Two-Granularity Tracking: Mediating Trajectory and Detection Graphs for tracking under Persistent Occlusions

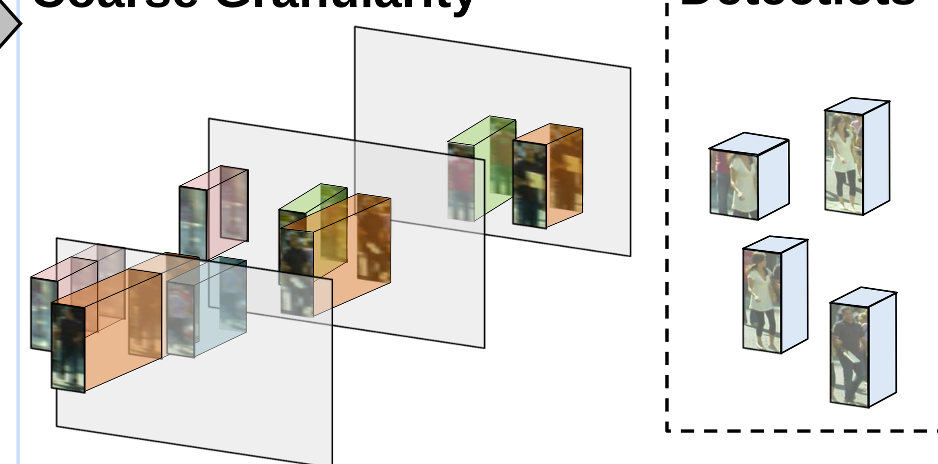
Katerina Fragkiadaki Weiyu Zhang Geng Zhang Jianbo Shi

University of Pennsylvania

Input Video:

Coarse Granularity

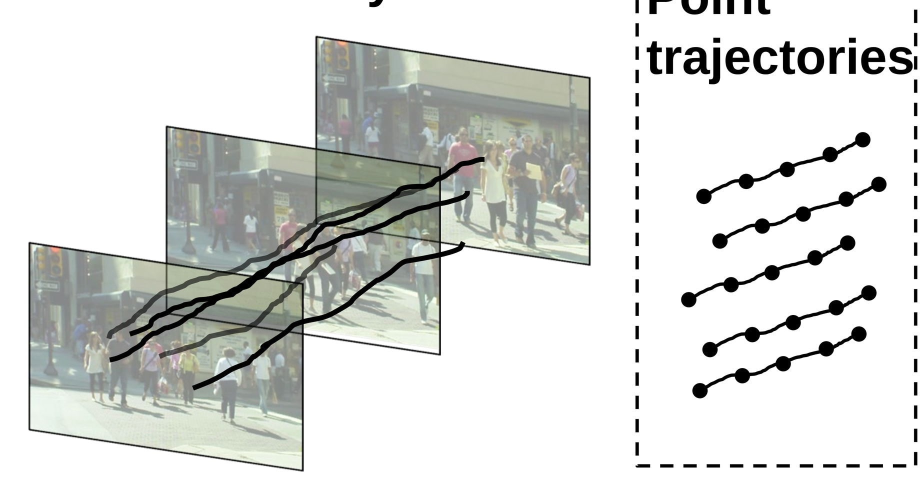
Detectlets



$\mathrm{dl}_p = (\{\mathrm{box}_p^t, t \in T_p\}, \mathbf{f}_p)$ $\mathbf{f}_p$: confidence score

