# FollowMe: Efficient Online Min-Cost Flow Tracking with Bounded Memory and Computation

Philip Lenz[1], Andreas Geiger[2], Raquel Urtasun[3]

[1]Karlsruhe Institute of Technology
lenz@kit.edu

[2]MPI for Intelligent Systems, Tübingen
andreas.geiger@tue.mpg.de

[3]University of Toronto
urtasun@cs.toronto.edu

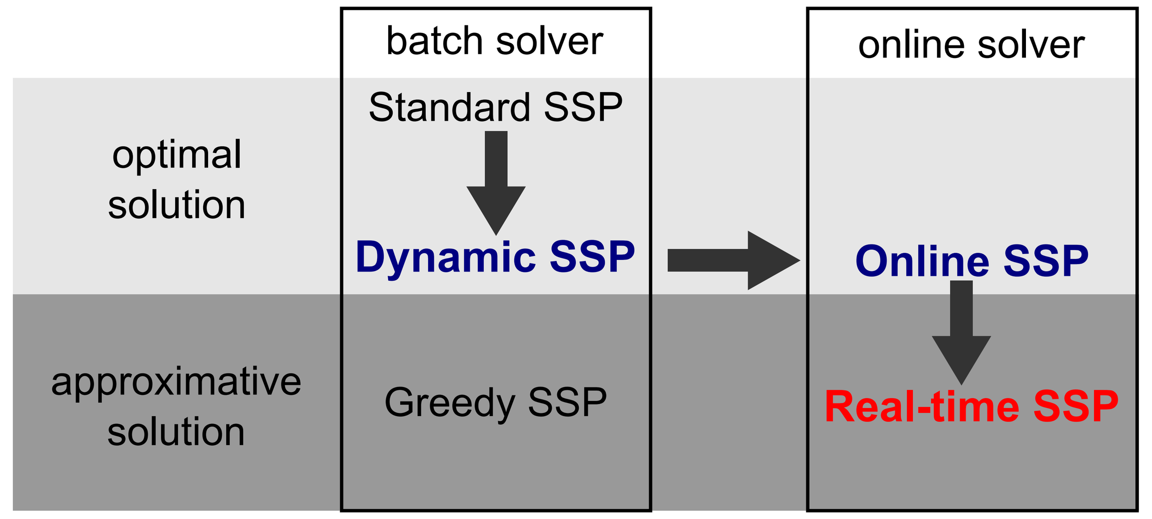
Project Page with Code
http://www.cs.toronto.edu/~boundTracking

## Abstract

One of the most popular approaches to multi-target tracking is tracking-by-detection. Current min-cost flow algorithms which solve the data association problem optimally have three main drawbacks: they are computationally expensive, they assume that the whole video is given as a batch, and they scale badly in memory and computation with the length of the video sequence. In this paper, we address each of these issues, resulting in a computationally and memory-bounded solution.

## Contribution

- Dynamic version of the **successive shortest-path algorithm**.
- Solves the data association problem optimally, reuses computation.
- Faster inference than standard solvers.
- Optimal data association for an online setting.
- Approximate online solution with **bounded memory and computation**.
- Capable of handling videos of arbitrary in real time.
- State-of-the-art results on KITTI and PETS2009 benchmarks.

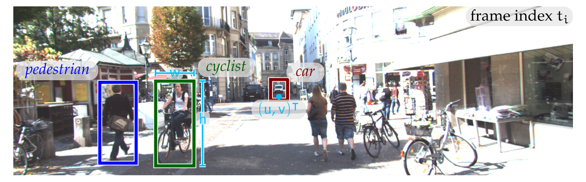



**Results on KITTI Tracking Benchmark.** State-of-the-art performance for significantly faster inference than standard solvers.

## Problem Statement

- Given a set of detections $\mathcal{X} = \{\mathbf{x}_i\}$ (car, pedestrian, and cyclist) per frame with $\mathbf{x}_i$ = (frame, position, size, appearance, detector score)
- Find the set of optimal trajectories $\mathcal{T} = \{T_k\}$ by maximizing the posterior probability $p(\mathcal{T}|\mathcal{X}) = p(\mathcal{T}) \prod_i (\mathbf{x}_i|\mathcal{T})$
- Zhang et al. [7] showed how data association with pairwise energies can be formulated as a network flow problem.
- Standard solvers can be leveraged to retrieve the global optimum.
- Their formulation solves for the globally optimal trajectories including their number, and hence implicitly solves the model selection problem.

