Multi-hypothesis Motion Planning for Visual Object Tracking

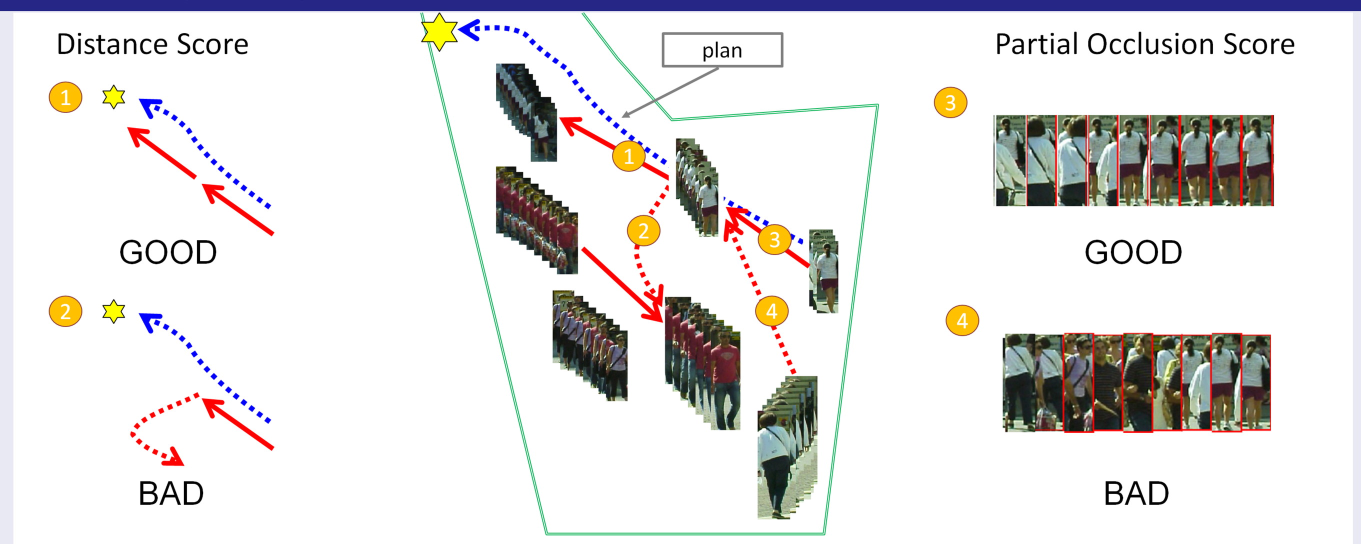
Haifeng Gong[†], Jack Sim[†], Maxim Likhachev[‡], Jianbo Shi[†]

[†] GRASP Lab, University of Pennsylvania [‡] Robotics Institute, Carnegie Mellon University

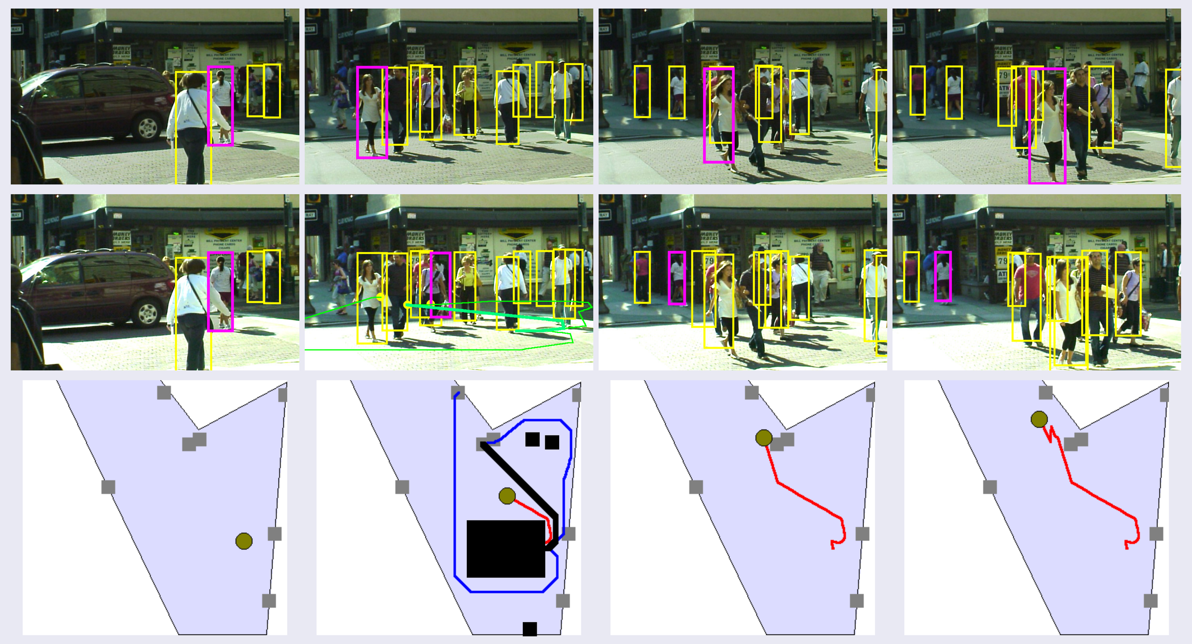
Motion planning as motion model for visual object tracking

