Immortal Tracker: Tracklet Never Dies

Qitai Wang[1,3] Yuntao Chen[2] Ziqi Pang[4] Naiyan Wang[2] Zhaoxiang Zhang[1,3]
[1] University of Chinese Academy of Sciences [2] Tusimple
[3] Institute of Automation, Chinese Academy of Sciences (CASIA)
[4] UIUC
{wangqitai2020, zhaoxiang.zhang}@ia.ac.cn
{chenyuntao08, winsty}@gmail.com ziqip2@illinois.edu

Abstract

Previous online 3D Multi-Object Tracking(3DMOT) methods terminate a tracklet when it is not associated with new detections for a few frames. But if an object just goes dark, like being temporarily occluded by other objects or simply getting out of FOV, terminating a tracklet prematurely will result in an identity switch. We reveal that ***premature tracklet termination*** *is the main cause of identity switches in modern 3DMOT systems. To address this, we propose Immortal Tracker, a simple tracking system that utilizes trajectory prediction to maintain tracklets for objects gone dark. We employ a simple Kalman filter for trajectory prediction and preserve the tracklet by prediction when the target is not visible. With this method, we can avoid 96% vehicle identity switches resulting from premature tracklet termination. Without any learned parameters, our method achieves a mismatch ratio at the 0.0001 level and competitive MOTA for the vehicle class on the Waymo Open Dataset test set. Our mismatch ratio is tens of times lower than any previously published method. Similar results are reported on nuScenes. We believe the proposed Immortal Tracker can offer a simple yet powerful solution for pushing the limit of 3DMOT. Our code is available at https://github.com/ImmortalTracker/ImmortalTracker.*

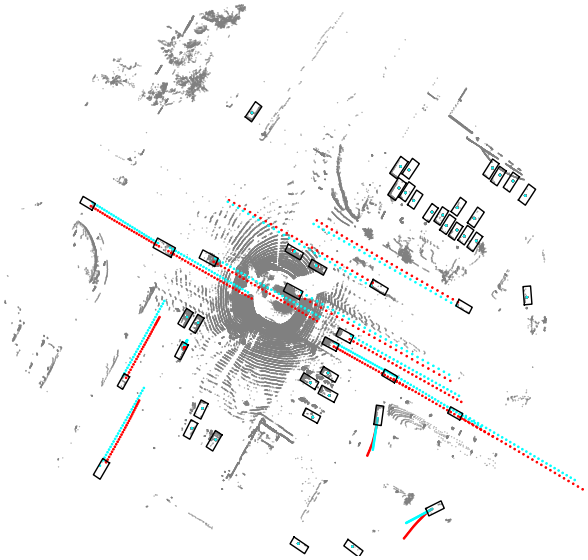


Figure 1. Bird's Eye View(BEV) tracking visualization of vehicles with their predicted trajectories. For each tracked vehicle, we plot their bounding boxes in the initial frame and predicted(marked in cyan) or ground-truth(marked in red) center locations for the next 30 frames. Small displacements are added to the overlapped trajectories for better visualization.

1. Introduction

Most state-of-the-art methods in 3D Multiple Object Tracking(3DMOT) follow the tracking-by-detection paradigm. They first obtain bounding boxes of objects in each frame with 3D detectors and then associate the detected objects between frames as tracklets. In the online tracking process, the same object is associated sequentially across frames.

Ideally, a tracklet should be maintained until its target leaves the scene. Previous works [1, 6, 11, 17, 24, 25, 27, 28] in 3DMOT usually adopt a simple delay-based mechanism for tracklet life management. They maintain a tracklet that loses its target for a few frames. If the tracklet fails to be associated with any detections during this period, it will be terminated and no longer participate in the detection-tracklet association in the future frames. This mechanism assumes that an object not detected for several frames is considered to have left the scene. However, objects go dark for different reasons. When an object is occluded or temporarily being out of FOV, it might be missing for frames as well and its corresponding tracklet will be terminated pre-

Method	IDS	Early Termination	Wrong Assosication
CenterPoint	2891	2890	1
Immortal Tracker(Ours)	114	113	1

Table 1. Statics on the vehicle identity switch cases reported by CenterPoint and Immortal Tracker on the validation set of WOD. We divide the identity switches into Early Termination(the ground truth trajectory of an object breaks into two tracklet) and Wrong Association(two tracklets belonging to different objects are confused during the tracking process).

maturely. In such a case if the object is detected again in the future, a new tracklet will be initialized, causing an identity switch. As shown in Table 1, 99.9% vehicle identity switches reported by the CenterPoint tracker on the validation set of Waymo Open Dataset (WOD) are the results of premature tracklet termination.

