GNN-PMB: A Simple but Effective Online 3D Multi-Object Tracker Without Bells and Whistles

Jianan Liu, Liping Bai, Yuxuan Xia, Member, IEEE, Tao Huang, Senior Member, IEEE, Bing Zhu, Member, IEEE, and Qing-Long Han, Fellow, IEEE

Abstract—Multi-object tracking (MOT) is among crucial applications in modern advanced driver assistance systems (ADAS) and autonomous driving (AD) systems. The global nearest neighbor (GNN) filter, as the earliest random vector-based Bayesian tracking framework, has been adopted in most of state-of-the-art trackers in the automotive industry. The development of random finite set (RFS) theory facilitates a mathematically rigorous treatment of the MOT problem, and different variants of RFS-based Bayesian filters have then been proposed. However, their effectiveness in the real ADAS and AD application is still an open problem. In this paper, it is demonstrated that the latest RFS-based Bayesian tracking framework could be superior to typical random vector-based Bayesian tracking framework via a systematic comparative study of both traditional random vector-based Bayesian filters with rule-based heuristic track maintenance and RFS-based Bayesian filters on the nuScenes validation dataset. An RFS-based tracker, namely Poisson multi-Bernoulli filter using the global nearest neighbor (GNN-PMB), is proposed to LiDAR-based MOT tasks. This GNN-PMB tracker is simple to use, and it achieves competitive results on the nuScenes dataset. Specifically, the proposed GNN-PMB tracker outperforms most state-of-the-art LiDAR-only trackers and LiDAR and camera fusion-based trackers, ranking the 3rd among all LiDAR-only trackers on nuScenes 3D tracking challenge leaderboard (https://bit.ly/3bQJ2CP) at the time of submission.

Index Terms—Autonomous driving, GNN-PMB, LiDAR, multi-object tracking, random finite set-based Bayesian filters, random vector-based Bayesian filters.

I. INTRODUCTION

MULTI-OBJECT tracking (MOT) is an integral and critical computational module for various systems, including autonomous vessels [1], roadside traffic monitoring [2], advanced driver assistance systems (ADAS) and autonomous driving (AD) [3], etc. Motivations of using a multi-object tracker include: 1) the tracker assigns and maintains a unique track ID for the same object throughout the life cycle of a tracking process; 2) the tracker rejects the false detection provided by the object detector; 3) the tracker sustains the tracking process when the tracked object fails to be detected over consecutive frames; and 4) the tracker refines the state information provided by the upstream module, e.g. bounding boxes estimated by object detector using mono-camera [4] or LiDAR [5], instance clusters estimated by radar [6], [7], etc., to support the downstream module like prediction [8]. Essentially, a multi-object tracker is a state estimator, or equivalently, a filter. In this paper, the concepts “multi-object tracker,” “tracker,” and “filter” are equivalent.

Depending on the employed sensor modalities, the task of MOT can be categorized into 2D MOT with camera [9], [10], 3D MOT with LiDAR only [11], [12], [13], and 3D MOT with both camera and LiDAR [14], [15]. Based on the processing order of detection and tracking, the task of MOT can be further categorized into tracking-before-detection [16], joint-detection-and-tracking [17], and tracking-by-detection [18]. In this paper, we focus on the tracking-by-detection 3D MOT with LiDAR for ADAS and AD application, which is the process when outputs of an object detector across frames are refined and connected by their corresponding track IDs, as shown by Fig. 1. Specifically, the object detector first provides the bounding box information, then the multi-object tracker refines the information provided by the object detector and assigns the appropriate ID to the bounding boxes.

However, the existing 3D MOT strategies and algorithms in literature are more complicated, requiring either additional data-driven feature extraction modules or heuristic tricks and rules for data association and track maintenance [15], [19], [20]. The rules may not be easily generalized in different scenarios or computationally feasible for embedded systems in the real-world. In contrast, this paper proposes a simple but effective online 3D multi-object tracker, namely Poisson multi-Bernoulli filter using the global nearest neighbor (GNN-PMB). The proposed GNN-PMB tracker is designed under random finite set (RFS) framework without requiring any feature extraction modules, heuristic tricks or additional rules, yet outperforms the...
A simple but effective 3D online LiDAR-based tracker, a systematic comparative study is presented for the first time on the nuScenes \cite{21} 3D LiDAR-based MOT benchmark dataset. The main contributions of this paper are listed as follows:

- A systematic comparative study is presented for the first time by employing several commonly-used Bayesian filters and the proposed GNN-PMB filter, with different object detectors, on the nuScenes benchmark dataset. This paper is based on real-world data and provides guidelines for designing the tracking framework in practice.

