Carnegie Mellon University
The Robotics Institute

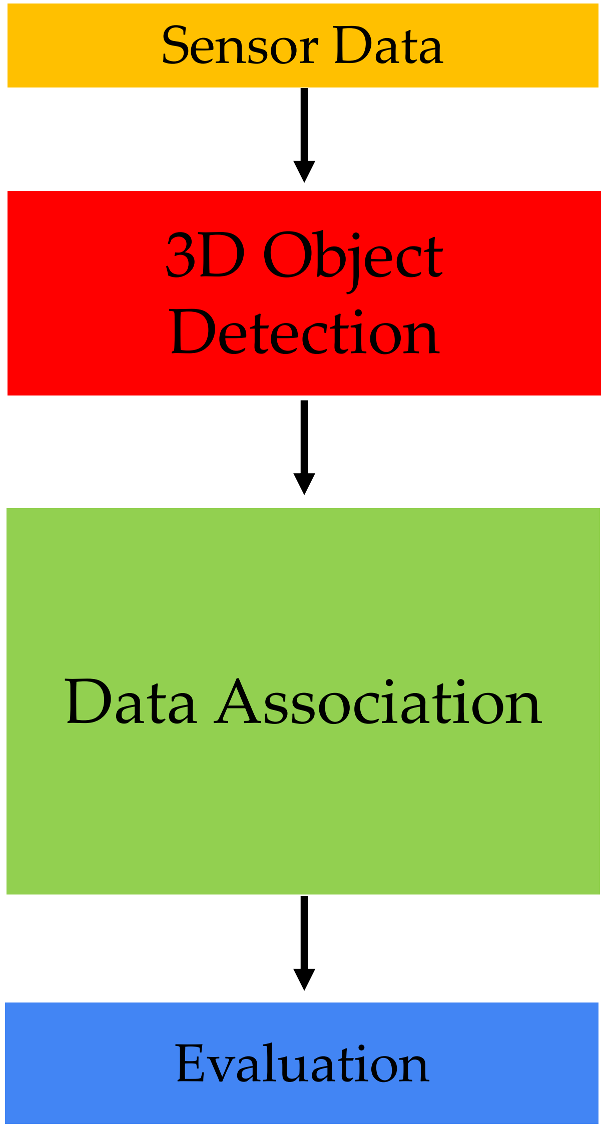
3D Multi-Object Tracking: A Baseline and New Evaluation Metrics

Xinshuo Weng, Jianren Wang, David Held, Kris Kitani
Robotics Institute, Carnegie Mellon University

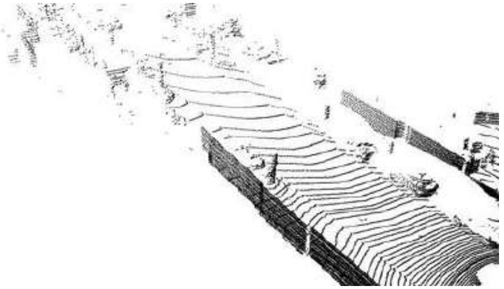
*IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**), 2020*

THE ROBOTICS INSTITUTE

Standard 3D MOT Pipeline



Standard 3D MOT Pipeline



LiDAR point clouds

Sensor Data

↓

3D Object Detection

↓

Data Association

↓

Evaluation

RGB frames

Standard 3D MOT Pipeline

Sensor Data

3D Object Detection

Data Association

Evaluation

Detection results

Standard 3D MOT Pipeline

Sensor Data

↓

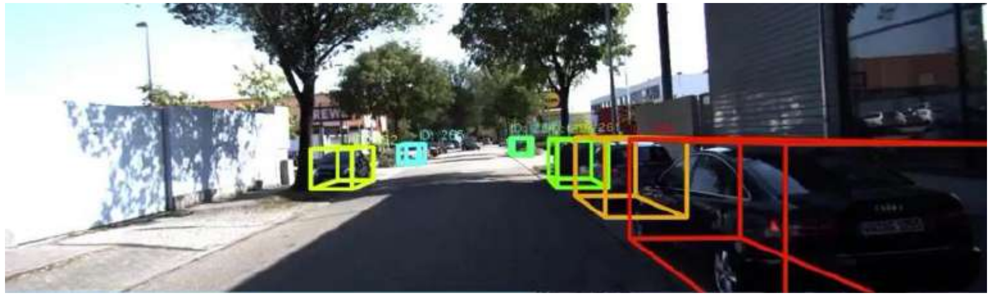
3D Object Detection

↓

Data Association

↓

Evaluation



3D MOT results

Standard 3D MOT Pipeline

Sensor Data

↓

3D Object Detection

↓

Data Association

↓

Evaluation ← Also important!

Evaluation:

1. MOTA: MOT accuracy
2. MOTP: MOT precision
3. IDS: # of identity switches
4. FRAG: # of trajectory fragments
5. ……

Standard 3D MOT Pipeline

Sensor Data

↓

3D Object Detection

↓

Data Association ← Limitation: ignore practical factors such as speed and system complexity

↓

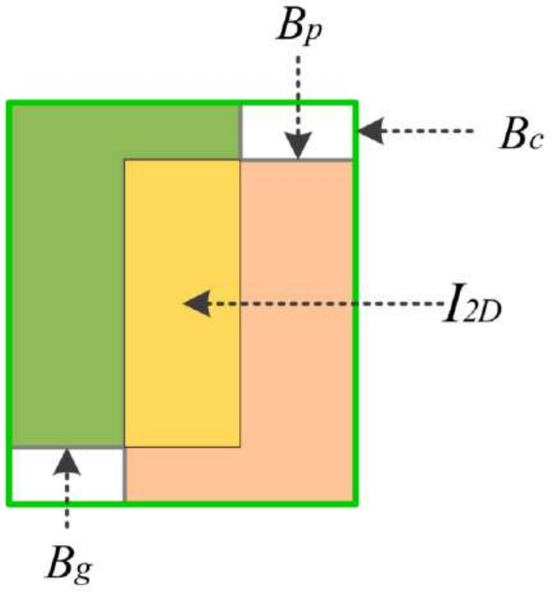
Evaluation ← Limitation: appropriate 3D MOT evaluation is not available

Our Contributions

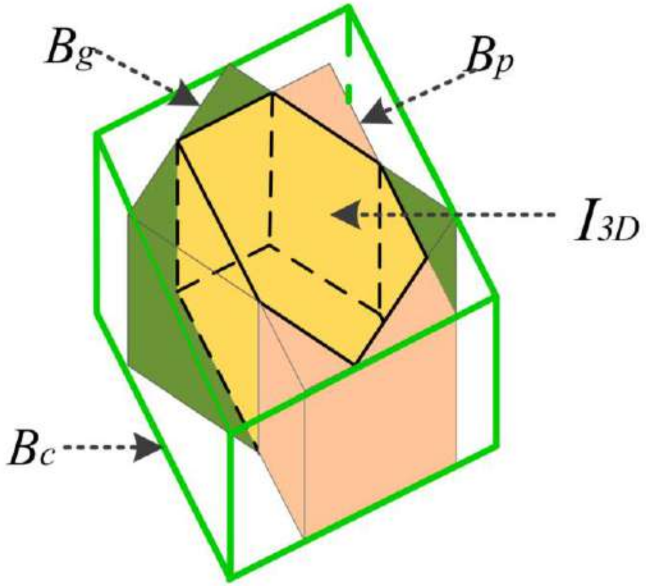
1. A 3D MOT evaluation tool along with three integral metrics
2. A strong and simple 3D MOT system with the fastest speed (207.4 FPS)

What are the Issues of 3D MOT Evaluation?