To overcome premature tracklet termination, we present Immortal Tracker, a simple tracking framework based on trajectory prediction for invisible objects. Instead of terminating an unassociated tracklet, we maintain it with its predicted trajectory. Therefore when a temporarily invisible object reappears on the predicted trajectory, it can be assigned to its original tracklet. We use a vanilla 3D Kalman filter(3DKF) [10] for trajectory prediction to highlight the simplicity and effectiveness of our predict-to-track paradigm. As shown in Figure 1, our simple 3DKF works surprisingly well for trajectory prediction on WOD.

Experiments show that our Immortal Tracker can significantly reduce identity switches when applied on the top of a wide range of off-the-shelf 3D detectors. Built upon the detection results from CenterPoint [27], we establish new SOTA results on 3DMOT benchmarks like Waymo Open Dataset and nuScenes. Our method achieves 60.6 level 2 MOTA for the vehicle class on Waymo Open Dataset. We also bring the mismatch ratio down to the 0.0001 level, tens of times lower than previous SOTA methods. On nuScenes, our method achieves a 66.1 AMOTA, outperforming all previous published LiDAR-based methods.

2. Related works

3D MOT. Following the tracking-by-detection paradigm, previous works in 3D multi-object tracking usually solve the tracking problem in the form of a bipartite graph matching process on top of off-the-shelf detectors. Inspired by early works in 2D MOT [7, 15, 16], various methods focus on strengthening the association between detections and tracklets by modeling their motions or appearance, or the combination of these both. [20] defines the state of Kalman Filter in 2D plane. AB3DMOT [24] presents a baseline method through combining 3D Kalman filter and the Hungarian algorithm [12] on top of PointR-CNN detector [22]. AB3DMOT employs 3D Intersection of Union (3D IoU) as its association metric, while Chiu et al [5] used Mahalabnobis distance [18] instead. SimpleTrack [19] generalized GIoU [21] to 3D for association. CenterPoint [27] learns to predict two-dimensional velocity of detected box following CenterTrack [29] and perform simple point-distance matching.

To further reduce confusion during the association process, GNN3DMOT [25] jointly extracts appearance and motion features with a Graph Neural Network to introduce feature interactions between objects. [5] proposes a probabilistic multi-modal system that contains trainable modules for 2D-3D appearance feature fusion, distance combination, and trajectory initialization. [11] combines 2D and 3D object evidence obtained from 2D and 3D detectors. Authors of [28] merge predictive models and objects detection features in a unified graph representation. In this paper, we use simple 3D IoU/GIoU metrics for data association.

Following early experience in 2D MOT [9, 15, 16], previous works [6, 11, 24, 25, 27, 28] in 3D MOT usually adopt count-based method for tracklet life-cycle management. For each frame, new tracklets are initialized with detections that are not associated with existing tracklets. Tracklets that lost their targets for serveral(typically less than 5) frames will be terminated. [1, 17] proposed to initialize and terminate tracks depending on their confidence score estimated from the confidence of their associated detections. However, they will still permanently terminate the tracklets that fail to be associated with new detections. In contrast, we show that through positively predicting and preserving trajectories of objects, the tracklets which lost their targets can be properly maintained for possible association in the future.

3. Method

Figure 2 shows the overview of our system. As an online tracking method, our method takes sequential detection results as input. For each frame, we predict the object locations for all tracklets with 3DKF. Then we perform Hungarian algorithm for bipartite matching between detections and predictions. The bipartite matching process will output the matched pairs and the unmatched tracklets or detections. The states of matched tracklets are updated by their corresponding detections, and unmatched tracklets are updated with their predicted object states. The unmatched detections will be initialized as new tracklets. Each part in the pipeline will be described in detail in the following sections.

Detection. By design, our Immortal Tracker is agnostic to the choice of detector. For our best-reported result, We employ CenterPoint [27] as our off-the-shelf detector. CenterPoint chooses a quite permissive IoU threshold for NMS to ensure a better recall. Therefore, we follow Sim-