- In crowded street scenes, frequent occlusions lead to ambiguous data association or 'drifting' in tracking.
- Many of these occlusions could be dealt with using a long-term motion model.
- We propose to construct a set of 'plausible' plans for each person.
  - multi-hypotheses,
  - no redundancy, no unnecessary loop,
  - no collisions with other objects.

Tracking with motion planning

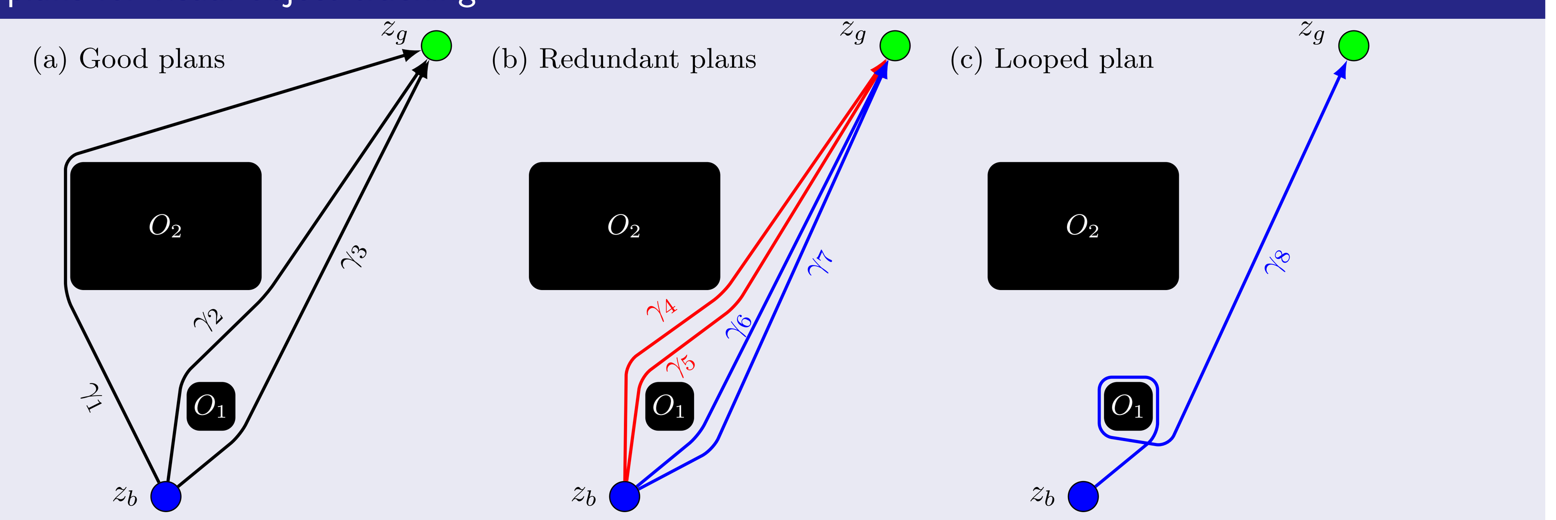


Tracking with multi-hypothesis motion planning



Top: tracking without planning. Middle: tracking with planning. Bottom, top view of tracking with planning. Note that we plan in advance, therefore, the obstacles are other objects a few frames ago.