Fine Granularity

Point trajectories



$\mathrm{tr}_i = \{(x_i^t, y_i^t), t \in T_i\}$

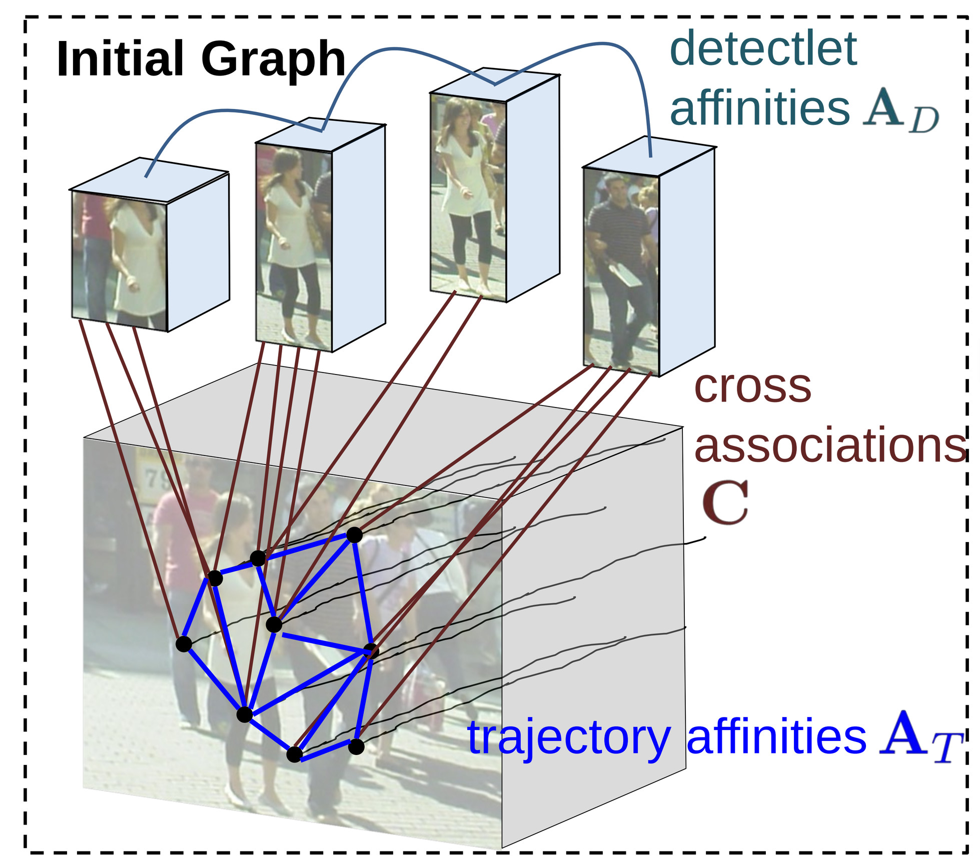
Coarse Granularity: Detectlets
Tracking by linking detections

- false positive /miss detection due to deformations /occlusions
- under fit /loose fit boxes

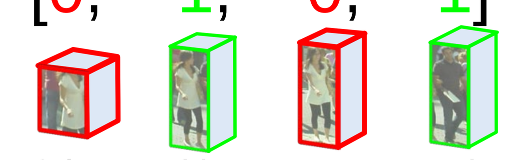
Fine Granularity: Point trajectories
Tracking by linking flow fields, segmenting by trajectory clustering