- Matching criteria: IoU (intersection of union)
- For the pioneering 3D MOT dataset KITTI, evaluation is performed in the 2D space
  - IoU is computed on the 2D image plane (not 3D)
- The common practice for evaluating 3D MOT methods is:
  - Project 3D trajectories onto the image plane
  - Run the 2D evaluation code provided by KITTI



IoU in 2D space



IoU in 3D space

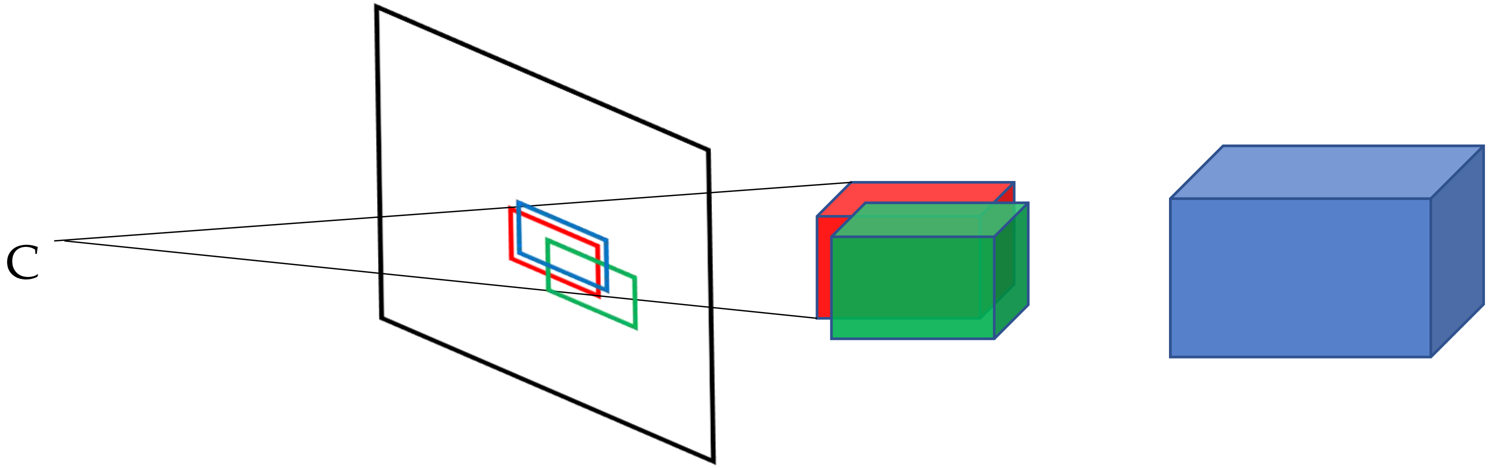
B_p: the predicted box
B_g: the ground truth box
B_c: the smallest enclosing box
I_{2D}, I_{3D}: the intersection

Image credit to Xu et al: 3D-GIoU

What are the Issues of 3D MOT Evaluation?

- Why is it not good to evaluate 3D MOT methods in the 2D space?

- Cannot measure the strength of 3D MOT methods
  - Estimated 3D information: depth value, object dimensionality (length, height and width), heading orientation
- Cannot fairly compare 3D MOT methods, why?
  - Not penalized by the wrong predicted depth value, length, heading as long as the 2D projection is accurate
  - Which predicted box is better, blue or green?
  - Conclusion: should not evaluate 3D MOT methods in the 2D space



Blue: the predicted box 1
Green: the predicted box 2
Red: the ground truth box

Our Solution: Upgrade the Matching Criteria to 3D

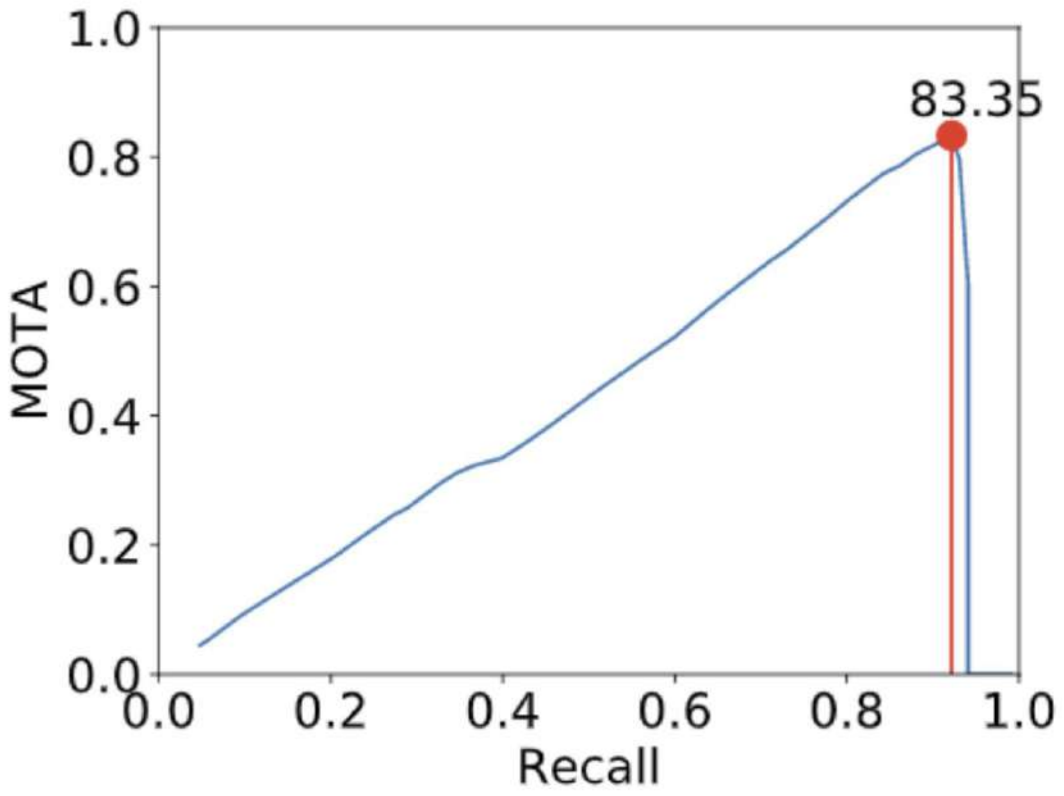
- Replace the matching criteria (2D IoU) in the KITTI evaluation code with 3D IoU
  - https://github.com/xinshuoweng/AB3DMOT (**800+ stars**)
- Work with nuTonomy collaborators and use our 3D MOT evaluation metrics in the nuScenes evaluation with the matching criteria of center distance
  - https://www.nuscenes.org/

Our released new evaluation code

nuScenes 3D MOT evaluation with our metrics

What are the Issues of Evaluation?