- A simple but effective 3D online LiDAR-based tracker, GNN-PMB, is proposed. The proposed GNN-PMB is designed in a unified and simple RFS framework, naturally allowing for a mathematically rigorous treatment of the data association hypothesis and track maintenance. It requires no extra feature extraction module, heuristic tricks, or rules for data association and track maintenance. Therefore, it is simple to tune and use, and more robust to detection errors than the random vector-based GNN framework with many heuristics, which has been widely applied in the practical ADAS and AD system.

- The performance of the proposed tracker is evaluated on the nuScenes \cite{21} 3D LiDAR-based MOT benchmark dataset. The main contributions of this paper are presented and discussed in Section IV. The proposed GNN-PMB tracker is detailed in Section V. Experimental results are provided and analyzed in Section VI. Finally, the conclusion is drawn in Section VII.

\section{Related Works}

In this paper, we focus on the task of 3D MOT with LiDAR only, which can be further differentiated into MOT without deep learning and MOT with deep learning. In this section, we briefly review these approaches. In addition, we also review MOT methods with LiDAR and Camera fusion.

\subsection{3D MOT With LiDAR Only}

1) \textit{LiDAR-based 3D MOT without Deep Learning:} Most current trackers for 3D MOT tasks in the ADAS and AD system utilize random vector-based GNN frameworks with many tricks and heuristics for track-and-detection data association and track maintenance. AB3DMOT \cite{11} establishes a baseline for LiDAR-based 3D MOT problem. In this algorithm, the track-and-detection association is computed based on the 3D intersection over union (IoU) score \cite{23}, and the track-and-detection association with the highest association score is regarded as the valid association scheme. Probabilistic 3D MOT \cite{12} which achieves better results than AB3DMOT, was the first work applying Mahalanobis distance instead of 3D IoU to evaluate track-and-detection association in the 3D MOT problem. SimpleTrack \cite{13} uses a generalized 3D intersection over union (GIOU) instead of 3D IoU as the track-and-detection association score, and it uses a non-maximum suppression (NMS) to process detection information provided by the object detector. This work has demonstrated that combining GIOU and NMS preprocessing improves the tracking result. ACK3DMOT \cite{24} proposes a joint probability function of appearance, geometry, and distance correlation among the detected bounding boxes and predicted objects to form the cost matrix for track-and-detection data association, leading to improved tracking performance when combined with an adaptive cubature Kalman filter. To improve track continuity, both ImmortalTracker \cite{25} and PC3T \cite{20} evaluate the similar idea of silently maintaining the tracks even when the tracks are no longer visible, which help reduce the ID switches and fragmented tracks. Score refinement is proposed in confidence-based 3D MOT \cite{26} for track maintenance, where the tracker achieves a low ID switch score and track fragmentation score.

2) \textit{LiDAR-Based 3D MOT With Deep Learning:} SimTrack \cite{27} firstly applies a RFS-based method, specifically the PMBM filter, to the LiDAR 3D MOT problem, where the data association problem is addressed by propagating multiple data association hypotheses with the highest likelihoods over time. Belief Propagation Tracker \cite{28}, was the first time formulated the MOT problem in its factor graph representation. Then, the sum-product message passing algorithm is used to compute the approximate marginal association probability iteratively.

There are also LiDAR only 3D MOT trackers that do not use GNN for data association. RFS-M3 \cite{27} firstly applies a RFS-based method, specifically the PMBM filter, to the LiDAR 3D MOT problem, where the data association problem is addressed by propagating multiple data association hypotheses with the highest likelihoods over time. Belief Propagation Tracker \cite{28}, was the first time formulated the MOT problem in its factor graph representation. Then, the sum-product message passing algorithm is used to compute the approximate marginal association probability iteratively.
and tracking from raw point clouds. A graph structure based on neural message passing was designed in OGR3MOT [30] to process detection and tracking in an online manner, where the data association is fully trainable. Neural enhanced belief propagation (NEBP) tracker [31] leverages the belief propagation tracker with a learned graph neural network, and it achieves the state-of-the-art performance on 3D LiDAR MOT.