- leaking affinities across objects w/ similar motion/disparity

Initial Graph

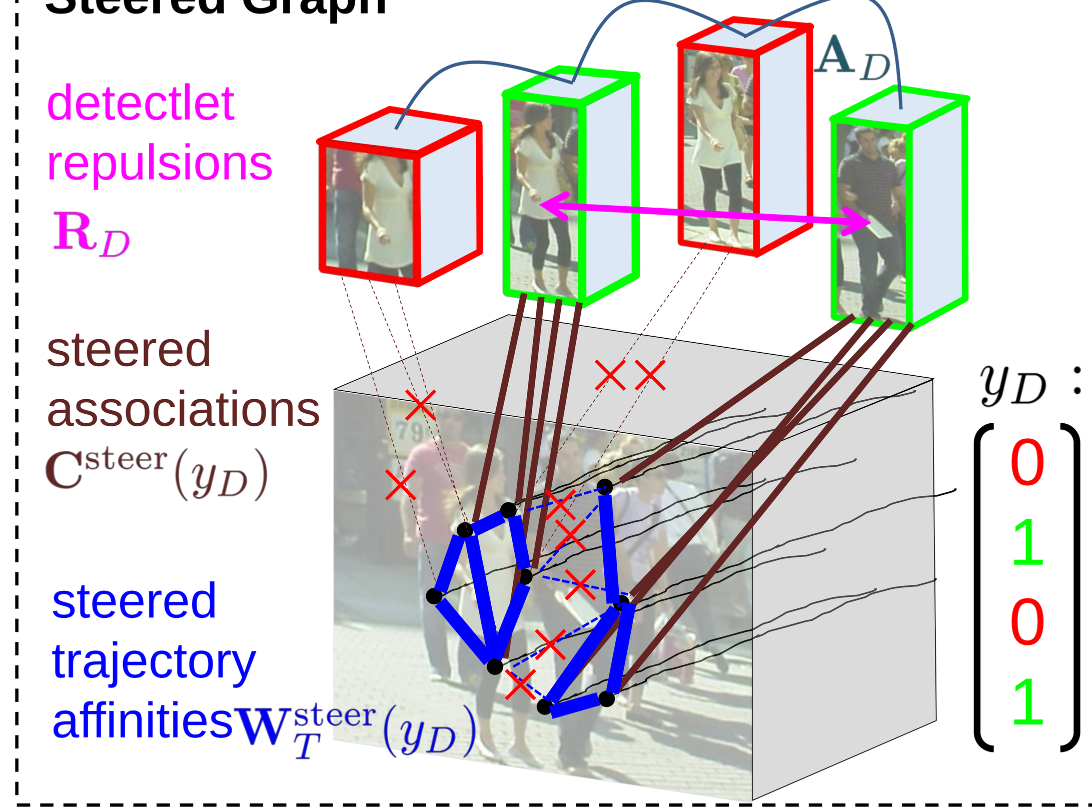


Detectlet Selection $y_D = [0, 1, 0, 1]$

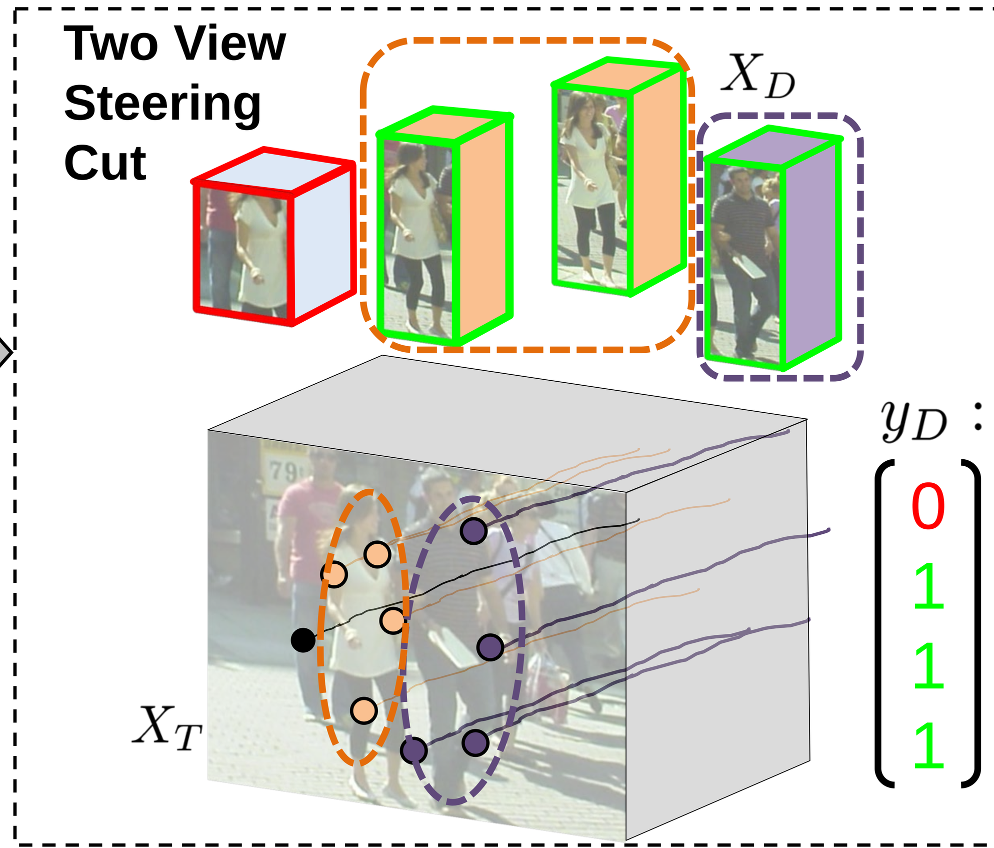


Detectlets can be true or false positives. y_D encodes the detectlets to be on/off. Initialized by confidence.