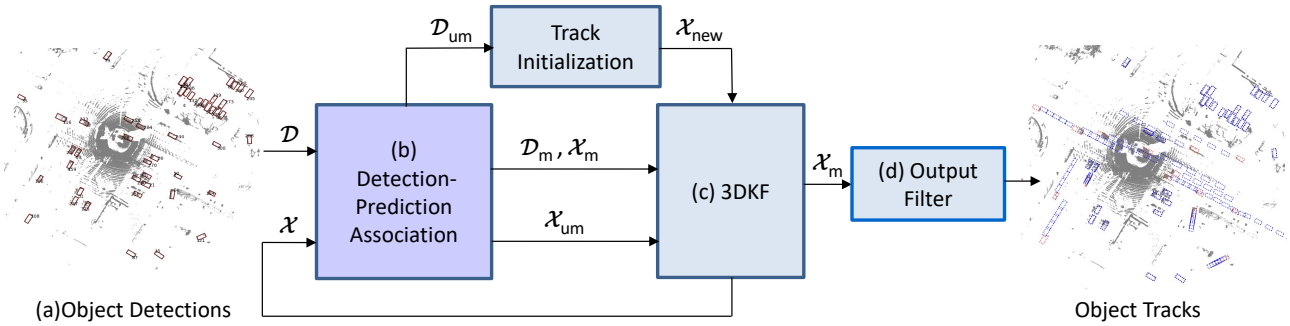


Figure 2. **Algorithm Pipeline.** (a)We employ an off-the-shelf 3D detector to obtain 3D detections $\mathcal{D}$ from input point clouds. (b)For data association, we compute 3D IoU or GIoU between detected 3D bounding boxes and boxes predicted by Immortal Tracker. Then we perform Hungarian matching based on 3D IoU/GIoU. (c)We use a vanilla 3D Kalman filter(3DKF) for prediction. Based on the outputs of Hungarian matching process, The states of matched tracklets $\mathcal{X}_{\text{m}}$ are updated by its corresponding detection $\mathcal{D}_{\text{m}}$. The unmatched tracklets $\mathcal{X}_{\text{um}}$ will be updated with their predicted object states and the unmatched detections $\mathcal{D}_{\text{um}}$ will be initialized as new tracklets $\mathcal{X}_{\text{new}}$. (d)Only the alive tracklets which have left their birth stage and successfully been matched in the current frame are adopted as outputs.

pleTrack [19] and perform a much stricter NMS on detection results to remove unwanted boxes before feeding them to our Immortal Tracker.

Trajectory Prediction. We use a vanilla 3D Kalman filter(3DKF) for trajectory prediction. Following AB3DMOT [24], we define the states of Kalman filter $Z = [x, y, z, \theta, l, w, h, \dot{x}, \dot{y}, \dot{z}]$ in 3D space, including object position, box size, box orientation and velocity. Different from AB3DMOT, our object positions are defined in the world frame like [27], instead of in the LiDAR frame. The change of reference frame is crucial to our Immortal Tracker, helping us reduce the total number of tracklets managed. During the tracking process, we alternate between predicting object state X with 3DKF and updating 3DKF state Z with incoming detection D. Both X and D are 7-dim vectors consisting of $[x, y, z, \theta, l, w, h]$. Note that we use X to denote both a tracklet and the latest prediction of its associated 3DKF. We may refer X as tracklet or prediction interchangeable hereafter.

Detection-Prediction Association. For association, we compute 3D IoU or GIoU [19, 21] between detected 3D bounding boxes and boxes predicted by 3DKF. We then perform Hungarian matching based on 3D IoU / GIoU. We discard low quality matchings if the 3D IoU / GIoU of a matching is lower than $\text{IoU}_{\text{thres}}$ or $\text{GIoU}_{\text{thres}}$. The inputs of the bipartite matching process is defined as:

$$\mathcal{D} = \{D^1, D^2, \cdots, D^p\} \tag{1}$$

$$\mathcal{X} = \{X^1, X^2, \cdots, X^q\} \tag{2}$$

where p and q are the number of detections and tracklets in the frame.

The output of the bipartite matching process is defined as:

$$\mathcal{D}_{\text{m}} = \{D^1_{\text{m}}, D^2_{\text{m}}, \cdots, D^k_{\text{m}}\} \tag{3}$$

$$\mathcal{D}_{\text{um}} = \{D^1_{\text{um}}, D^2_{\text{um}}, \cdots, D^{p-k}_{\text{um}}\} \tag{4}$$

$$\mathcal{X}_{\text{m}} = \{X^1_{\text{m}}, X^2_{\text{m}}, \cdots, X^{k_t}_{\text{m}}\} \tag{5}$$

$$\mathcal{X}_{\text{um}} = \{X^1_{\text{um}}, X^2_{\text{um}}, \cdots, X^{q-k}_{\text{um}}\} \tag{6}$$

where $\mathcal{X}_{\text{m}}$ and $\mathcal{D}_{\text{m}}$ are the k matched pairs of predictions and detections. $\mathcal{X}_{\text{um}}$ are the unmatched tracklets and $\mathcal{D}_{\text{um}}$ are the unmatched detections.

The matched detections $\mathcal{D}_{\text{m}}$ will be used to update the 3DKF states of corresponding tracklets $\mathcal{X}_{\text{m}}$. The unmatched tracklets $\mathcal{X}_{\text{um}}$ will use their own predictions to update 3DKF.

Tracklet Birth and Preservation. Following the common tracklet initialization strategy, we initialize a set of tracklets $\mathcal{X}_{\text{new}}$ for detections in $\mathcal{D}_{\text{um}}$. For these new tracklets, we will mark them as *alive* if they are successfully associated with detections for M_{hits} times in future frames. Otherwise, they stay at the *birth* stage.

Different from previous works [1, 6, 11, 17, 24, 25, 27, 28] which terminate a tracklet after several frames since its last successful association, our Immortal Tracker always maintains tracklets for objects even if they are invisible. The maintained tracklets will repeatedly predict their locations in the future frames with their last estimated states. In this way, we maintain tracklets for all objects observed at least once. Our tracklets never die once created and hence are **immortal**. And when the missing object reappears near its predicted trajectory, we can associate it with its corresponding tracklet again and update its tracklet with the observation instead of prediction.