## Graph Construction




2D Bounding Boxes as Input Detections [2]

Observation model

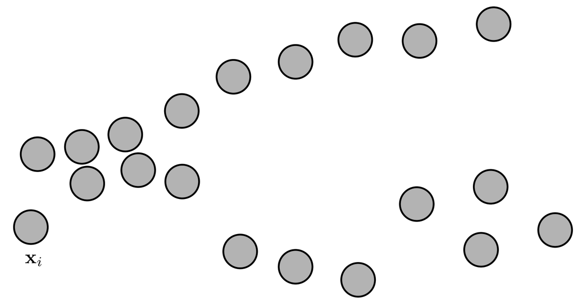
$$p(\mathbf{x}_i|\mathcal{T}) = \begin{cases} P_i & \text{if } \exists T_k \in \mathcal{T} \wedge i \in T_k \\ 1 - P_i & \text{otherwise} \end{cases}$$

Pair-wise prior over trajectories

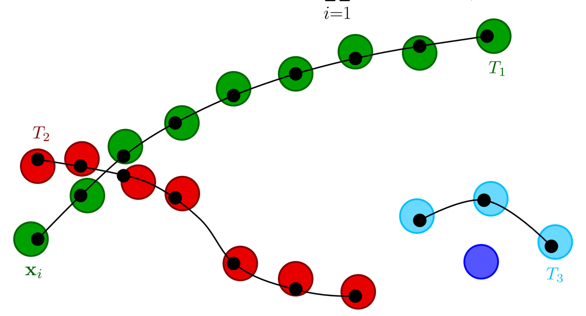
$$p(\mathcal{T}) \propto \prod_{T \in \mathcal{T}} \Psi(T) \prod_{T,T' \in \mathcal{T}} [T \cap T' = \emptyset]$$

Unary factors

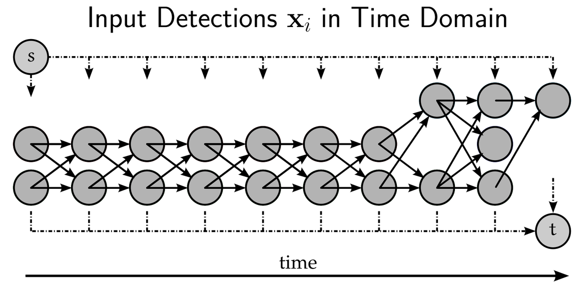
$$\Psi(T) = \Psi_{en}(\mathbf{x}_{o_1}) \Psi_{ex}(\mathbf{x}_{o_l}) \prod_{i=1}^{l-1} \Psi_{li}(\mathbf{x}_{o_i}, \mathbf{x}_{o_{i+1}})$$