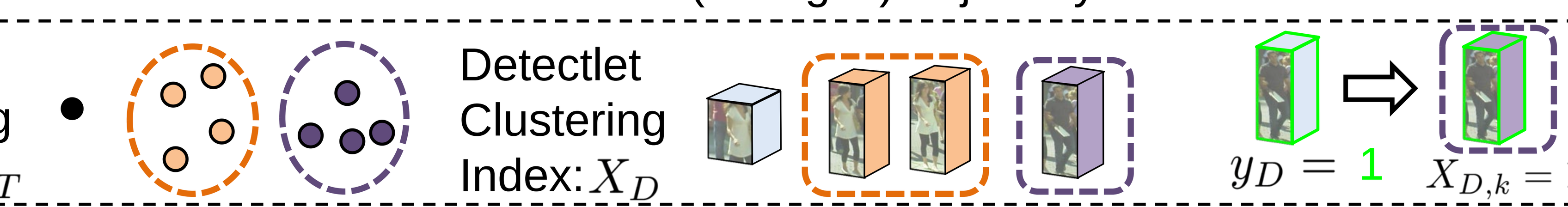
Steered Graph



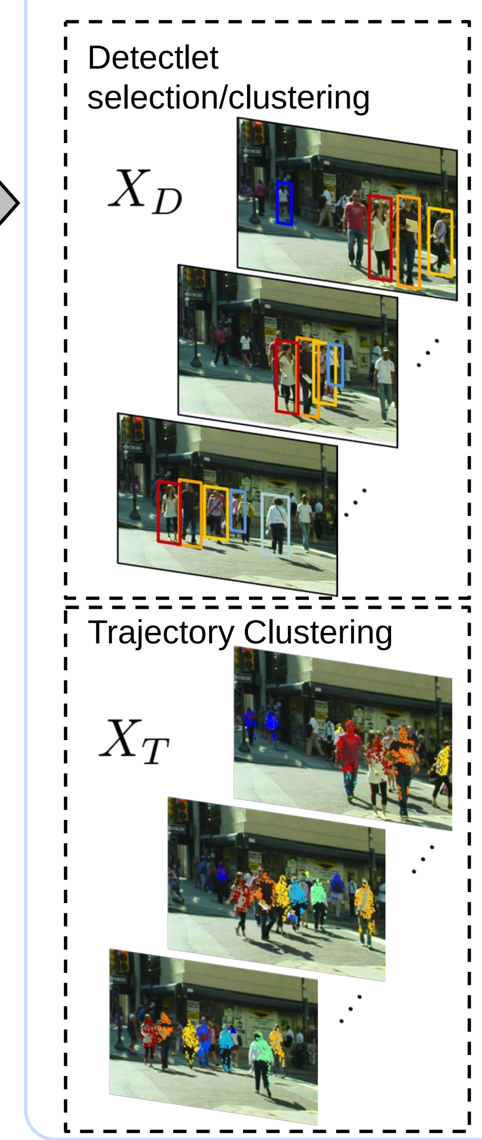
Two View Steering Cut



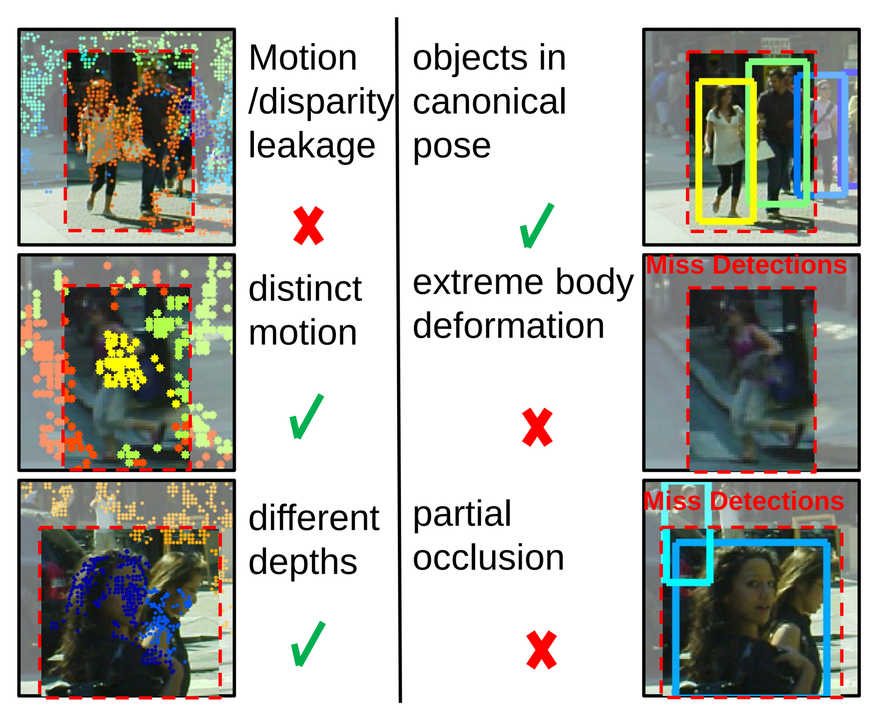
Intuition: We cast two granularity tracking as **selection** and **clustering** in the graph of detectlets and trajectories. Detectlets' selection steers (changes) trajectory affinities and cross-associations.