In each frame, we output tracklets in $\mathcal{O} = \{X_i \in$

$\mathcal{X}_{\text{m}} \cap \mathcal{X}_{\text{alive}}\}$, where $\mathcal{X}_{\text{alive}}$ are tracklets in the alive state. Therefore the predictions of unmatched tracklets will not cause false positives. Algorithm 1 summarizes our simple tracking algorithm.

Algorithm 1 Tracking Pipeline

Input: 3D Object Detections $\mathcal{D}_{0:t}$
Output: Output Tracks $\mathcal{O}_{0:t}$
Frame number $t \leftarrow 1$
Tracks set $\mathcal{X}_0 \leftarrow \{\}$
while not the end of tracking **do**
$\mathcal{X}_t \leftarrow \text{Prediction}(\mathcal{X}_{t-1})$
$\mathcal{D}_{\text{m}}, \mathcal{X}_{\text{m}}, \mathcal{D}_{\text{um}}, \mathcal{X}_{\text{um}} \leftarrow \text{Match}(\mathcal{D}_t, \mathcal{X}_t)$
$\mathcal{X}_{\text{m}} \leftarrow \text{Update}(\mathcal{D}_{\text{m}})$
$\mathcal{X}_{\text{new}} \leftarrow \text{Initialize}(\mathcal{D}_{\text{um}})$
$\mathcal{O}_t \leftarrow \{X_i \in \mathcal{X}_{\text{m}} \cap \mathcal{X}_{\text{alive}}\}$
$\mathcal{X}_{t+1} \leftarrow \mathcal{X}_{\text{m}} \cup \mathcal{X}_{\text{um}} \cup \mathcal{X}_{\text{new}}$
$t \leftarrow t + 1$
end while

4. Experiments

4.1. Datasets and Metrics.

We evaluate our method on Waymo Open Dataset(WOD) and nuScenes.

Waymo Open Dataset [17] contains 1000 driving video sequences each lasts 20 seconds at 10 FPS. 798 / 202 / 150 sequences are used for training, validation and testing, respectively. Point clouds and ground truth 3D boxes of objects in vehicle, pedestrian and cyclist classes are provided for each frame. Following the official evaluation metrics specified in [17], we report Multiple Object Tracking Accuracy (MOTA) [2], False Positives(FP), Miss and Mismatch for objects in the L2 difficulty. Readers may refer to [2] for detailed descriptions about these metrics.

nuScenes [3] contains 1000 driving sequences in total with LiDAR scans and ground truth 3D box annotations provided at 20 FPS and 2 FPS, respectively. We report AMOTA [23], MOTA, and identity switches(IDS) for nuScenes. AMOTA is computed by integrating MOTA over different recalls, which is used as the primary metric for evaluating 3DMOT on nuScenes.

4.2. Baselines

CenterPoint [27] is both a 3D detection method and a 3DMOT method. Here we refer them as CenterPoint-DET and CenterPoint-3DMOT, respectively. We compare our proposed method with CenterPoint-3DMOT since we use the same CenterPoint-DET detection results. Also, CenterPoint-3DMOT serves as a very strong baseline on both WOD and nuScenes.

CenterPoint++ is our modified tracking implementation of CenterPoint. CenterPoint-3DMOT only performs constant velocity trajectory prediction and nearest box center matching. It is unfair to compare it with our method which uses 3DKF for trajectory prediction and 3D IoU / GIoU for matching. Therefore we set another baseline named CenterPoint++, which uses 3DKF and 3D IoU / GIoU matching like ours. The main difference between CenterPoint++ and Immortal Tracker is that a traditional tracklet termination mechanism is used in CenterPoint++. We terminate a tracklet in CenterPoint++ after A_{max} frames since its last successful association following previous works [6,11,24,25,27,28]. All the other hyper-parameters of CenterPoint++ are set the same as our Immortal Tracker for a fair comparison.

4.3. Implementation Details

We take CenterPoint detection boxes as input. We drop boxes with scores below 0.5 on WOD. We keep all boxes on nuScenes. We also perform 3D IoU-based NMS on detected bounding boxes with 0.25 / 0.1 3D IoU thresholds on WOD / nuScences, respectively. In detection-prediction association, we use 3D IoU / GIoU between detections and predictions on WOD/nuScenes, respectively. Then we perform Hungarian algorithm over the 3D IoU distance matrix. We drop all detection-prediction associations with 3D IoU below 0.1 or 3D GIoU below -0.5. For tracklet initialization, we set the hits to birth $M_{\text{hits}} = 1$ on both WOD and nuScenes. For CenterPoint++, we set $A_{\text{max}} = 2$.

4.4. Main Results

We compare Immortal Tracker to previous 3DMOT methods on both WOD and nuScenes.

