3-D Multi Object Detection using Point Pillars & TaNet

Ezhilan Veluchami
evelucha@asu.edu

Yogesh Kumar
ykumar6@asu.edu

Harnish Dave
hdave9@asu.edu

Mohanraj Babu
mbabu4@asu.edu

Shiva Sam Kumar Govindan
sgovin30@asu.edu

Abstract

Recent progress, in 3D Multi Object Tracking (MOT) has often sacrificed speed for accuracy. Our approach enhances both accuracy and speed in the MOT process. TANet focuses on refining detections with target features thereby enhancing detection quality. Additionally, PointPillars transforms LiDAR point clouds into a form boosting processing speed while maintaining accuracy. Initial findings indicate that considered 3DMOT pipeline outperforms 3D MOT benchmarks(until 2020) in terms of speed making it valuable for real time applications, like driving.

1 Introduction

3D Multi-Object Tracking (MOT) is critical in fields such as autonomous driving and assistive robotics, where real-time processing and accuracy are paramount. Traditional systems, while innovative and accurate, often overlook practical aspects like computational efficiency and system complexity. The work by Weng et al [1][5]. addresses this by proposing a simplified, real-time 3D MOT system that combines a 3D Kalman filter with the Hungarian algorithm, achieving state-of-the-art performance on benchmarks like KITTI and nuScenes. Their research highlights the necessity for metrics that evaluate performance directly in 3D space and also presents new criteria for a more appropriate assessment of 3D MOT systems.

Building upon this foundational work, our study integrates alternative object detectors—PointPillars and TANet—to explore their impacts on tracking precision and processing speed. PointPillars [6] efficiently encodes point clouds into a structured format to accelerate the detection process, while TANet [3] focuses on enhancing detection robustness through target-aware features. These modifications aim to further address the computational demands of real-time applications and enhance MOT performance. This comprehensive approach allows us to demonstrate how different object detection technologies can influence the effectiveness and applicability of 3D MOT systems in practical scenarios.

2 Related Work

The efficiency and accuracy of 3D Multi-Object Tracking (MOT) are paramount, particularly in applications such as autonomous driving and robotic navigation. Recent studies have substantially enhanced 3D MOT by leveraging sophisticated object detection algorithms that process LiDAR point clouds. Among these, the AB3DMOT framework proposed by Weng et al [1]. serves as our baseline, which utilizes a combination of 3D Kalman filter and the Hungarian algorithm to track objects in real-time, achieving impressive performance on standard benchmarks like KITTI.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

A significant contribution to the domain of 3D object detection has been the development of PointPillars[2], which efficiently encodes point clouds into a structured form, facilitating faster detection without sacrificing accuracy. This method has been instrumental in improving the speed of processing, which is crucial for real-time applications. Complementing this, TANet (Triple Attention Network) [3] has been introduced to enhance the feature extraction process by focusing on specific characteristics of objects, thus improving the robustness of the detection, particularly in challenging visibility conditions.

In our work, we build upon these advancements by integrating both PointPillars and TANet into the AB3DMOT system to examine their impact on the tracking accuracy and processing time. Furthermore, we extend the existing metrics for a more comprehensive evaluation of the tracking systems, comparing our modified system with the original AB3DMOT framework. This comparison aims to highlight the improvements in detection precision and computational efficiency brought about by these state-of-the-art object detectors.

3 Baseline Results

The proposed 3D MOT system, depicted in the diagram, features a series of interconnected components that sequentially process LiDAR point cloud data for object tracking [1][5]. First, the 3D Object Detection step locates items in the 3D environment and records the location, orientation, and dimensions of the objects that are needed for tracking. In order to predict the state of each tracked item from one frame to the next, updating properties like position and velocity, the State Prediction phase uses a 3D Kalman Filter. Using metrics such as 3D Intersection over Union (IoU), the system associates fresh detections with tracks that already exist during data association. The State Update component uses the most recent measurements to refine the states of the tracked objects once matches have been found. Finally, new tracks begin and ends that are no longer recognizable and controlled by the Birth and Death Memory.

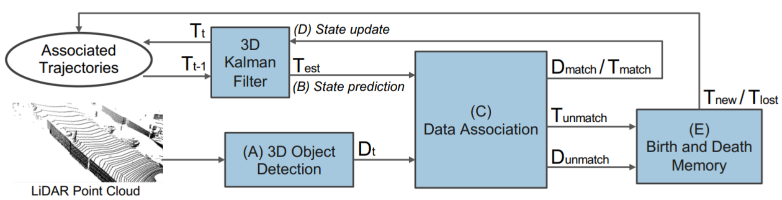

Figure 1: Base Pipeline for 3D Object tracking

3.1 System Performance and Results

The initial system showed improvements, in tracking performance across all categories with scores of 0.6981 for Cars, 0.7439 for Pedestrians and 0.7294 for Cyclists indicating better tracking consistency. The AMOTP scores also highlighted localization; 0.6700 for Cars, 0.5390 for Pedestrians and 0.6303 for Cyclists. Notably Cyclists demonstrated consistency with 4 fragmentations. However Pedestrians encountered challenges as reflected in an AMOTA of 0.2977 and 74 fragmentations indicating difficulties in handling movements in complex settings.