Plausible plans for visual object tracking



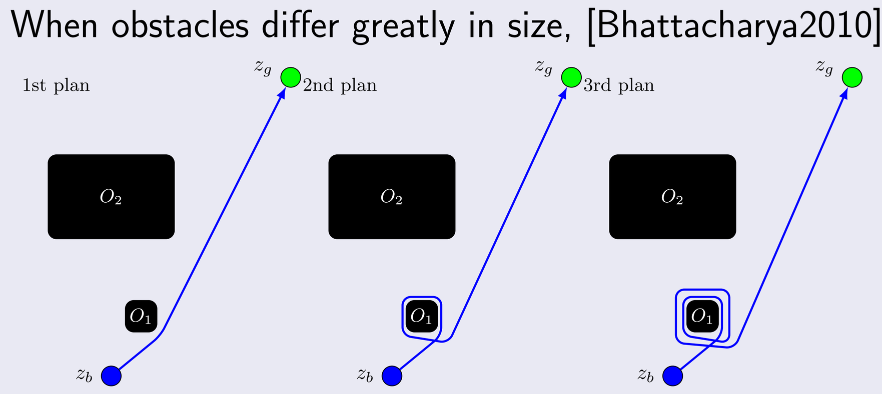
Examples of plausible plans and bad plans for visual object tracking. O_1 and O_2 are two obstacles. γ_i are possible paths. z_b and z_g are the start point and goal respectively.

Homotopy-class planning [Bhattacharya2010]

- Let z be a point in the complex plane, z_b the start point and z_g the goal of an agent (where it is intended to go). A path $\gamma(s)$ is a complex function of arc length parameter $s \in [0, T]$, with constraints $\gamma(0) = z_b$ and $\gamma(T) = z_g$.
- A complex *obstacle marker function* is defined as $F(z) = \frac{f_0(z)}{(z-\zeta_1)(z-\zeta_2)\cdots(z-\zeta_N)}$ where $f_0(z)$ is a complex Homomorphic function and ζ_i is a point in obstacle i.
- **Cauchy Integral Theorem** Two trajectories $\gamma_1(s)$ and $\gamma_2(s)$ connecting the same pair of points lie in the same homotopy class if and only if $\int_{\gamma_1} F(z)dz = \int_{\gamma_2} F(z)dz$.
- Therefore they use the L-value, defined as $L(\gamma) = \int_\gamma F(z)dz$ to index homotopy classes.

Drawbacks of [Bhattacharya2010]

1. When obstacles differ greatly in size, [Bhattacharya2010] performs poorly.



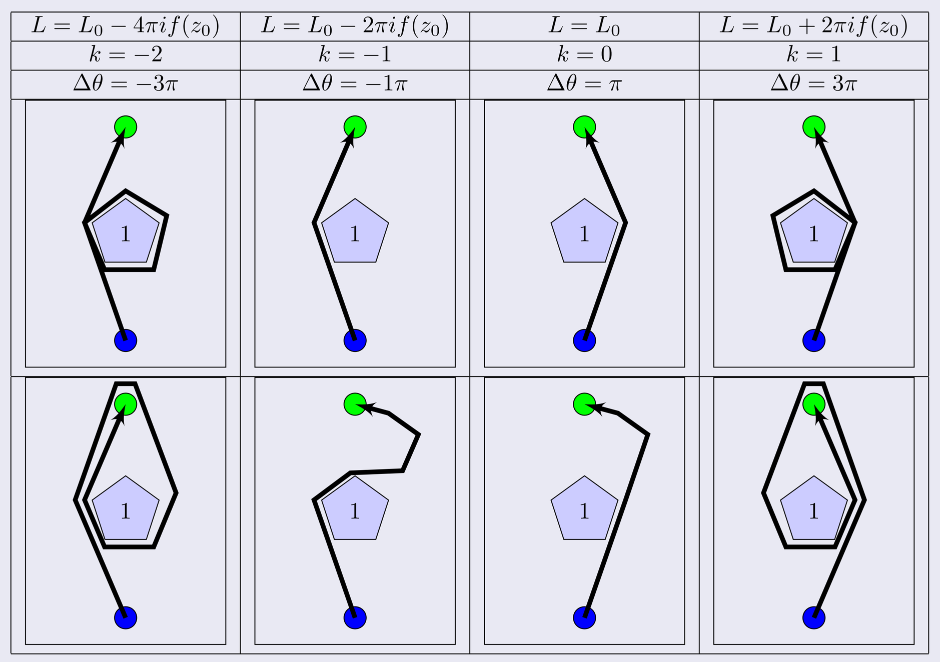
It might loop around small obstacles before taking bigger obstacles into account.