Results on WOD. Table 2 and 3 show comparisons of tracking results between our method, our baselines and other state-of-the-art methods on the WOD test and validation sets. For the vehicle class, our method reduces the mismatch ratio to a new level of 0.0001, several times lower than previous works. It is noteworthy that our method performs much better than CenterPoint++ in the mismatch metric. This indicates the majority of mismatch cases we have reduced are results of maintaining the tracklets with predicted trajectories rather than employing 3DKF as motion model. Besides the substantial improvement our proposed method achieved on the mismatch ratio, we also achieve a 0.4/0.9 MOTA improvement over CenterPoint++ for vehicle/pedestrian, respectively, outperforming all previously published methods.

Results on nuScenes. Table 4 and 5 shows our results on nuScenes test and validation sets. We compare our Immortal Tracker with published LiDAR-only works for a fair comparison. All methods in Table 4 and 5 are based on the same CenterPoint detection results, allowing a fair comparison between different tracking approaches. Our method achieves a 2.3 AMOTA improvement over CenterPoint and

Method	Vehicle				Pedestrian			
	MOTA%↑	FP%↓	Miss%↓	Mismatch%↓	MOTA%↑	FP%↓	Miss% ↓	Mismatch%↓
AB3DMOT [17]	40.1	16.4	43.4	0.13	37.7	11.6	50.2	0.47
CenterPoint [27]	59.4	9.4	**30.9**	0.32	56.6	**9.3**	33.1	1.07
SimpleTrack* [19]	60.3	8.8	**30.9**	0.08	60.1	10.7	28.8	0.40
CenterPoint++	60.2	**8.4**	31.2	0.24	59.7	10.3	29.2	0.85
Immortal Tracker(Ours)	**60.6**	8.5	31.0	**0.01**	**60.6**	11.0	**28.3**	**0.18**

Table 2. State-of-the-art comparisons for 3D MOT on WOD test set. * represents for works that are not peer-reviewed yet

Method	Vehicle				Pedestrian			
	MOTA%↑	FP%↓	Miss%↓	Mismatch%↓	MOTA%↑	FP%↓	Miss% ↓	Mismatch%↓
CenterPoint [27]	55.1	10.8	33.9	0.26	54.9	**10.0**	34.0	1.13
SimpleTrack* [19]	56.1	10.4	**33.4**	0.08	57.8	10.9	30.9	0.42
CenterPoint++	56.1	**10.2**	33.5	0.25	57.4	11.1	30.6	0.94
Immortal Tracker(Ours)	**56.4**	**10.2**	**33.4**	**0.01**	**58.2**	11.3	**30.5**	**0.26**

Table 3. State-of-the-art comparisons for 3D MOT on WOD validation set. * represents for works that are not peer-reviewed yet

Method	AMOTA%↑	MOTA%↑	IDS↓
StanfordIPRL-TRI [6]	55.0	45.9	950
CenterPoint [27]	63.8	53.7	760
CBMOT [1]	64.9	54.4	557
OGR3MOT* [28]	65.6	55.4	**288**
CenterPoint++	65.8	**55.7**	609
Immortal Tracker(Ours)	**66.1**	55.2	365

Table 4. State-of-the-art comparisons for 3D MOT on nuScenes test set. Only published online methods which are Lidar-based are reported. * represents for works that are not peer-reviewed yet.

Method	AMOTA%↑	MOTA%↑	IDS↓
CenterPoint [27]	66.5	56.2	562
CBMOT [1]	67.5	58.3	494
OGR3MOT* [28]	69.3	**60.2**	**262**
CenterPoint++	68.7	59.2	519
Immortal Tracker(Ours)	**70.2**	60.1	385

Table 5. State-of-the-art comparisons for 3D MOT on nuScenes validation set. Only published online methods which are Lidar-based are reported. * represents for works that are not peer-reviewed yet.

a 0.3 improvement over CenterPoint++.

4.5. Ablation Studies

We analyze design choices and hyper-parameter settings of Immortal Tracker on the WOD validation set for the vehicle class. Results show that our Immortal Tracker is robust to a wide range of design choices and hyper-parameters.

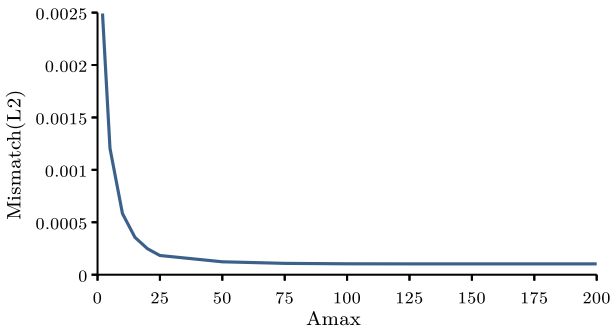


Figure 3. Details for mismatch ratio vs $A_{\max}$ on WOD validation set.