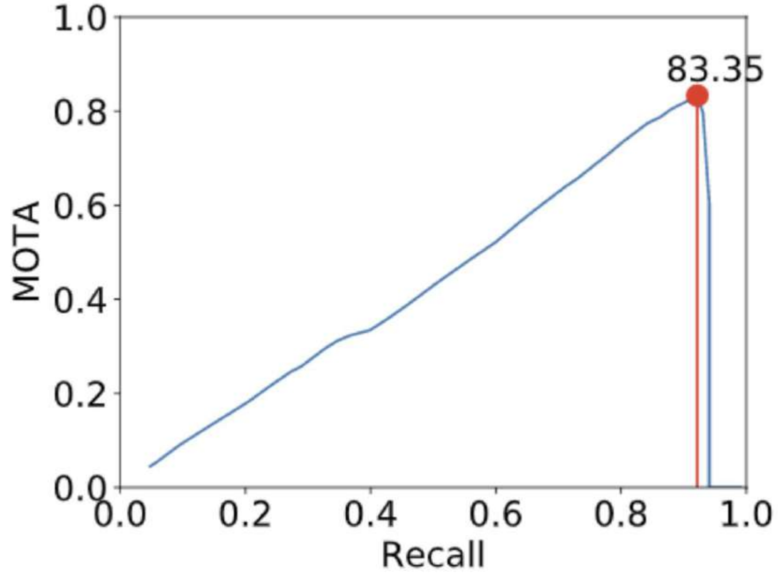
- Are we done with the evaluation? Can we further improve the current metrics?
  - E.g., MOTA (multi-object tracking accuracy)
  - $MOTA = 1 - \frac{FP + FN + IDS}{num_{gt}}$
- Performance is measured at a single recall point



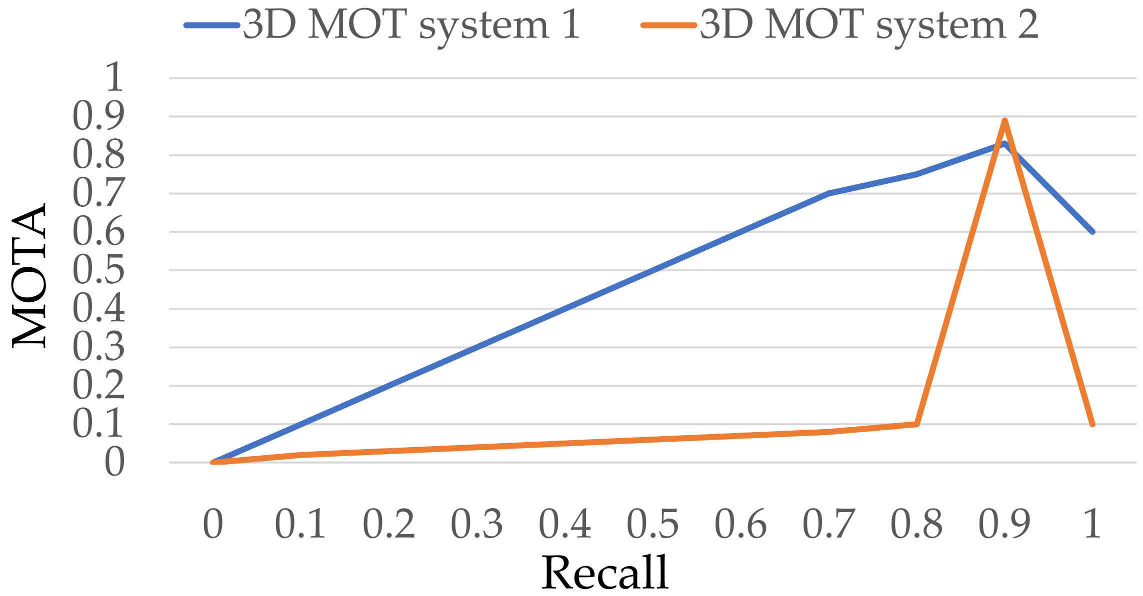
MOTA over Recall curve

What are the Issues of Evaluation?

- Why is it not good to evaluate at a single recall point?
- Consequences
  - The confidence threshold needs to be carefully tuned, requiring non-trivial effort
    - Sensitive to different detectors, different dataset, different object categories
  - Cannot understand the full spectrum of accuracy of a MOT system
    - Which MOT system is better, blue or orange?
    - The orange one has higher MOTA at its best recall point (r = 0.9)
    - The blue one has overall higher MOTA at many recall points
    - Ideally, we want as high performance as possible at all recall points

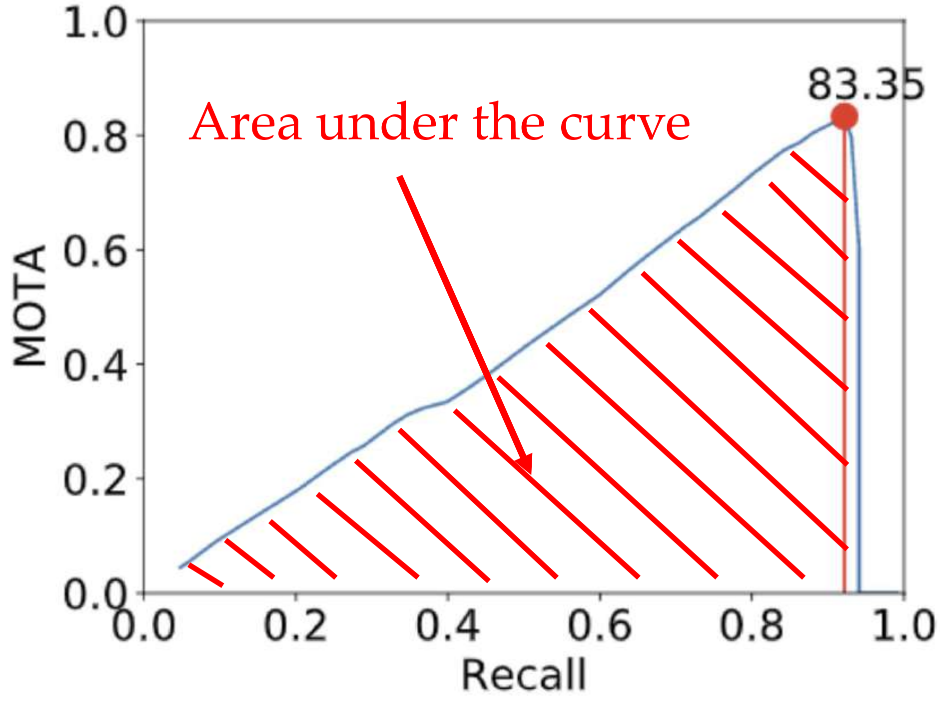


MOTA over Recall curve



Our Solution: Integral Metrics

- MOTA is measured at a single point on the curve
- What can we do to improve the evaluation metrics?
- Compute the integral metrics through the area under the curve, e.g., average MOTA (AMOTA)
  - Analogous to the average precision (AP) in object detection
  - Can measure the full spectrum of MOT accuracy