2. Obstacle marker function must be carefully chosen for numeric stability of L-values.
3. The representation of state space is an infinite augmented graph.

From L-value to winding numbers

- We propose replacing L-value with a more informative index, that incorporates the number of loops around obstacles.
- This allow us to screen out any paths with many loops, which are unlikely to be the paths that people actually take.
- The L-value of a plan γ with respect to a single obstacle is $L = \int_\gamma \frac{f(z)}{z-z_0}dz$.
- L-values for a single obstacle must be in the discrete set of $\{k * 2\pi i f(z_0) + L_0 : k \in \mathbb{Z}\}$.
- Thus we can use k (winding number) to distinguish homotopy classes with respect to one obstacle which

Example of winding numbers



- $k > 0$ indicates a path to the right of the obstacle that includes k loops around it.
- $k < -1$ indicates a path to the left of the obstacle that includes $-k-1$ loops around it.
- For a plausible path, the values of k will likely be 0 or -1, meaning 'go-right' or 'go-left' around the obstacle.

Vector of winding numbers

Definition By letting k_i be the k-value associated with the i-th obstacle, we can denote a homotopy class with respect to all obstacles as an integer vector (vector of winding numbers, or k-**vector**) $\mathbf{k} = (k_1, k_2, \cdots, k_N)^T$.

Theorem Two trajectories γ_1 and γ_2 with k-vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ connecting the same points lie in the same homotopy class if and only if $\mathbf{k}_1 = \mathbf{k}_2$.

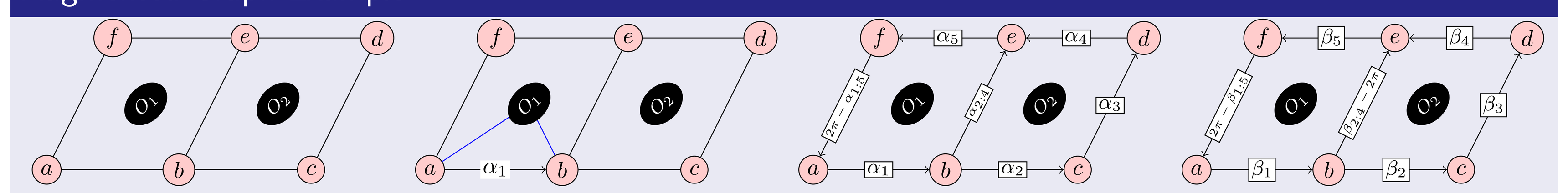
From winding numbers to winding angles

- A path γ can be written in parametric form, $\gamma(s) = z_0 + r(s)\exp[i\theta(s)]$.
- The obstacle marker function can be a constant $f(z) = 1$.
- Then L-value can be computed in closed form as $L = \text{Const} + i[\theta(T) - \theta(0)]$.
- The imaginary part $\Delta\theta = \theta(T) - \theta(0) = \Delta\theta_0 + 2k\pi$ may differ by $2k\pi$, where k is also a winding number.
- We call $\Delta\theta$ the winding angle of γ w.r.t. obstacle z_0.