Output Result



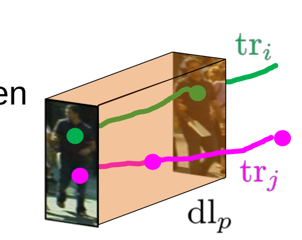
Two tracking granularities are complimentary



Initial Graph: $\mathbf{W} = \begin{bmatrix} \mathbf{A}_T & \mathbf{C} \\ C^\top & \mathbf{A}_D \end{bmatrix}$

Cross-Associations $\mathbf{C}$

Encode spatial overlap between detectlets and trajectories



complete overlap: $\mathbf{C}(\mathrm{tr}_i, \mathrm{dl}_p) = 1$

partial overlap: $\mathbf{C}(\mathrm{tr}_j, \mathrm{dl}_p) = 0$

Encode similarity between detectlets, based on trajectory anchoring:

Trajectory Affinities $\mathbf{A}_T$

Encode motion/disparity similarity between trajectories:

$$\mathbf{A}_T(\mathrm{tr}_i, \mathrm{tr}_j) = \exp\left(\left[-\mathrm{dst}_{i,j}\left(\mathrm{vel}_{i,j}^2/\sigma_v^2 + \mathrm{dsp}_{i,j}^2/\sigma_d^2\right)\right]\right)$$

$\mathrm{vel}_{i,j}, \mathrm{dsp}_{i,j}$: max velocity /disparity difference $\mathrm{dst}_{i,j}$: max Euclidian distance

Detectlet Affinities $\mathbf{A}_D$

$$\mathbf{A}_D(\mathrm{dl}_p, \mathrm{dl}_q) = \exp\left(-\frac{\mathrm{mean}_i |\mathbf{loc}(\mathrm{tr}_i, \mathrm{box}_p^{t_1}) - \mathbf{loc}(\mathrm{tr}_i, \mathrm{box}_q^{t_2})|^2}{\sigma_l^2}\right)$$

Steered Graph: $\mathbf{W}^{\mathrm{steer}}(y_D) = \begin{bmatrix} \mathbf{W}_T^{\mathrm{steer}}(y_D) & \mathbf{C}^{\mathrm{steer}}(y_D) \\ (\mathbf{C}^{\mathrm{steer}}(y_D))^\top & \mathbf{A}_D \end{bmatrix}$

Detectlet Repulsions $\mathbf{R}_D$

Set repulsions between overlapping in time detectlets, denoting they belong to distinct objects:

$$\mathbf{R}_D(\mathrm{dl}_p, \mathrm{dl}_q; y_D) = y_D(p) y_D(q) \delta(|T_p \cap T_q| > 0)$$

Steered Trajectory Affinities $\mathbf{W}_T^{\mathrm{steer}}(y_D)$, Associations $\mathbf{C}^{\mathrm{steer}}(y_D)$

Detectlet repulsion induces trajectory repulsions between associated trajectories:

$$\mathbf{R}_T(\mathrm{tr}_i, \mathrm{tr}_j; y_D) = \max_{p,q} \mathbf{C}_{i,p}(1 - \mathbf{C}_{i,q}) \mathbf{R}_D(\mathrm{dl}_p, \mathrm{dl}_q; y_D) \mathbf{C}_{j,q}(1 - \mathbf{C}_{i,p})$$

Attractions on the repulsive trajectory links is cancelled:

$$\mathbf{W}_T^{\mathrm{steer}}(y_D) = (\mathbf{1}_{n_T \times n_T} - \mathbf{R}_T(y_D)) \bullet \mathbf{A}_T$$

Cancel associations to non-selected detectlets: $\mathbf{C}^{\mathrm{steer}}(y_D) = \mathbf{C}\mathrm{Diag}(y_D)$

Steering Cut:

$$\max_{y_D, X, K} \sum_{k=1}^{K} \mathrm{align}(X_{T,k}, X_{D,k}) \cdot \mathrm{confidence}(X_{D,k}) \cdot \mathrm{ncut}(\mathbf{W}^{\mathrm{steer}}(y_D), X_k)$$

$$\text{s.t.} \quad \sum_{k=1}^{K} X_{T,k} = \mathbf{1}_{n_T}, \quad \sum_{k=1}^{K} X_{D,k} = y_D.$$

