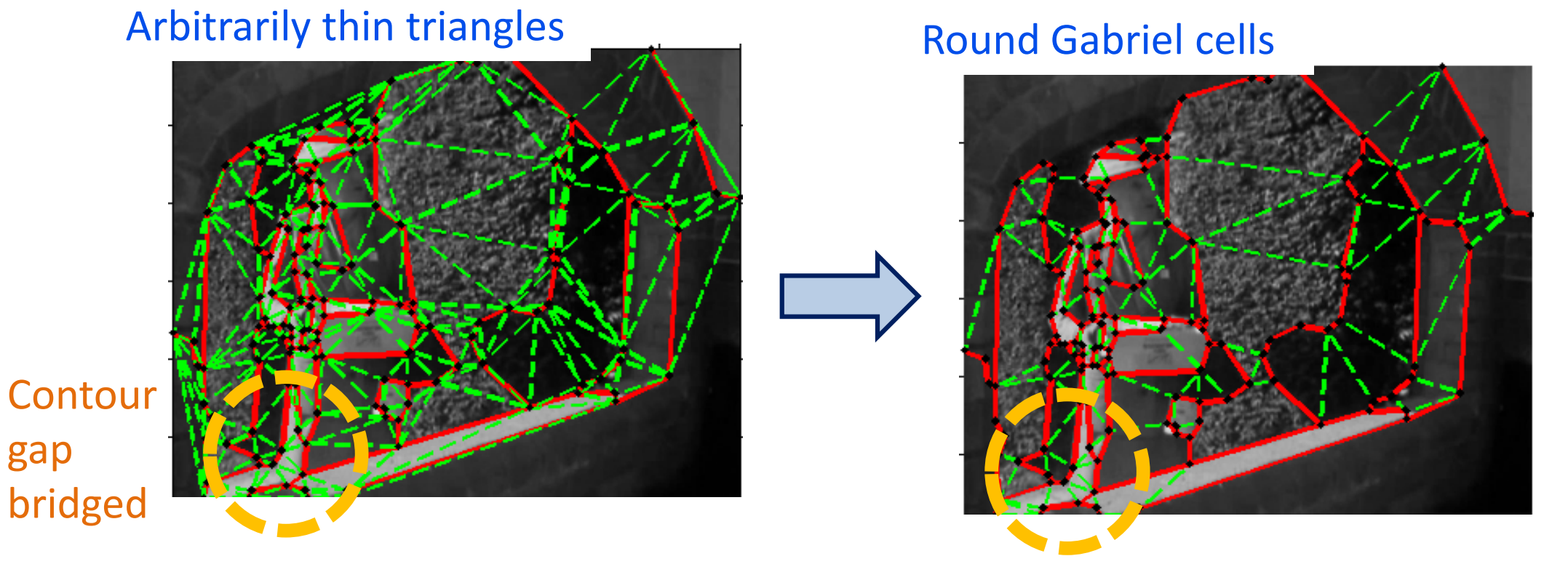
Our contribution

Constrained Gabriel graph as a **superpixel map .** Contour gaps due to missing boundaries are bridged according to **geometric reasoning** ,without resorting to tiny superpixels.

Constrained Gabriel Graphs



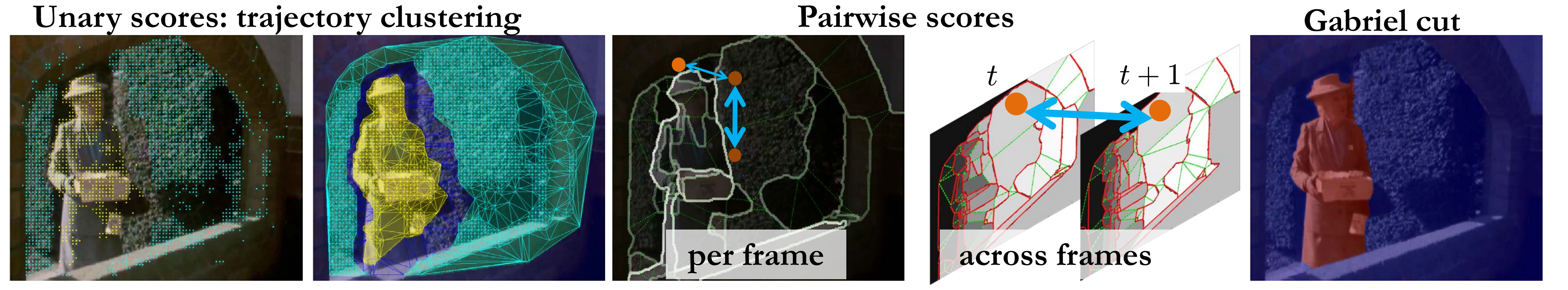
Two points are neighbors according to the **Gabriel property** if **their circumcirle does not contain any other points**.



A (constrained) Gabriel graph is obtained from the corresponding (constrained) Delaunay Triangulation by deleting Delaunay edges violating the Gabriel property.

Gabriel Cut

We compute video segmentation via graph cuts on Gabriel cells of all frames simultaneously.



Delaunay triangles whose *vertices have the same trajectory cluster label* indicate areas of reliable unary potentials. The rest (shown in blue), indicate unreliable unary potentials.

Pairwise potentials on Gabriel cells of the same frame depend on **Pb**. Pairwise potentials on Gabriel cells across neighboring frames depend on **trajectory sharing**.

Pixel segmentation result.