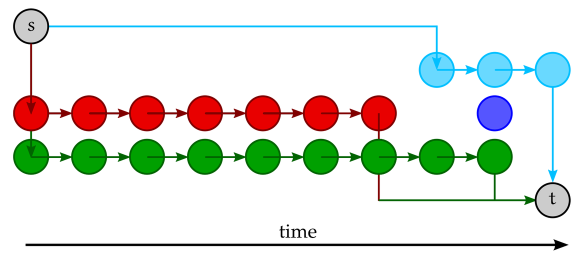

Input Detections $\mathbf{x}_i$ in Time Domain




Final World Association

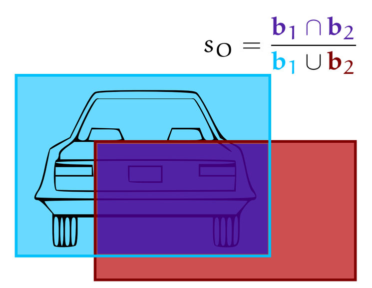



Possible Trajectories as Graph Representation

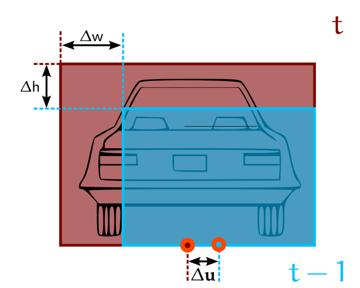



Optimal Association as Shortest Paths

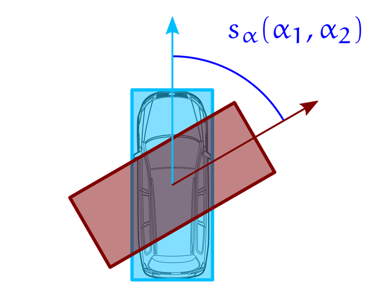
## Association Features



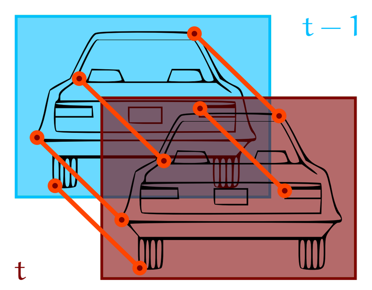

Bounding Box Overlap



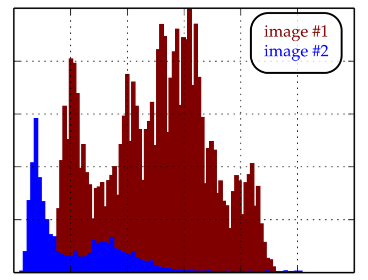

Positional and Size Similarity



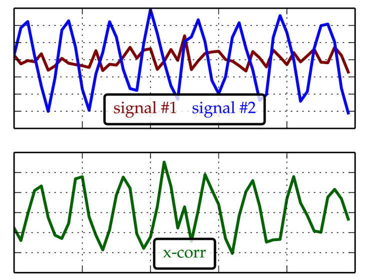

Orientation Similarity




Flow Overlap



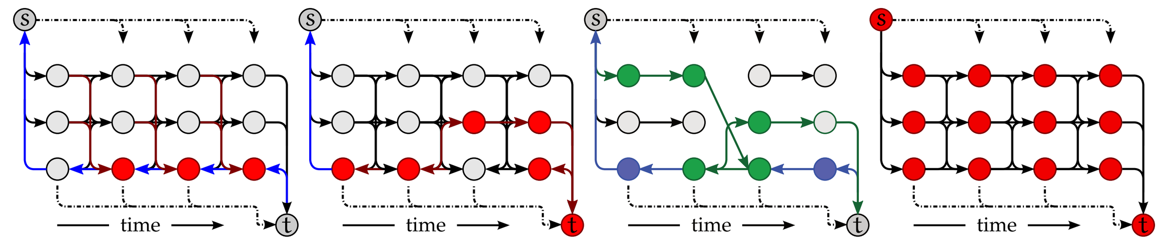

Color Histogram Similarity




Template Matching

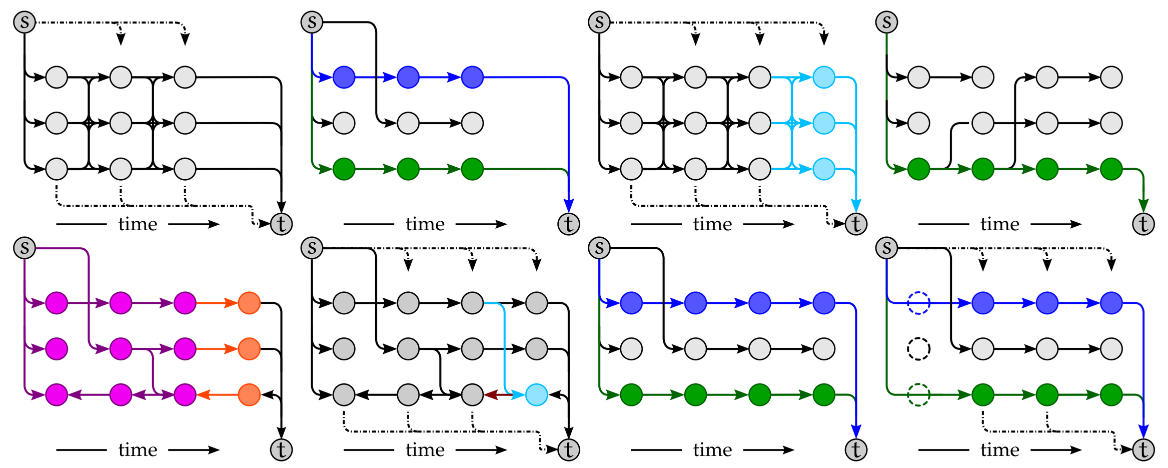
## Dynamic Successive Shortest Path Algorithm