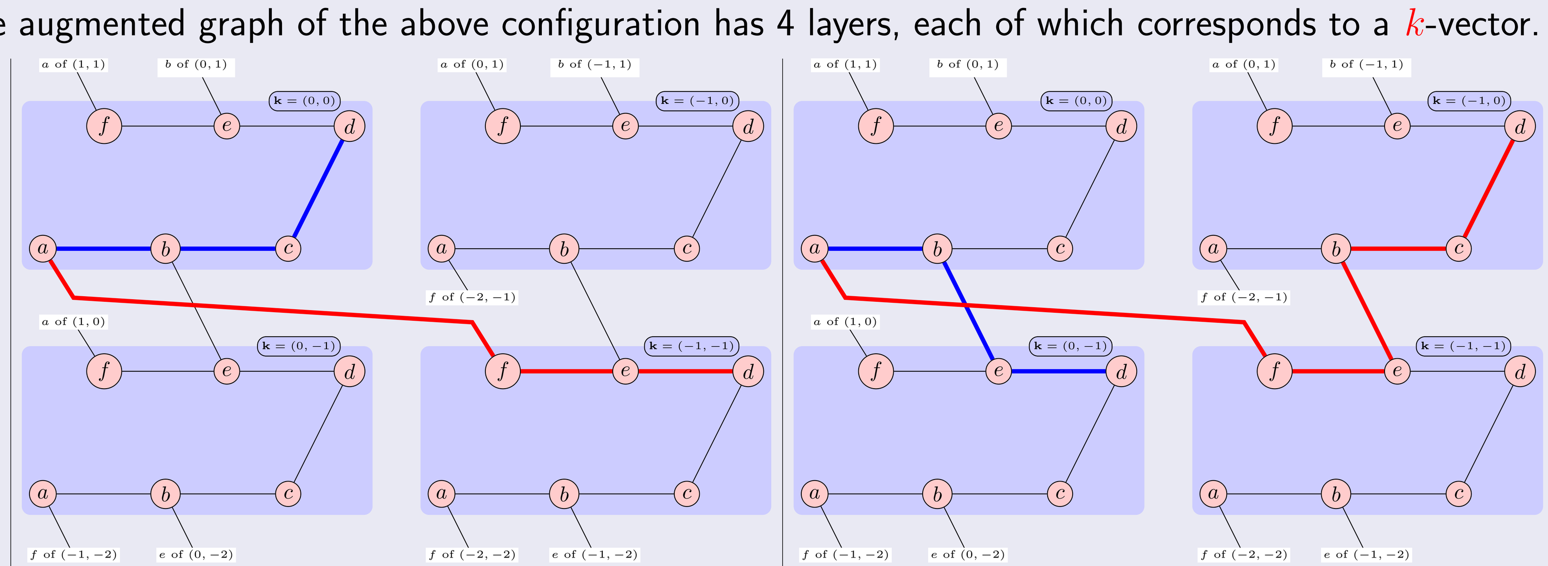
Augmented Graph

- Like [Bhattacharya2010], we use a graph based search algorithm, but we search on a finite graph.
- We begin with neighborhood graph G, in which each grid point on ground not occupied by an obstacle is a vertex, and each pair of neighboring points are connected by an edge.
- Each vertex in G is represented by its coordinate on ground z.
- We augment this graph with winding angle to create an augmented graph $\bar{G}$.
- We equip both vertices with winding angles and edges with increments of winding angles.

Augmented Graph Example



The augmented graph of the above configuration has 4 layers, each of which corresponds to a k-vector.



- Blue bold lines, planned path from start to the goal in homotopy class $(0, 0)$.
- Red bold lines, planned path from start to the goal in homotopy class $(-1, -1)$.
- Blue bold lines, planned path from start to the goal in homotopy class $(0, -1)$.
- Red bold lines, planned path from start to the goal in homotopy class $(-1, 0)$.

Tracking by Planning

- We test our motion model in a batchmode tracking by detection framework.
- Tracking a person in the visible state leads to a short trajectory that we call a tracklet.
- A conservative threshold is used to terminate the trajectory when the tracking score becomes too low.
- After termination, the same person may be picked up again by the detection algorithm, and tracked to produce associated tracklets.
- After tracklets are obtained, we can link them using both appearance and planning consistency.