MOTA over Recall curve

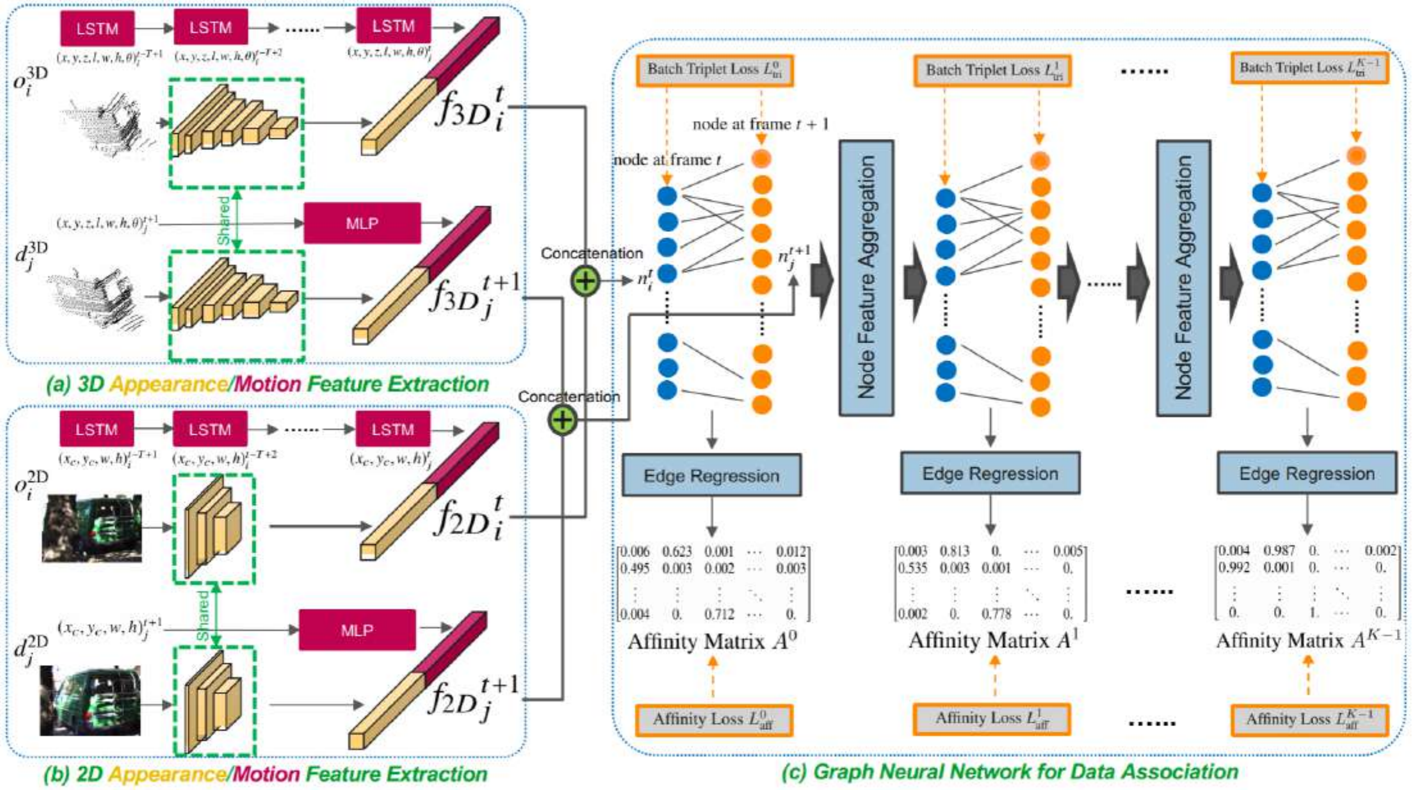
Our Contributions

1. A 3D MOT evaluation tool along with three integral metrics
2. A strong and simple 3D MOT system with the fastest speed (207.4 FPS)

Limitation of Prior Work

- Prior work often ignores practical factors
  - Computational efficiency
  - System complexity

- Consequences
  - Difficult to tell which part contributes the most to performance
  - Not ready to be deployed in time-critical systems

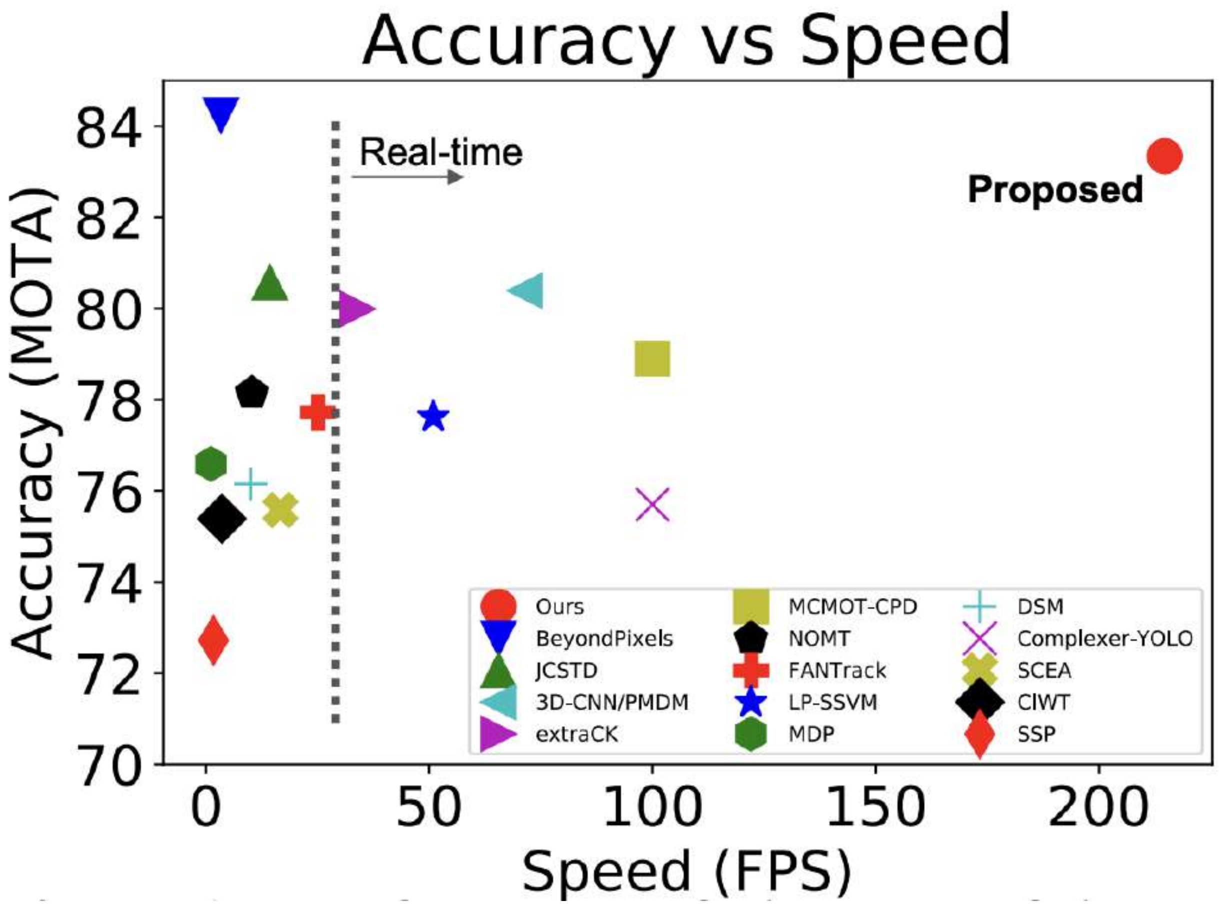


1. A giant neural network for feature extraction
2. Runs at about 5 FPS

Weng et al. GNN3DMOT: Graph Neural Network for 3D Multi-Object Tracking with 2D-3D Multi-Feature Learning. CVPR 2020