- *Dynamic* algorithm. Performs computations only when necessary.
- Nodes are held in a priority queue.
- The most promising node is relaxed in each iteration.
- The queue is dynamically updated depending on previous updates.

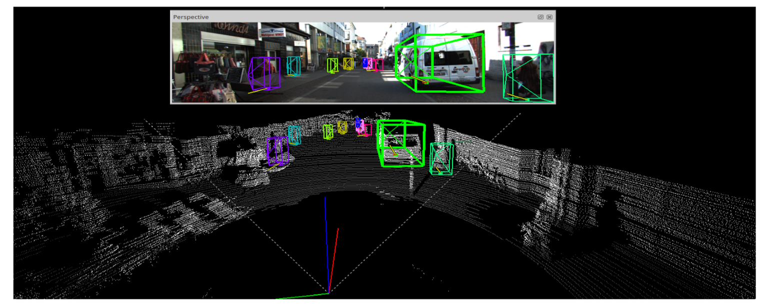



## Online Memory-bounded SSP

- Most of the past can be neglected for a tight approximation.
- Only trajectories that were optimal are retained.
- Simply deleting edges from the graph is suboptimal and discards previous computations.
- In order to "remember" known paths, costs of clipped trajectories are summed up and new entry edges added to the graph.




## Ground Truth: KITTI Tracking Dataset



- 3D Ground Truth Tracklets
- 1249 Car Tracklets
- 459 Pedestrian Tracklets

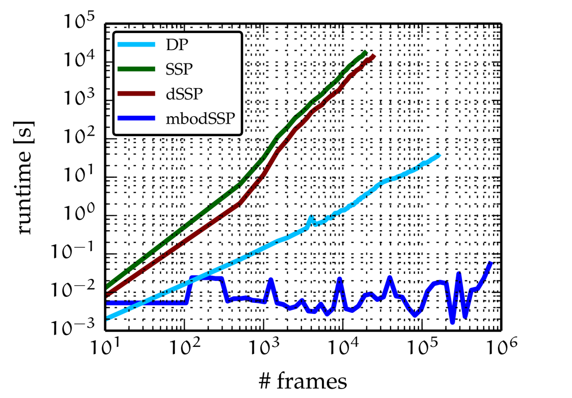
## Qualitative Results

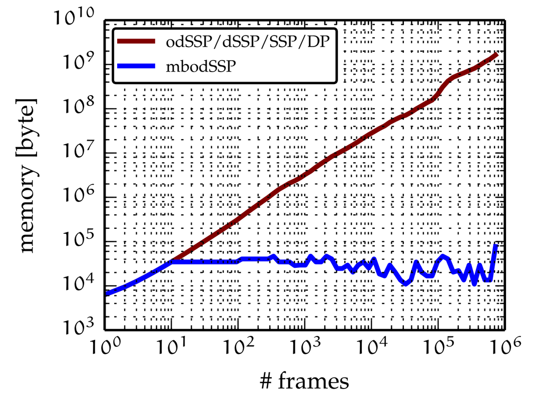
**KITTI-Tracking Benchmark.** Stable and consistent results for long-term and complex scenes including challenging illumination or partial occlusions.