Criteria for tracklets linking by planning

- Assume that we have a set of tracklets $\mathcal{T} = \{F_1, \cdots, F_{N_{\text{Tr}}}\}$.
- Each tracklet is described by 3D point series.
- We then link and extend these tracklets, $\mathcal{T}$, into complete trajectories.
- To link tracklets into plausible goal-directed obstacle-avoiding paths, we design the following criterion for tracking: $\max_L \epsilon(L) = \sum_{i,j:L_{i,j}=1}[S_{\text{App}}(i,j) + \alpha S_{\text{Plan}}(i,j)]$.
- $S_{\text{App}}(i,j)$ measures appearance similarity between tracklets F_i and F_j,
- $S_{\text{Plan}}(i,j)$ measures 1) how consistent F_i and F_j are with a plausible goal directed path; and 2) how partial occlusion in the gap can be explained by appearance of F_i and F_j.

Planning score

- The planning score is given by finding the best planned path to fill the gap between tracklet i and j.
- The best path is compatible with tracklet i and tracklet j geometrically, and allows possible partial matches by appearance during occlusions.
- We use the following score:

$$S_{\text{Plan}}(i,j) = \max_{r \in \text{paths}} -\text{Dist}(r, F_i) - \text{Dist}(r, F_j) + S_{\text{Occl}}(F_i, F_j, r),$$

  where $\text{Dist}(r, F_i)$ is the distance between path r and tracklet F_i and $S_{\text{Occl}}(F_i, F_j, r)$ is the score for picking up the partial occlusions along the gap.
- To reduce computation, we prune paths whose costs are higher than the minimal one above a threshold.

Experiment Setting

Street Scene To test our algorithm we have collected a video from a moving vehicle in an urban city.

Binocular Sensor The stereo images were collected at 1024×768 resolution and 6 FPS.

3D 3D scene layout/goal estimation, and camera ego-motion computation.

Detection 3D people detection (based on [Felzenszwalb2008]).

Goals We estimated building planes and ground plane in each frame and intersected them to get street side lines. The goals are estimated by intersecting the street side lines, plus infinity points along the street.

Obstacles We only track people, but detect cars as dynamic obstacles. When planning for a specified object, other objects are regarded as obstacles.

Data and Result Comparison

	# obj	# frames	# BB	#Occl. BB
seq #1	13	169	1139	471
seq #2	12	60	532	130
seq #3	7	35	210	125
seq #4	4	40	148	51
seq #5	5	112	211	46
seq #6	5	41	170	17
seq #7	2	27	54	16
Total	48	484	2464	856

Test Videos with 3 difficulty levels according to the number of occluded bounding boxes. (BB = Bounding boxes.)

		miss rate	fa rate	id switch
seq #1	PLAN	0.413	0.089	9
	LINEAR	0.442	0.070	8
	LTA	0.488	0.214	8
seq #2	PLAN	0.259	0.193	0
	LINEAR	0.330	0.199	4
	LTA	0.366	0.310	6
seq #3	PLAN	0.311	0.223	1
	LINEAR	0.340	0.200	2
	LTA	0.476	0.445	6
seq #4	PLAN	0.176	0.00	0
	LINEAR	0.176	0.110	0
	LTA	0.270	0.212	0
seq #5	PLAN	0.137	0.032	0
	LINEAR	0.123	0.016	0
	LTA	0.189	0.090	0
seq #6	PLAN	0.147	0.194	0
	LINEAR	0.153	0.152	6
	LTA	0.211	0.394	5
seq #7	PLAN	0.056	0.00	0
	LINEAR	0.056	0.00	0
	LTA	0.203	0.157	0