AB3DMOT: A Baseline for 3D Multi-Object Tracking

- Motivation
  - Reduce system complexity of 3D MOT methods
  - Increase the computational efficiency (i.e., run time speed)
- Simple design: 3D Kalman filter + Hungarian algorithm
  - 3D Kalman filter
    - Extension of standard 2D Kalman filter
    - Add object's 3D property into the state space
- High speed:
  - 207.4 FPS on the KITTI dataset for Cars
  - 470.1 FPS on the KITTI dataset for Pedestrians
  - 1241.6 FPS on the KITTI dataset for Cyclists
- Strong 3D MOT performance competitive to more complicated systems



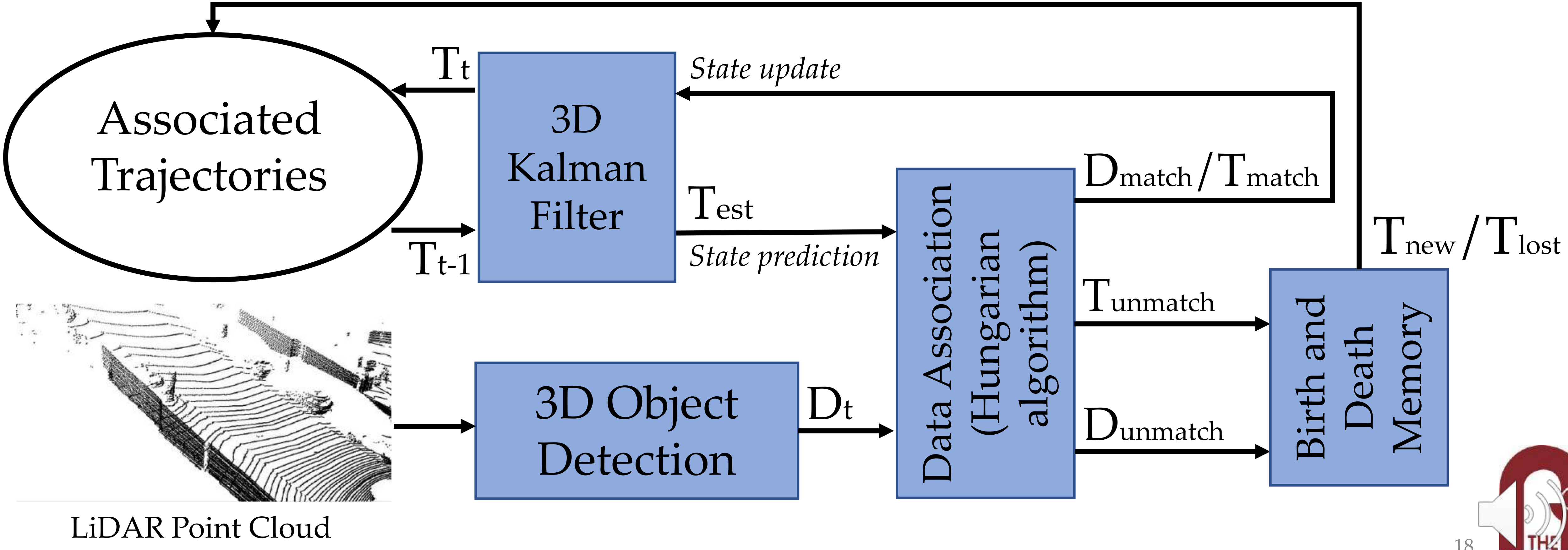
KITTI MOT leaderboard by end of 2019

AB3DMOT: A Baseline for 3D Multi-Object Tracking

- System pipeline (5 modules)
  - 3D object detection
  - Hungarian algorithm
  - Birth and death memory

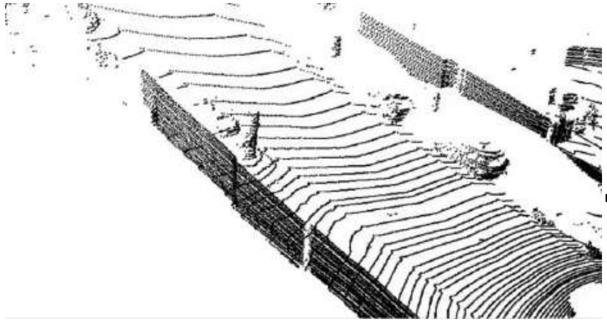
3D Kalman filter: state prediction

3D Kalman filter: state update



AB3DMOT: A Baseline for 3D Multi-Object Tracking

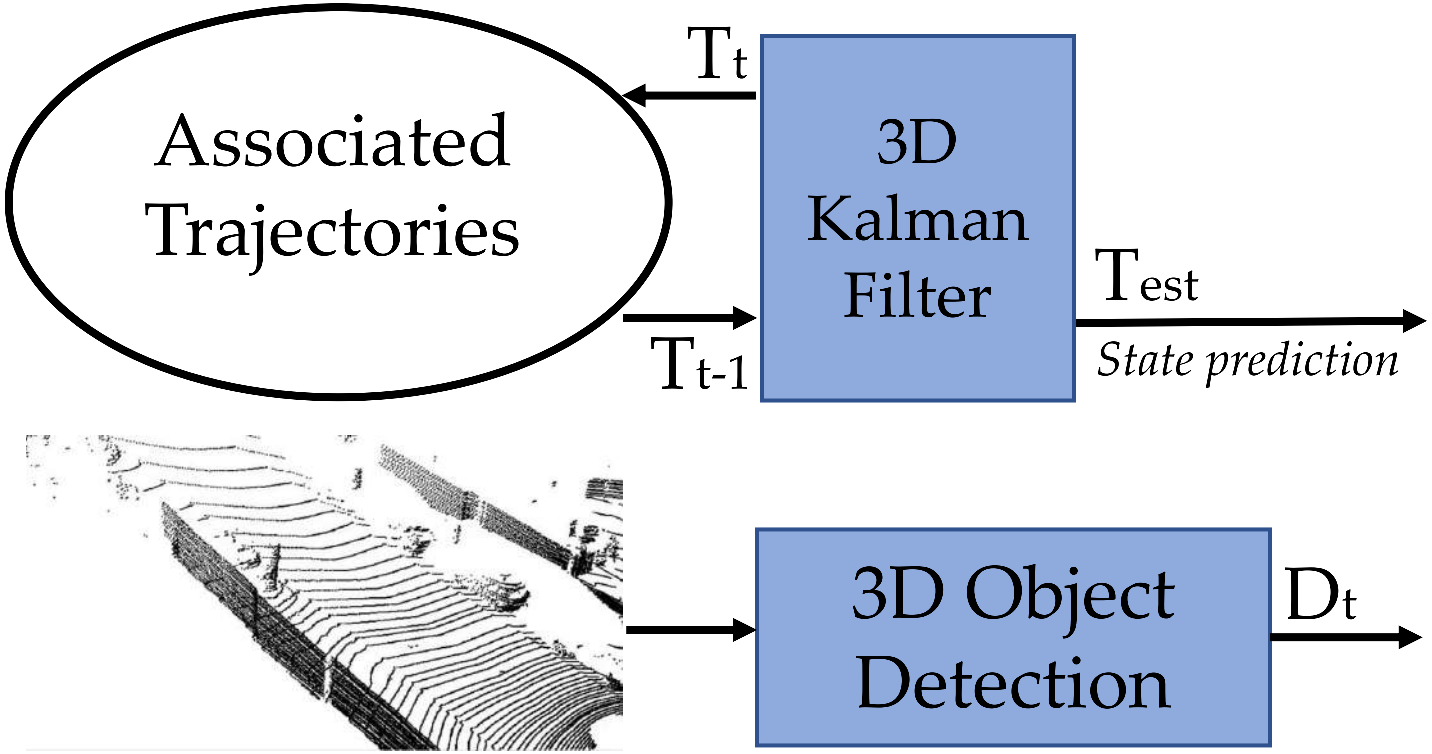
- System pipeline
  - 3D object detection module detects the objects' bounding boxes D_t from the LiDAR point cloud at the current frame t