## Quantitative Results

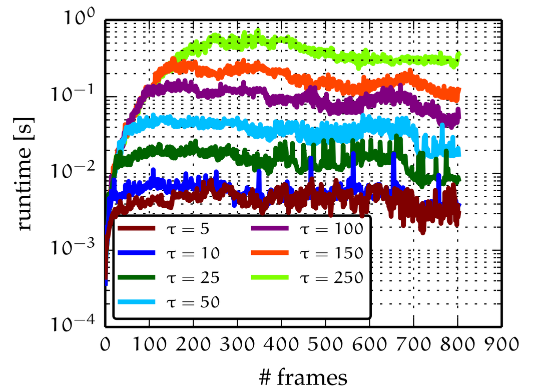
| | HM | [1] | [4] | [5] | [3] | mbodSSP | SSP | mbodSSP* | SSP* |
|---|---|---|---|---|---|---|---|---|---|
| MOTA | 0.42 | 0.35 | 0.48 | 0.44 | 0.52 | 0.52 | **0.54** | 0.67 | 0.67 |
| MOTP | 0.78 | 0.75 | 0.77 | 0.78 | 0.78 | 0.78 | 0.78 | 0.79 | 0.79 |
| F1 | 0.60 | 0.61 | 0.67 | 0.62 | 0.69 | **0.70** | **0.71** | 0.83 | 0.83 |
| FAR | **0.048** | 0.46 | 0.18 | 0.053 | 0.083 | 0.14 | 0.11 | 0.34 | 0.40 |
| MT | 0.077 | 0.11 | 0.14 | 0.11 | 0.14 | **0.15** | **0.21** | 0.34 | 0.41 |
| ML | 0.42 | 0.34 | 0.34 | 0.39 | 0.35 | **0.30** | **0.27** | 0.10 | 0.090 |
| IDS | 12 | 223 | 125 | 2738 | 33 | **0** | **7** | 117 | 194 |
| Frag. | 578 | 624 | **401** | 3241 | 540 | 708 | 717 | 894 | 977 |

**Quantitative Results on the KITTI-Car Dataset.** Comparison of our proposed methods to four state of the art methods and a HM baseline implementation. Detections: DPM [2] reference detections and Regionlets [6] (marked with a star).




**Run Time and Memory Comparison.** Computational performance of all solvers using one long sequence. From left to right: Mean runtime and idealized memory consumption for every solver.




| $\tau$ | 5 | 10 | 15 | 20 | 50 | 100 |
|---|---|---|---|---|---|---|
| MOTA | 0.50 | **0.52** | 0.48 | 0.49 | 0.51 | **0.52** |
| MOTP | **0.78** | **0.78** | **0.78** | **0.78** | **0.78** | **0.78** |
| F1 | 0.69 | 0.70 | 0.68 | 0.69 | 0.70 | **0.71** |
| FAR | 0.15 | **0.14** | **0.14** | **0.14** | **0.14** | **0.14** |
| MT | 0.14 | 0.15 | 0.16 | 0.17 | **0.18** | 0.15 |
| ML | 0.30 | 0.30 | 0.34 | 0.33 | 0.30 | **0.26** |
| IDS | 2 | **0** | **0** | **0** | 4 | 5 |
| Frag. | 703 | 708 | **690** | 701 | 710 | 712 |

**Online Approximation Performance.** Impact of different values for the history length $\tau$.

## References

[1] Anton Andriyenko, Konrad Schindler, and Stefan Roth. Discrete-continuous optimization for multi-target tracking. In *CVPR*, 2012.

[2] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *PAMI*, 32:1627–1645, 2010.

[3] Andreas Geiger, Martin Lauer, Christian Wojek, Christoph Stiller, and Raquel Urtasun. 3D traffic scene understanding from movable platforms. *PAMI*, 36(5):1012–1025, 2014.

[4] A. Milan, S. Roth, and K. Schindler. Continuous energy minimization for multitarget tracking. *PAMI*, 36(1):58–72, 2014.

[5] Hamed Pirsiavash, Deva Ramanan, and Charless C. Fowlkes. Globally-optimal greedy algorithms for tracking a variable number of objects. In *CVPR*, 2011.

[6] Xiaoyu Wang, Ming Yang, Shenghuo Zhu, and Yuanqing Lin. Regionlets for generic object detection. In *ICCV*, 2013.

[7] L. Zhang, Y. Li, and R. Nevatia. Global data association for multi-object tracking using network flows. In *CVPR*, 2008.