B. 3D MOT With LiDAR and Camera Fusion

In principle, the performance of MOT with LiDAR only input can be improved by considering LiDAR and Camera fusion [14], [15], [19]. The Probabilistic 3D Multi-Modal MOT fuses features from 2D images and 3D LiDAR point clouds to capture the appearance and geometric information of objects [14]. In addition, a new metric that combines the Mahalanobis and feature distances is used for the track-and-detection association. This results show that, by incorporating the camera information, the Probabilistic 3D Multi-Modal MOT tracker achieves better tracking results than the Probabilistic 3D multi-object tracker [12] that only uses 3D LiDAR detection as input. Similar to DeepFusionMOT [22], EagerMOT [15] increases the tracking performance, compare to Probabilistic 3D Multi-Modal MOT, by utilizing the two-stage data association scheme for 3D detection fused from both LiDAR and camera and 2D detection from the camera, respectively. Later, AlphaTrack [19] further improves the performance of EagerMOT by adding a feature extractor along with the object detector. The feature extractor takes image information and the LiDAR point cloud information as inputs and provides information for the track-and-detection association.

III. BASIS OF RFS-BASED METHOD

A. Set Notation

In RFS-based methods, object states (e.g., position, orientation, size of bounding box, etc.) and their corresponding measurements are represented in the form of finite sets. It is assumed that there are \( n_k \) objects at time \( k \). Let \( x^i_k \) denotes the state of \( i \)-th object at time \( k \). Thus, the multi-object state at time \( k \) can be represented as a finite set \( X_k = \{ x_1^k, \ldots, x_{n_k}^k \} \). The cardinality of this set is \( |X_k| = n_k \). In addition, assume there are \( p_k \) measurements at time \( k \). Let \( Z_k = \{ z_1^k, \ldots, z_{p_k}^k \} \) denote the set of measurements at time step \( k \) and let \( Z^k = \{ Z_1^k, \ldots, Z_k^k \} \) denotes the sequence of all the measurement sets received so far up to and including time step \( k \).

B. Key Random Processes

Two random processes have prominent roles in the RFS-based methods: Poisson point process (PPP) and Bernoulli process. The PPP is a random set of points where the cardinality of the set is Poisson distributed. Similarly, the Bernoulli process is a random set of points where the cardinality of the set is distributed based on the Bernoulli process.

A PPP can be described using its intensity function \( \lambda(\cdot) \), and in this way its RFS density is

\[
\rho_{\text{PPP}}(X) = e^{-\int \lambda(x) dx} \prod_{x \in X} \lambda(x),
\]

where \( X \) is a finite set and its cardinality \( |X| \) is Poisson distributed with mean \( \lambda = -\int \lambda(x) dx \).

A Bernoulli process with existence probability \( r \) and existence-conditioned probability density function (PDF) \( f(\cdot) \) has RFS density

\[
\rho_{\text{Ber}}(X) = \begin{cases} 
1 - r & X = \emptyset; \\
rf(x) & X = \{x\}; \\
0 & \text{otherwise}.
\end{cases}
\]

The cardinality \( |X| \) is Bernoulli distributed with parameter \( r \). It is worth mentioning that the existence probability is used for the proposed GNN-PMB tracker to indicate the likelihood that the object exists at the current time step.

A multi-Bernoulli (MB) RFS \( X \) is the union of a finite number of independent Bernoulli processes \( X_1, \ldots, X_n \), and its density is

\[
\rho_{\text{MB}}(X) = \sum_{\mathcal{U} \subseteq \{1, \ldots, n\}} \rho_{\text{Ber}}(X_i)
\]

where \( \mathcal{U} \) denotes the disjoint union and \( \rho_{\text{Ber}}(\cdot) \) is the density of the \( i \)-th Bernoulli component as shown in (2).