LiDAR Point Cloud

AB3DMOT: A Baseline for 3D Multi-Object Tracking

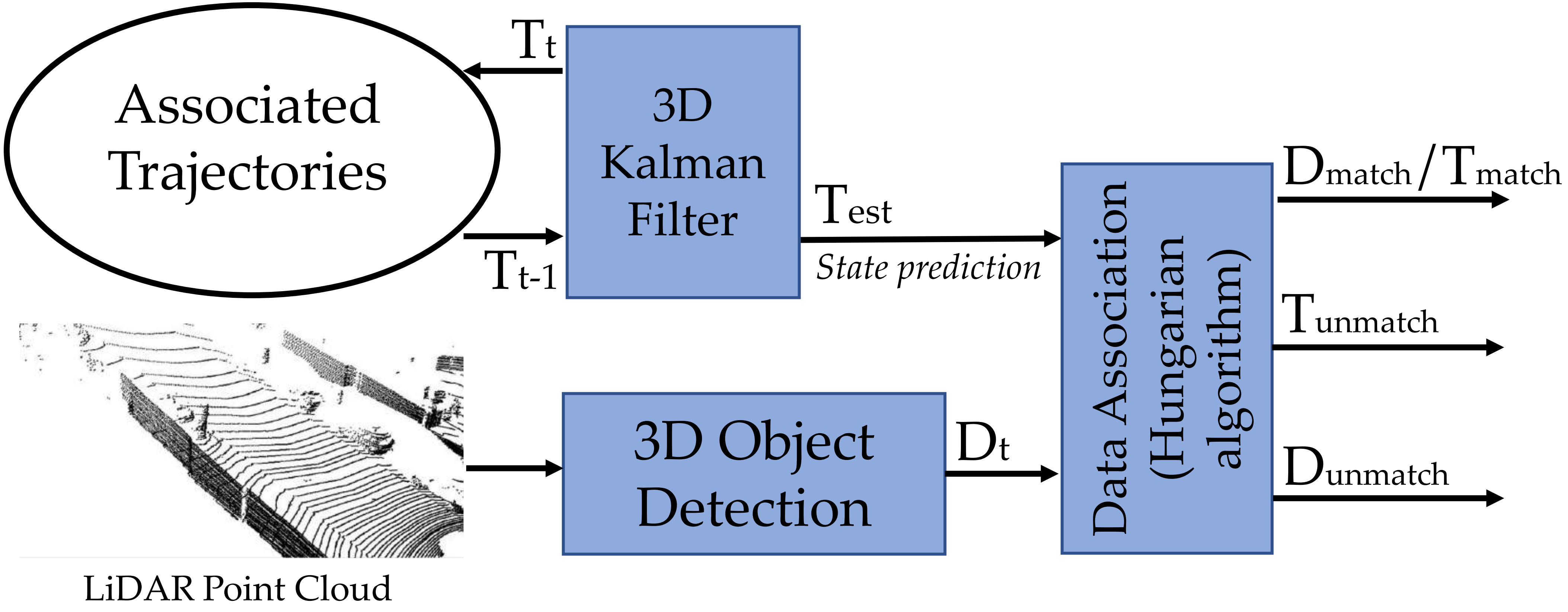
- System pipeline
  - 3D Kalman filter predicts the state of trajectories T_{t-1} in the last frame to the current frame t as T_{est} during the state prediction step



LiDAR Point Cloud

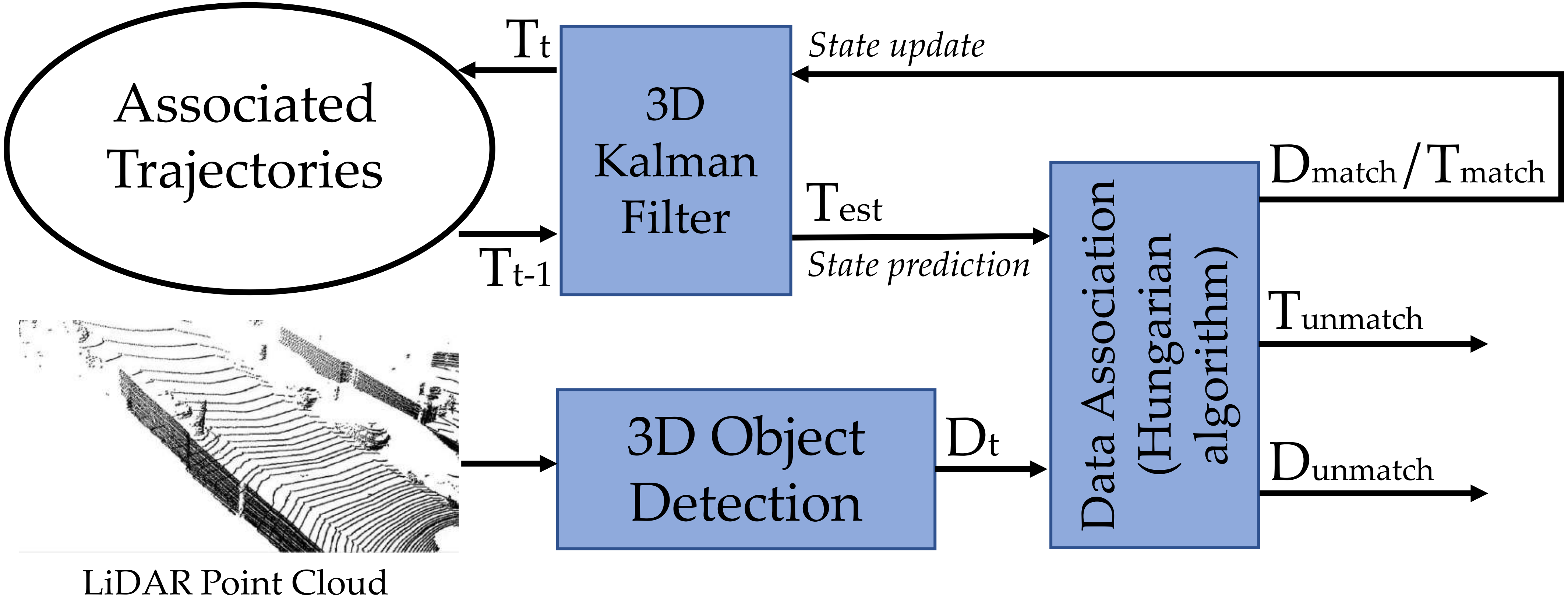
AB3DMOT: A Baseline for 3D Multi-Object Tracking