The performance of base 3D detectors. Table 6 summarizes the performances of our proposed method on top of different 3D detectors on the WOD validation set. We take detection results of several 3D detectors including PointPillars+PPBA [4, 13] , SECOND [26], LiDAR R-CNN [14] and CenterPoint that take voxelized or Bird's Eye View (BEV) representations of point clouds for object detection, and RangeDet [8] which takes range view representation of collected 3D informations as input. The results show that our method is agnostic to 3D detectors and can reduce the mismatch rates to 0.0001 level on top of various detectors.

When to terminate a tracklet. Immortal Tracker permanently maintains tracklets for objects gone dark, which is the main difference between Immortal Tracker and CenterPoint++. For ablation experiment, here we introduce the trajectory prediction as well as tracklets preservation mech-

Detector	MOTA%↑	FP(%)↓	Miss(%)↓	Mismatch(%)↓
PointPillars+PPBA [4, 13]++	40.9	9.3	49.5	0.34
PointPillars+PPBA + Immortal Tracker	41.3	9.4	49.3	0.01
SECOND [26]++	50.2	8.6	40.9	0.31
SECOND + Immortal Tracker	50.6	8.7	40.8	0.01
RangeDet [8]++	50.5	10.2	39.1	0.25
RangeDet + Immortal Tracker	50.7	10.2	39.1	0.02
Lidar R-CNN [14]++	53.7	7.3	38.1	0.84
Lidar R-CNN + Immortal Tracker	54.9	8.2	36.9	0.02
CenterPoint++	56.1	10.2	33.5	0.25
CenterPoint + Immortal Tracker	56.4	10.2	33.4	0.01

Table 6. Ablation study for different detectors on WOD validation set. Here ++ represents for we apply the same baseline tracker described in CenterPoint++ on top of the detector.

M_{hits}	MOTA%↑	FP%↓	Miss%↓	Mismatch%↓
0	56.4	10.3	33.4	0.01
1	56.4	10.2	33,4	0.01
2	56.4	10.2	33.4	0.01
3	56.3	10.2	33.5	0.01

Table 7. Ablation study for the minimum birth count M_{hits} on WOD validation set.

$\text{IoU}_{\text{thres}}$	MOTA%↑	FP%↓	Miss%↓	Mismatch%↓
0.7	51.5	**9.6**	38.1	0.86
0.5	54.7	10.0	34.9	0.39
0.3	56.2	10.2	33.5	0.08
0.1	**56.4**	10.2	**33.4**	**0.01**

Table 8. Ablation study for the association threshold $\text{IoU}_{\text{thres}}$ on WOD validation set.

$\text{NMS}_{\text{thres}}$	MOTA%↑	FP%↓	Miss%↓	Mismatch%↓
0	56.3	10.2	33.5	**0.01**
0.25	56.4	10.2	33.4	**0.01**
0.5	56.4	10.2	33.4	0.02
0.75	56.2	10.2	33.4	0.15

Table 9. Ablation study for the NMS threshold $\text{NMS}_{\text{thres}}$ on WOD validation set.

anisms into CenterPoint++, but still with a max preservation age A_{max} for the tracklets. Then we increase A_{max} from 2 to study the influence of premature tracklet termination.

As shown in Figure 3, the mismatch ratio drop significantly and monotonically when we extend the life of tracklets by increasing A_{max}. When we preserve tracklets with predicted trajectories for 50 frames the mismatch ratio reaches 0.000123, 20 times lower than when the tracklets are preserved by predictions for only 2 frames. And when the life of tracklet continues to grow from 50 frames, up to 17% of remaining mismatch cases can further be avoided. This indicates that preserving tracklets by trajectory prediction can significantly reduce mismatch cases without introducing significant tracklets confusions.

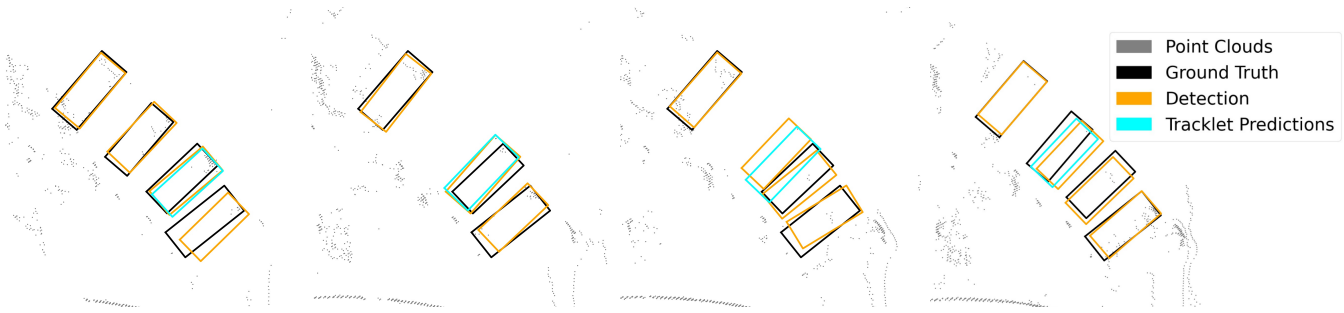
The minimum birth count M_{hits}. We also explore the influence of the minimum birth count M_{hits} on the performance of our proposed method. The results are shown in Table 7. Clearly Immortal Tracker is robust to M_{hits}.