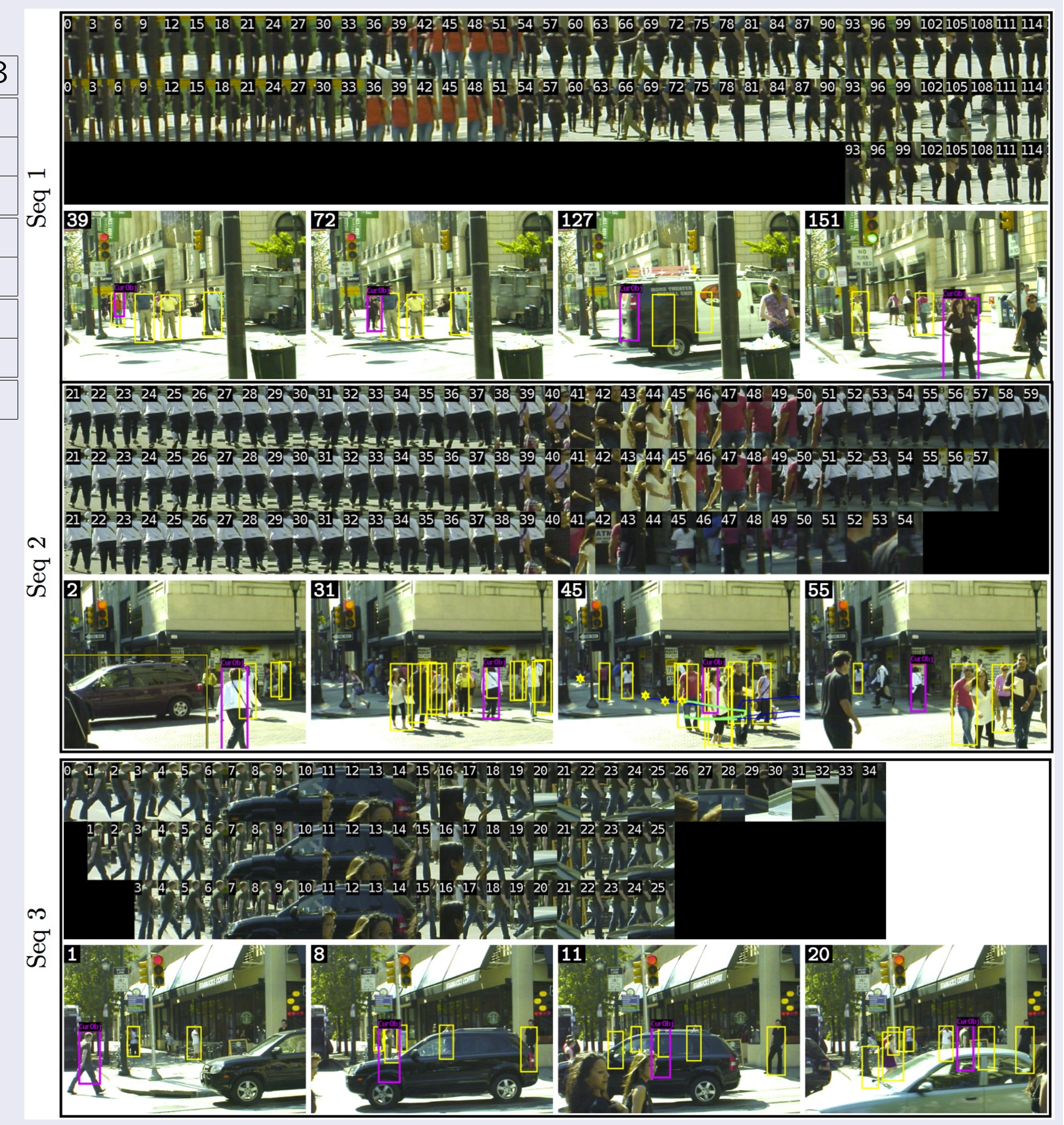


Image patches and bounding boxes over time. Each panel shows the bounding boxes of a pedestrian in two parts. The top parts show the image patches of ground truth (1st row), PLAN results (2nd row) and LINEAR results (3rd row). The number on each box is the frame number. They are trimmed on left or right for better visual effects. The bottom parts show video frames superimposed with bounding boxes. The magenta bounding boxes are current objects of interests. Yellow bounding boxes are other objects. The bold green lines are the planned routes that the objects follow. The thinner green lines are other planned paths (after pruning) that are not followed by the people.