C. Bayesian Recursion and Multi-Object Models

In Bayesian MOT filtering, the multi-object posterior \( f_{k|k}(X_k|Z^k) \) is critical because it captures all the information of the set \( X_k \) of object states at time step \( k \) conditioned on all the measurements. The posterior is computed by recursively applying the Chapman-Kolmogorov prediction

\[
f_{k|k-1}(X_k|Z^{k-1}) = \int \Phi_{k|k-1}(X_k|X_{k-1}) f_{k-1|k-1}(X_{k-1}|Z^{k-1}) dX_{k-1}
\]

and the Bayes update

\[
f_{k|k}(X_k|Z^k) \propto G(Z_k|X_k) f_{k|k-1}(X_k|Z^{k-1}),
\]

where \( G(Z_k|X_k) \) is the multi-object transition density for modeling the dynamics of multiple objects, and \( G(Z_k|X_k) \) is the multi-object measurement likelihood for modeling the measurement of multiple objects. In this paper, the standard multi-object dynamic model \( \Phi_{k|k-1}(X_k|X_{k-1}) \) is based on the following assumptions [32]:

- Single object with state \( x_{k-1} \) at time \( k - 1 \) moves to a new state \( x_k \) with a Markov transition density \( \phi(x_k|x_{k-1}) \). If \( x_{k-1} \) is the state of object \( i \) at time \( k - 1 \) and \( x_k \) is the new state of object \( i \) at time \( k \), then \( \phi(x_k|x_{k-1}) \) is the transition density of object \( i \) from \( x_{k-1} \) to \( x_k \).
- Single object with state \( x_k \) at time \( k \) has a probability \( 1 - P^S(x_k|x_{k-1}) \) of leaving the sensor’s field-of-view, where the superscript \( S \) refers to survival.
- The state of newborn objects \( X^0_k \) at time \( k \) appears in sensor’s field-of-view according to a PPP with intensity \( \lambda^0_k(\cdot) \).
• The appearing/disappearing of newborn/existing objects and the object motions are conditionally independent of the previous multi-object state $X_{k-1}$.

• The set $X_k$ of objects at time $k$ is the union of the set $X^o_k$ of existing objects at time $k$ and the set $X^b_k$ of newborn objects, i.e., $X_k = X^o_k \cup X^b_k$, where the superscript $o$ refers to existing, and superscript $b$ refers to newborn.

Specifically, the multi-object transition density from time $k-1$ to $k$, for the proposed GNN-PMB tracker is given by a convolution of a PPP density for newborn objects $X^b_k$ at time $k$ and a multi-Bernoulli density for existing objects $X^o_k$ inherited from time $k-1$ as:

$$\Phi_{k|k-1}(X_k|X_{k-1}) = \sum_{x_k \in X^o_k \cup X^b_k} f^\text{ppp}_{k}(X^b_k) f^\text{mb}_{k|k-1}(X^o_k|X_{k-1}) \cdot (6)$$

The standard multi-object measurement model $G(Z_k|X_k)$ is made with the following assumptions:

• The measurement set $Z_k$ at time step $k$ consists of measurements $Z^c_k$ generated by the state set $X_k$ of objects and clutter measurements $Z^e_k$, i.e., $Z_k = Z^c_k \cup Z^e_k$, where the superscript $c$ refers to the set generated by the set of objects, and the superscript $e$ refers to the set generated by the clutter process.

• $Z^c_k$ and $Z^e_k$ are statistically independent.

• No measurement is generated by more than one object.

• Given a state set $X_k$ of objects, each object state $x_k \in X_k$ is either detected with probability $P^D(x_k)$, where the superscript $D$ stand for detection probability, and generates a single measurement $z_k$ with measurement likelihood $g(z_k|x_k)$, or misdetected with probability $1 - P^D(x_k)$.

• $Z^e_k$ follows a PPP with intensity $\lambda^e_k(\cdot)$.

Specifically for GNN-PMB tracker, the multi-object measurement likelihood $G(Z_k|X_k)$ is given by a convolution of a PPP density for clutter measurements $Z^e_k$ and a multi-Bernoulli density for object-oriented measurements $Z^o_k$ by

$$G(Z_k|X_k) = \sum_{z_k \in Z^e_k \cup Z^o_k} f^\text{ppp}_{k}(Z^e_k) f^\text{mb}_{k}(Z^o_k|X_k) \cdot (7)$$

IV. A SYSTEMATIC STUDY OF BAYESIAN MOT METHODS

In this section, we present a systematic comparative study of common Bayesian MOT methods from mainly two aspects: 1) track maintenance and 2) approximation methods for computational tractability.