Association Threshold. We find the association threshold $\text{IoU}_{\text{thres}}$ is a major influence factor to the performance of Immortal Tracker. As shown in Table 8, when we increase $\text{IoU}_{\text{thres}}$, Immortal Tracker reports significantly more mismatch cases with a lower MOTA.

Redundancy of input detections. In our method, we perform NMS on detection results with a relatively low IoU threshold compared to CenterPoint. The motivation is that considering rigid objects will not overlap in 3D space, the overlapped 3D detections have a high probability to be false positives and shall be discarded. Our experiment results in Table 9 show when we decrease the NMS threshold for a more strict suppression, the mismatch ratio drops by a large margin. This result supports our assumption.

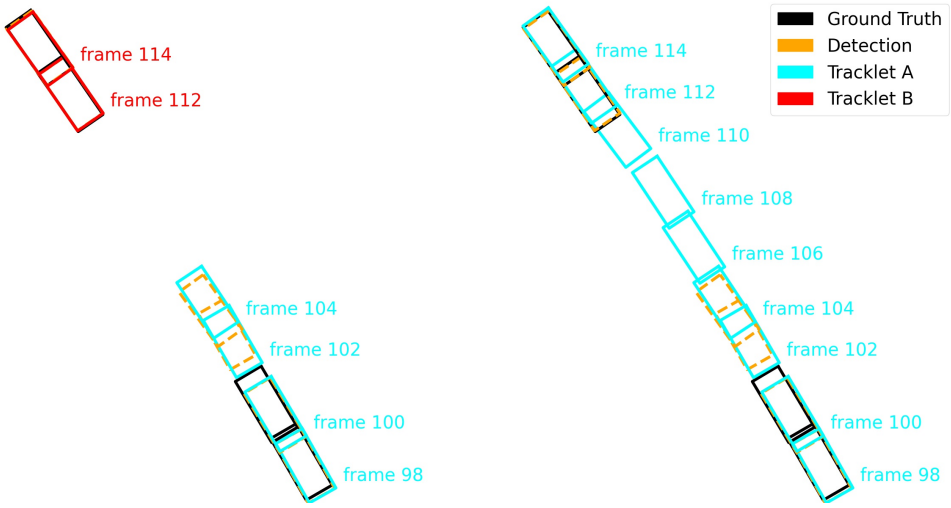
4.6. Qualitative Results.

Fig 4 shows the BEV visualization of the only remaining wrong association case in Immortal Tracker. Fig 5 visualizes one of the prevented premature terminations in Immortal Tracker.



(a) frame 108 (b) frame 112 (c) frame 114 (d) frame 118

Figure 4. Visualization of the remaining wrong association case. Outputs of the wrongly associated tracklet are drawn in cyan. The ground truth bounding boxes are drawn in black and the detection results are drawn in brown. The tracklet is misled by sequential inaccurate or false positive detections from its initial location to another. Such a case is extremely rare in our method.



(a) CenterPoint++ (b) Immortal Tracker

Figure 5. Visualization of a prevented premature termination. In this case the vehicle is not detected in frame 100-112. We plot the estimated bounding boxes of tracklets every two frames. Tracklets with different identities are drawn in different colors. For a better view we do not plot the ground truth bounding boxes in frame 100-112. (a)In CenterPoint++, the tracklet(marked in cyan) is terminated when the object is temporarily not observed, causing an identity switch. (b)While in Immortal Tracker, the initial tracklet is preserved until the object is detected again.

5. Conclusion and Future Direction

In this work, we found that identity switches in 3D MOT are almost solely caused by premature tracklet termination widely existing in modern 3D MOT methods. We proposed using trajectory prediction to preserve tracklets for invisible objects. We found using a simple 3D Kalman filter for trajectory prediction could reduce identity switch cases by 96%. Our method provided valuable insights for handling long-standing challenges in tracking like long-term occlusion. However, due to the limitation of existing 3DMOT benchmarks, we can not verify the effectiveness of our predict-to-track paradigm in long-term($\gg$ 20 seconds) and/or more interactive(*e.g.* the famous Shibuya crossing) scenarios. In the future, we want to set up a more challenging benchmark for 3DMOT and explore the synergy between trackers and more sophisticated trajectory predictors.