- System pipeline
  - Detections D_t and trajectories T_{est} are associated using the Hungarian algorithm



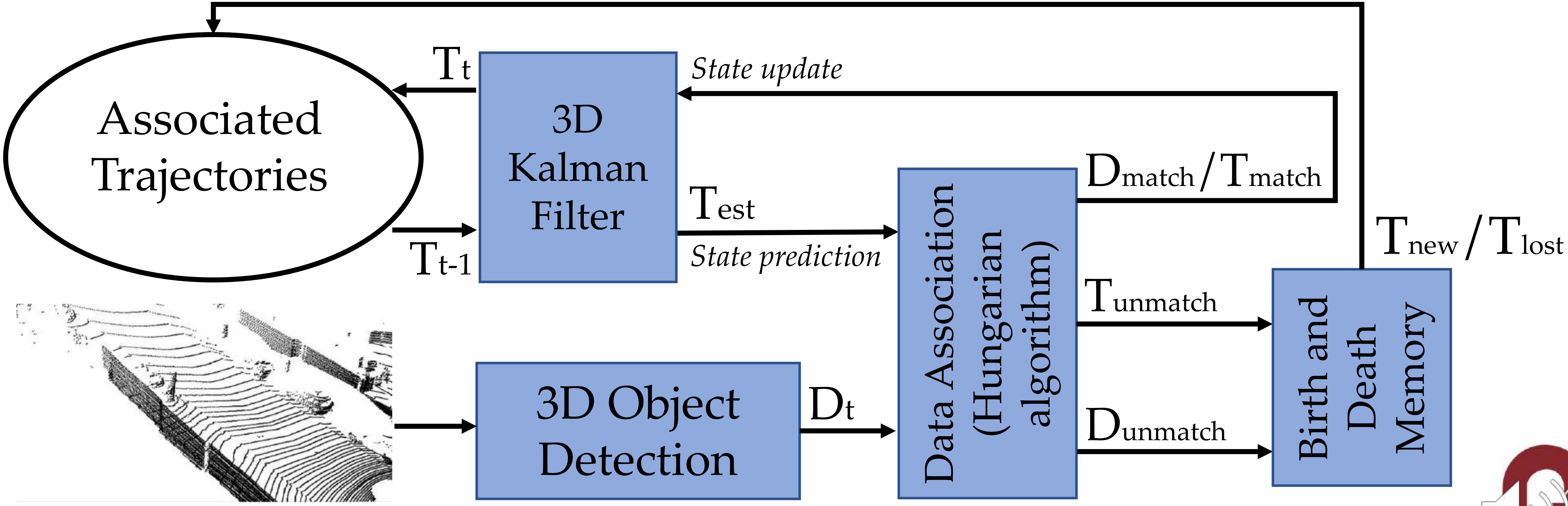
AB3DMOT: A Baseline for 3D Multi-Object Tracking

- System pipeline
  - State of matched trajectories T_{match} is updated based on the corresponding matched detections D_{match} to obtain the final trajectory outputs T_t in the current frame t



AB3DMOT: A Baseline for 3D Multi-Object Tracking

- System pipeline
  - Unmatched detections $D_{unmatch}$ and unmatched trajectories $T_{unmatch}$ are used to create new trajectories T_{new} and delete disappeared trajectories T_{lost}



LiDAR Point Cloud

Quantitative Results

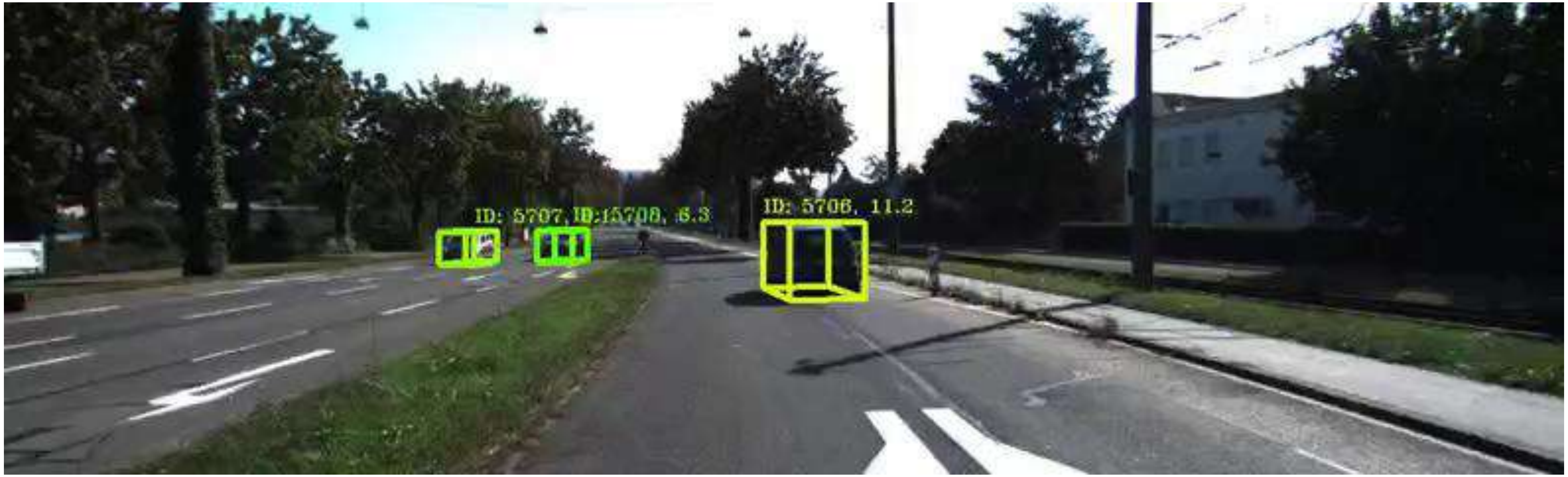
3D MOT Evaluation on KITTI for Cars

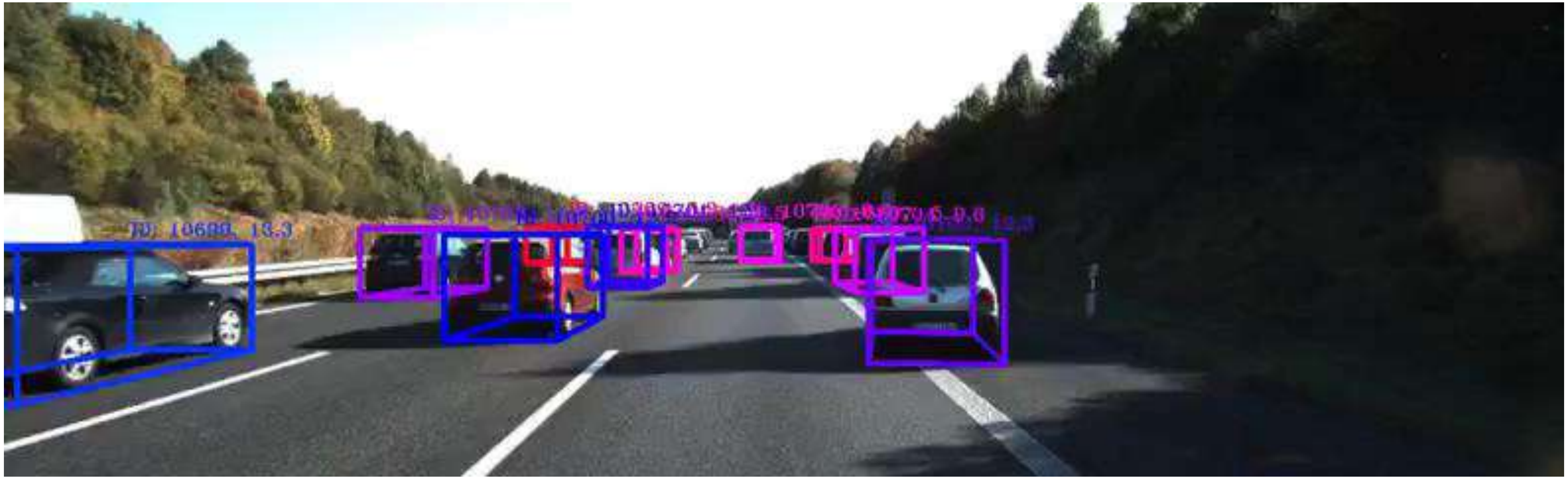
- Our 3D MOT system runs at the fastest speed without the need of a GPU
- Our simple system outperforms two more complicated 3D MOT systems