Alignment score between point trajectories and detectlets

$$\mathrm{align}(X_{T,k}, X_{D,k}) = \frac{|\{i| \ X_{T,k}(i) = 1 \wedge \exists j, \mathbf{C}(\mathrm{tr}_i, \mathrm{dl}_j) = 1\}|}{|\{i| \ X_{T,k}(i) = 1 \vee \exists j, \mathbf{C}(\mathrm{tr}_i, \mathrm{dl}_j) = 1\}|}$$

Confidence score of detectlets in each joint cluster

$$\mathrm{confidence}(X_{D,k}) = \mathbf{f}^\top X_{D,k}$$

Normalized-cut score

$$\mathrm{ncut}(\mathbf{W}^{\mathrm{steer}}(y_D), X_k) = \frac{X_k^\top \mathbf{W}^{\mathrm{steer}}(y_D) X_k}{X_k^\top \mathrm{Diag}(\mathbf{W}^{\mathrm{steer}}(y_D)\mathbf{1}_n) X_k}$$

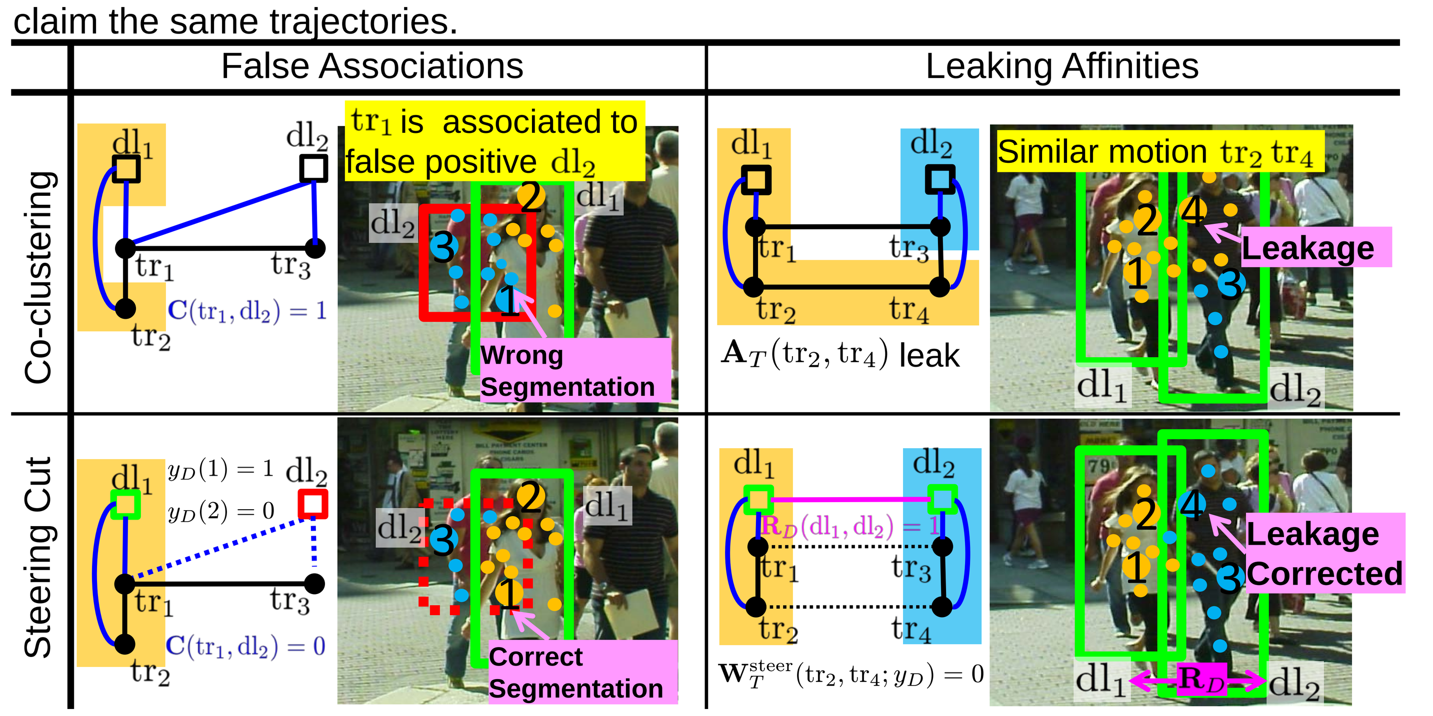
We solve two-view steering cut by **multiple** steered segmentations:

1. Initialize y_D by sampling high-confident detectlets. Compute $\mathbf{W}^{\mathrm{steer}}(y_D)$
2. Vary K and solve multiple ncuts, score clusters' alignment and confidence.
3. Populate the tracking solution by greedily selecting non-overlapping clusters .
4. Update y_D with the detectlets chosen in the tracking solution.