References

[1] Nuri Benbarka, Jona Schröder, and Andreas Zell. Score refinement for confidence-based 3d multi-object tracking. *arXiv preprint arXiv:2107.04327*, 2021. 1, 2, 3, 5

[2] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008:1–10, 2008. 4

[3] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Gi-

ancarlo Baldan, and Oscar Beijbom. nuscenes: A multi-modal dataset for autonomous driving. In *CVPR*, 2020. 4

[4] Shuyang Cheng, Zhaoqi Leng, Ekin Dogus Cubuk, Barret Zoph, Chunyan Bai, Jiquan Ngiam, Yang Song, Benjamin Caine, Vijay Vasudevan, Congcong Li, et al. Improving 3d object detection through progressive population based augmentation. In *ECCV*, 2020. 5, 6

[5] Hsu-kuang Chiu, Jie Li, Rareş Ambruş, and Jeannette Bohg. Probabilistic 3d multi-modal, multi-object tracking for autonomous driving. In *ICRA*, 2021. 2

[6] Hsu-kuang Chiu, Antonio Prioletti, Jie Li, and Jeannette Bohg. Probabilistic 3d multi-object tracking for autonomous driving. *arXiv preprint arXiv:2001.05673*, 2020. 1, 2, 3, 4, 5

[7] Caglayan Dicle, Octavia I Camps, and Mario Sznaier. The way they move: Tracking multiple targets with similar appearance. In *CVPR*, 2013. 2

[8] Lue Fan, Xuan Xiong, Feng Wang, Naiyan Wang, and Zhaoxiang Zhang. Rangedet: In defense of range view for lidar-based 3d object detection. *arXiv preprint arXiv:2103.10039*, 2021. 5, 6

[9] Jiawei He, Zehao Huang, Naiyan Wang, and Zhaoxiang Zhang. Learnable graph matching: Incorporating graph partitioning with deep feature learning for multiple object tracking. In *CVPR*, 2021. 2

[10] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960. 2

[11] Aleksandr Kim, Aljoša Ošep, and Laura Leal-Taixé. Eagermot: Real-time 3d multi-object tracking and segmentation via sensor fusion. CVPR-Workshops, 2020. 1, 2, 3, 4

[12] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955. 2

[13] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *CVPR*, 2019. 5, 6

[14] Zhichao Li, Feng Wang, and Naiyan Wang. Lidar r-cnn: An efficient and universal 3d object detector. In *CVPR*, 2021. 5, 6

[15] Yang Liu, Qinghua Hu, Yuexian Zou, and Wenwu Wang. Labelled non-zero particle flow for smc-phd filtering. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5197–5201. IEEE, 2019. 2

[16] Yang Liu, Wenwu Wang, Jonathon Chambers, Volkan Kilic, and Adrian Hilton. Particle flow smc-phd filter for audio-visual multi-speaker tracking. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 344–353. Springer, 2017. 2

[17] Yang Liu, Wenwu Wang, and Volkan Kilic. Intensity particle flow smc-phd filter for audio speaker tracking. *arXiv preprint arXiv:1812.01570*, 2018. 1, 2, 3, 4, 5

[18] Prasanta Chandra Mahalanobis. On the generalized distance in statistics. National Institute of Science of India, 1936. 2

[19] Ziqi Pang, Zhichao Li, and Naiyan Wang. SimpleTrack: Understanding and rethinking 3d multi-object tracking. *arxiv 2111.09621*, 2021. 2, 3, 5

[20] Abhishek Patil, Srikanth Malla, Haiming Gang, and Yi-Ting Chen. The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes. In *ICRA*, 2019. 2

[21] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *CVPR*, 2019. 2, 3

[22] Shaoshuai Shi, Xiaogang Wang, and Hongsheng Li. Pointrcnn: 3d object proposal generation and detection from point cloud. In *CVPR*, 2019. 2

[23] Xinshuo Weng and Kris Kitani. A baseline for 3d multi-object tracking. *arXiv preprint arXiv:1907.03961*, 1(2):6, 2019. 4

[24] Xinshuo Weng, Jianren Wang, David Held, and Kris Kitani. 3d multi-object tracking: A baseline and new evaluation metrics. *IROS*, 2020. 1, 2, 3, 4

[25] Xinshuo Weng, Yongxin Wang, Yunze Man, and Kris Kitani. Gnn3dmot: Graph neural network for 3d multi-object tracking with multi-feature learning. *arXiv preprint arXiv:2006.07327*, 2020. 1, 2, 3, 4

[26] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018. 5, 6

[27] Tianwei Yin, Xingyi Zhou, and Philipp Krähenbühl. Center-based 3d object detection and tracking. *CVPR*, 2021. 1, 2, 3, 4, 5

[28] Jan-Nico Zaech, Dengxin Dai, Alexander Liniger, Martin Danelljan, and Luc Van Gool. Learnable online graph representations for 3d multi-object tracking. *arXiv preprint arXiv:2104.11747*, 2021. 1, 2, 3, 4, 5

[29] Xingyi Zhou, Vladlen Koltun, and Philipp Krähenbühl. Tracking objects as points. In *ECCV*, 2020. 2