Method	Input Data	Matching criteria	**sAMOTA**↑	AMOTA↑	AMOTP↑	MOTA↑	MOTP↑	IDS↓	FRAG↓	FPS↑
mmMOT [32] (ICCV′19)	2D + 3D	$IoU_{thres} = 0.25$	70.61	33.08	72.45	74.07	78.16	10	55	4.8 (GPU)
FANTrack [17] (IV′20)	2D + 3D	$IoU_{thres} = 0.25$	82.97	40.03	75.01	74.30	75.24	35	202	25.0 (GPU)
Ours	3D	$IoU_{thres} = 0.25$	93.28	45.43	77.41	86.24	78.43	0	15	207.4 (CPU)

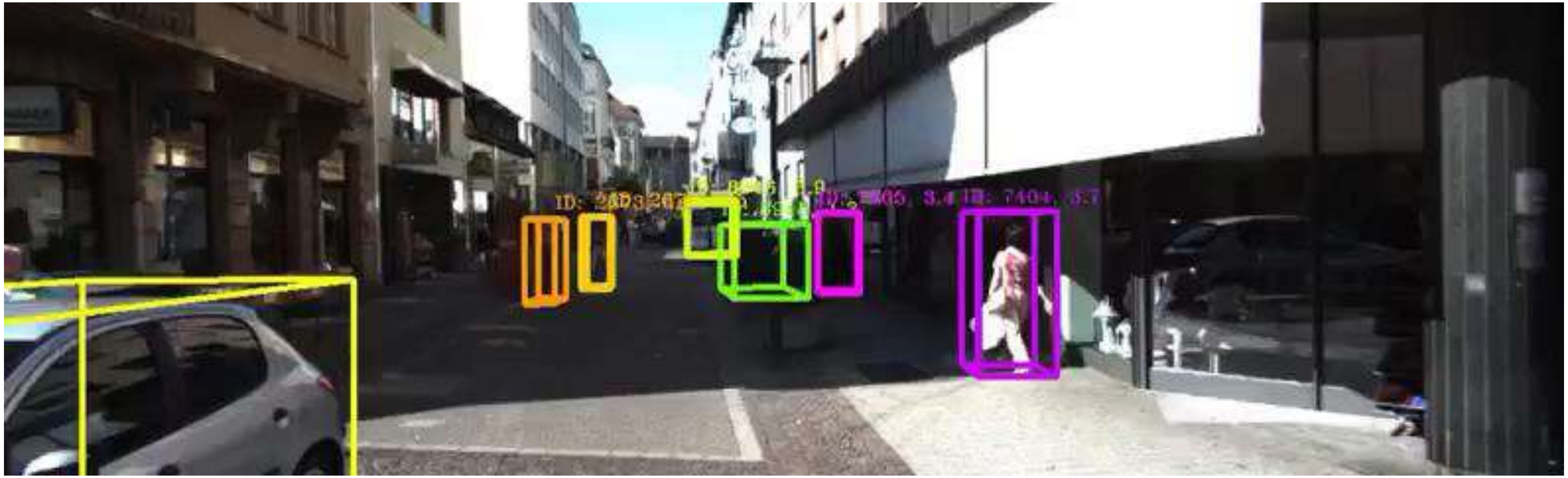
Qualitative Results

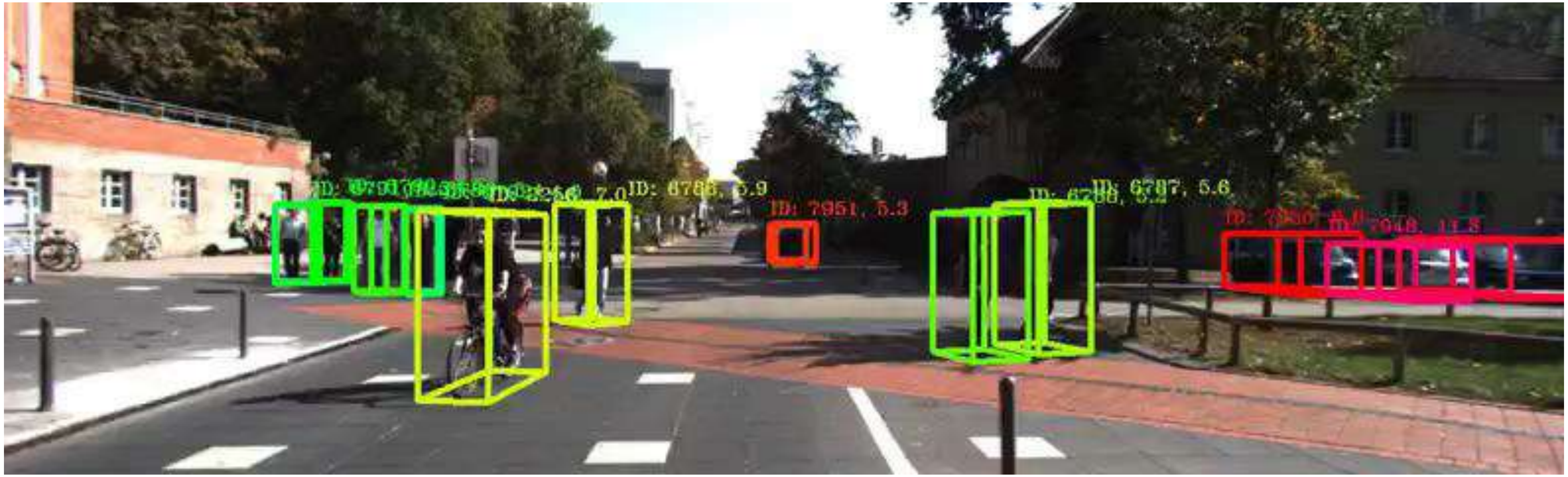
Qualitative Results for Cars





Qualitative Results for Pedestrians / Cyclists





Carnegie Mellon University
The Robotics Institute

3D Multi-Object Tracking: A Baseline and New Evaluation Metrics

Xinshuo Weng, Jianren Wang, David Held, Kris Kitani

Robotics Institute, Carnegie Mellon University

IEEE/RSJ International Conference on Intelligent Robots and Systems (***IROS***), 2020