Steering cut vs Co-clustering

In steering cut the graph changes according to detectlet selection:

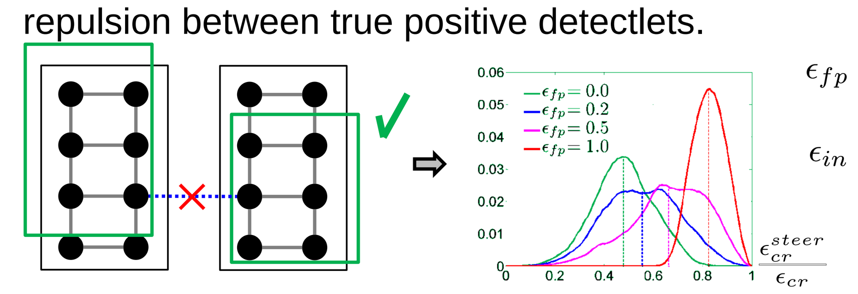
- Affinity contradictions between trajectories and detectlets are resolved by induced trajectory repulsions.
- False associations to non-selected detectlets are cancelled so multiple detectlets cannot claim the same trajectories.



Steered Graph vs Initial Graph

Graph steering:

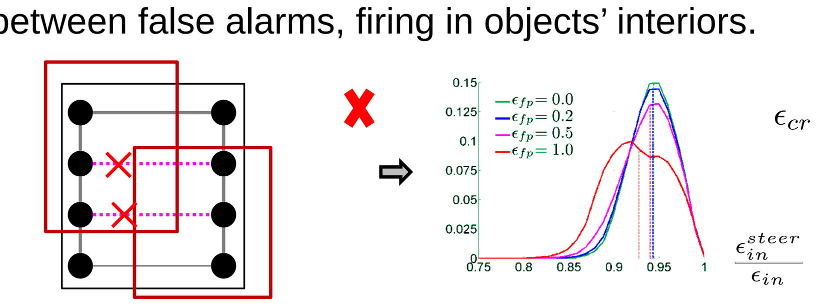
- Correctly cancels cross-object leaking affinities in case of repulsion between true positive detectlets.



ϵ_{fp} : false alarm detectlet rate

ϵ_{in} : intra-object affinity rate

- Erroneously cancels intra-object affinities in case of repulsions between false alarms, firing in objects' interiors.