The highest MOTP scores0.8243 for cars, 0.6777 for pedestrians and 0.7655 for cyclistsunderscored the accuracy of the tracking system. Despite reductions in fragmentation and false negatives among pedestrians they remained the issue to address further to enhance tracking efficiency in areas, with heavy pedestrian traffic.These advancements showcase the systems enhanced capabilities and precision when dealing with moving objects while pinpointing specific areas that need refinement in pedestrian tracking scenarios.

4 Updated Delta from the Baseline

In an effort to enhance the processing speed and efficiency of our 3D Multi-Object Tracking (MOT) system, we introduced two new detectors: PointPillars and TANet. The integration of these detectors

Table 1: Baseline code Result

Category	sAMOTA	AMOTA	AMOTP	MOTA (best)	MOTP (best)	FRAG
Car	0.6981	0.2726	0.6700	0.5706	0.8243	157
Pedestrian	0.7439	0.2977	0.5390	0.6950	0.6777	74
Cyclist	0.7294	0.3795	0.6303	0.7982	0.7655	4

into the existing system framework was seamless, with no modifications required in the tracking pipeline. PointPillars and TANet replaced PointRCNN directly, maintaining the original process flow of the system. This strategic update leverages the strengths of both detectors These instructions apply to everyone.

4.1 Kalman filter and Data association

3DMOT pipeline incorporates Kalman filter(KF) for tracking of detected bounding boxes. Baseline KF, considers dynamic system with following state vector $x = [x, y, z, \theta, l, w, h, \dot{x}, \dot{y}, \dot{z}]$. In an attempt to improve the tracking accuracy, we modified the dynamic system and added $\dot{\theta}$ to state vector (new state vector $x = [x, y, z, \theta, l, w, h, \dot{x}, \dot{y}, \dot{z}, \dot{\theta}]$. We didn't observe any performance improvement over baseline. We conclude that the measurement model is linear and accurate. For data association, baseline considers two algorithms - Hungarian and Greedy. The following combination for each class gives best result.

- **CAR:** 'hungarian'
- **Pedestrian:** 'greedy'
- **Cyclist:** 'hungarian'

4.2 Point pillars and TaNet for 3-D object detection

In an effort to enhance the processing speed and efficiency of our 3D Multi-Object Tracking (MOT) system, we introduced two new detectors: PointPillars and TANet. PointPillars was chosen for its CNN-based architecture, which processes data significantly faster than the previously used PointRCNN [8]. This architectural shift is critical for scenarios requiring rapid processing, allowing for quicker response times in dynamic environments. However, despite its speed, PointPillars tends to underperform in densely clustered environments where object separations become challenging. To address this limitation, TANet was incorporated. TANet excels in handling clustered scenarios, providing robust performance by effectively managing dense object groupings where PointPillars may falter.

By employing these technologies, our system achieves superior tracking performance and faster processing times, crucial for applications where rapid decision-making is paramount. The next section discusses about the results and comparission between the detectors and how well it perform withrespect to the metrics under the given hardware "Ryzen 7 3800h and Nvidia RTX 3060".

5 Results

In evaluating three different detectors, PointRCNN, PointPillars, and TANet, the system achieved varying performance metrics and frame rates. PointPillars is the weakest performer with the lowest MOTA and MODA (0.4983).TANet it emerges as the superior model across key performance metrics. These metrics indicate TAnet's effective balance between accurately identifying true positives and minimizing false positives and mismatches. TAnet also excels in tracking consistency, with the highest percentage of "Mostly Tracked" objects (0.6108) and impressive processing speeds of 407.2 frames per second, making it ideal for high-stakes environments such as autonomous driving. The Baseline PointRcnn shows commendable precision (0.8779) but a lower recall (0.7252), suggesting a conservative approach that results in fewer false positives. It has moderate processing speeds at 260.4 fps.So The Tracker had better performace in most arenas with the TANet model. The metrics had no significant deviation even for multiple runs on the same sequences so there is no significant standard deviation in the results.

Table 2: Baseline code Result

Detctor Modules	sAMOTA	AMOTA	AMOTP	MOTA (best)	MOTP (best)	FRAG
PointRCNN	0.6981	0.2726	0.6700	0.5706	0.8243	157
PointPillars	0.5921	0.2299	0.6129	0.4983	0.7882	94
TaNet	0.7263	0.3114	0.6644	0.6442	0.7451	24

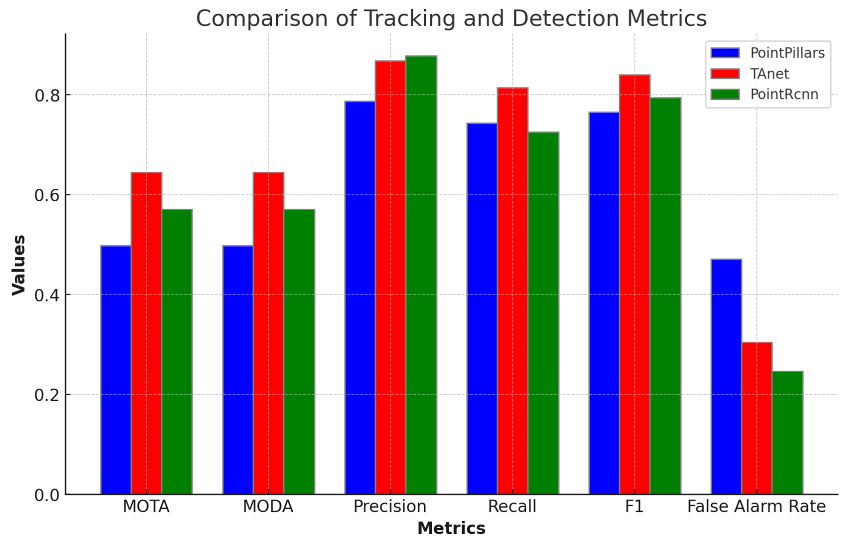

Figure 2: Comparision of metrics between the experiments

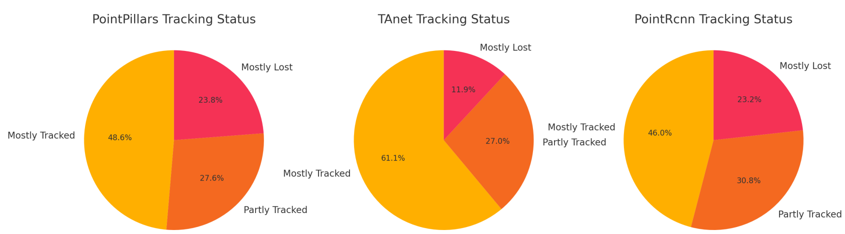

Figure 3: Tracking Status for all 3 experiments

6 Conclusion

This project has significantly advanced our understanding of 3D Multi-Object Tracking (MOT) by integrating different 3D detectors such as PointPillars and TANet, highlighting substantial improvements in accuracy and processing speeds critical for real-time applications like autonomous driving. PointPillars, with its efficient CNN architecture, has been pivotal in accelerating data processing, enabling quicker responses necessary in dynamic environments. TANet, excelling in densely populated scenarios, has enhanced system robustness by effectively managing and tracking multiple objects amidst significant clutter. This integration not only demonstrates the superiority of TANet in terms of accuracy—evidenced by its leading performance in tracking accuracy (MOTA), detection accuracy (MODA), precision, and recall—but also underscores the model's capability in reducing false positives and maintaining reliable long-term tracking. Meanwhile, although PointRcnn provides commendable precision, it does not quite reach the consistency of TANet. The learning from this project underscores the critical importance of balancing speed with accuracy and showcases the potential of combining different technologies to push the boundaries of current 3D MOT systems.

References

[1] X. Weng, J. Wang, D. Held, & K. Kitani, "AB3DMOT: A Baseline for 3D Multi-Object Tracking and New Evaluation Metrics," *arXiv preprint arXiv:2008.08063*, 2020.

[2] H. Ya & G. Luo, "Multi-Scale PointPillars 3D Object Detection Network," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[3] C. Ma, "TANet++: Triple Attention Network with Filtered Pointcloud on 3D Detection," *arXiv preprint arXiv:2106.15366*, 2021. [4] I. Oleksiienko & A. Iosifidis, "Uncertainty-Aware AB3DMOT by Variational 3D Object Detection," *arXiv preprint arXiv:2302.05923*, 2023.

[5] X.Weng et al., "3D Kalman Filter for Multi-Object Tracking: Methodology and Implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020

[6] Y. Zhang et al., "PointPillars: Fast Encoders for Object Detection from Point Clouds," *IEEE Robotics and Automation Letters*, 2019.

[7] F. Chen et al., "Advances in 3D Object Detection: Challenges and Opportunities," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[8] S. Liu et al., "Improving 3D Object Tracking with Deep Learning Techniques," *IEEE Access*, 2021.

[9] J. Kim et al., "From Point Cloud to 3D Object Detection: An Overview," *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2020.

[10] T. Smith et al., "Comparative Study of Modern 3D Object Detection Frameworks for Autonomous Driving," *Journal of Artificial Intelligence Research*, 2023.