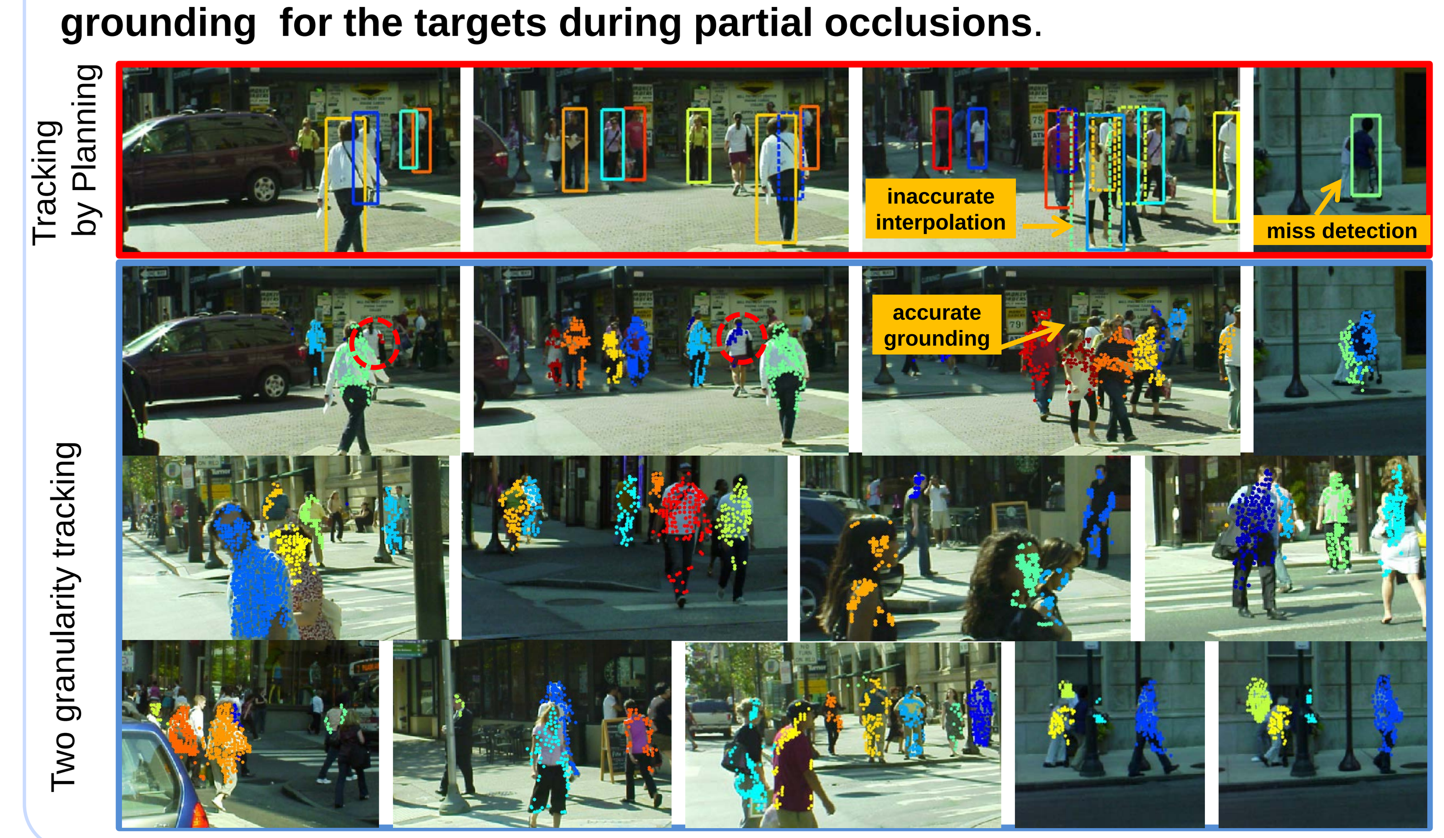


ϵ_{cr} : cross-object affinity rate

False alarms are randomly distributed in object interiors and cancel fewer links during steering. As such, **graph steering considerably improves affinity accuracy on average**!

Experiments

Two granularity tracking tolerates detection sparsity and provides **accurate grounding for the targets during partial occlusions**.

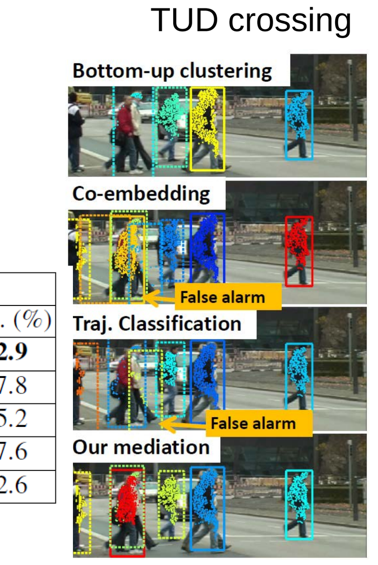


Two granularity tracking

- Correct leakages in bottom-up trajectory clustering, due to stationary/similar motion.
- Track under partial occlusions with accurate spatial mask.

	UrbanStreet				TUD crossing			
	MD(%)	FA(%)	ID-sw.	Acc.(%)	MD(%)	FA (%)	ID-sw.	Acc. (%)
Our Method	50.3	15.6	73	**30.1**	12.3	4.5	0	**82.9**
Co-embedding	57.7	47.0	72	-8.2	14.6	25.1	27	57.8
Trajectory classification	61.0	23.7	71	11.6	18.7	14.5	17	65.2
Bottom-up clustering	78.7	11.5	19	8.5	32.5	8.8	12	57.6
Detectlets	82.6	19.4	49	-4.7	42.6	7.3	81	42.6
Our Method*	44.7	12.0	28	**37.5**				
Gong et al.*	76.5	24.3	36	-6.8				

TUD crossing



UrbanStreet Dataset: new tracking dataset with accurate object mask