A. Global and Local Hypotheses

The main challenge of MOT is the unknown data association due to the unknown correspondence between objects and measurements. Therefore we start by giving a unified terminology of data association hypotheses. Consider the data association at time step $k$. A local hypothesis is defined as a pair of the object-to-measurement association at time $k$, and a global hypothesis is a valid collection of local hypotheses, explaining the association of every object and measurement at time $k$.

To further elaborate, let us consider the example illustrated in Fig. 2(a) where there are two objects $T_1,T_2$, and three measurements $m_1,m_2,m_3$. The global hypotheses describing their associations are represented using binary association matrices where each entry represents a possible local hypothesis, and we use dummy notation $m_0$ for misdetections. Each association matrix must satisfy: 1) each column must sum to one, and 2) each row must sum to one or zero. An all-zero row, except for $m_0$, means that the corresponding measurement belongs to the clutter. For the two global hypotheses shown in Fig. 2(b), Global Hypothesis 1 represents that object $T_1$ is misdetected and object $T_2$ is associated with measurement $m_1$, whereas Global Hypothesis 2 represents that object $T_1$ is associated to measurement $m_2$ and object $T_2$ is associated to measurement $m_1$.

B. Different Bayesian MOT Methods

Bayesian MOT methods can be generally categorized into 1) MOT methods based on random vectors and 2) MOT methods based on RFSs.

1) Vector-Based MOT Methods: Vector-based MOT methods describe the multi-object states and measurements by random vectors, and the most representative methods are GNN filter [33], joint probabilistic data association (JPDA) filter [34], [35], and multiple hypothesis tracking (MHT) filter [36].

Due to the unknown data associations, the number of global hypotheses increases hyper-exponentially over time. The GNN filter only keeps the most likely global hypothesis at each time step. The JPDA filter first computes the (approximate) marginal object-to-measurement association probabilities and then merges local hypotheses corresponding to the same object. MHT seeks to find the most likely global hypothesis over a sliding window of consecutive time steps, which involves the propagation of multiple global hypotheses over time.

2) Set-Based MOT Methods: Set-based MOT methods describe the multi-object states and measurements by RFSs. There are lots of RFSs-based MOT methods in the literature. Early developments include methods that avoid explicitly handling the data association uncertainty, such as probability hypothesis
density (PHD) filter [37] and cardinalized PHD (CPHD) filter [38] using moment approximations.

The PHD and CPHD filters approximate the multi-object posterior by a PPP and an i.i.d. cluster process, respectively, both in the sense of minimizing the Kullback-Leibler divergence. The CPHD filter is computationally heavier than the PHD filter, but it yields better performance when the signal-to-noise ratio is low.

A significant trend in RFSs-based MOT is the development of filters based on multi-object conjugate priors, which means that the multi-object posterior has the same functional form as the predicted distribution (and the prior). A typical example is the Poisson multi-Bernoulli mixture (PMBM) filter [39], which gives the close-form solution for the standard multi-object models introduced in Section III-C. In the PMBM filter, both the prediction and the update preserve the PMBM form of the density without approximation:

$$f^{\text{pmbm}}_{k|k'}(X_k|Z^k) = \sum_{X_k^{u|k} \in X_k^{d}} f^{\text{ppp}}_{k|k}(X_k^{u}) f^{\text{pmbm}}_{k|k'}(X_k^{d}),$$  \( (8) \)

$$f^{\text{pmbm}}_{k|k'}(X_k^{d}) = \sum_{h=1}^{H_{k'}} w_{h} f^{h}_{k|k'}(X_k^{d}),$$  \( (9) \)

where \( k' \in \{k-1, k\} \). In (8), the set \( X_k^{u} \) of undetected objects that have not yet been detected is described by a PPP; whereas the set \( X_k^{d} \) of detected objects that have been detected at least once is described by a multi-Bernoulli mixture (MBM). In the MBM (9), each multi-Bernoulli component corresponds to a unique global hypothesis for the detected objects. The \( h \)-th multi-Bernoulli component has density \( f^{h}_{k|k'}(\cdot) \) and weight \( w_{h} \), which satisfies that \( \sum_{h=1}^{H_{k'}} w_{h} = 1 \). If there is only a single multi-Bernoulli component in (9), i.e., when \( H_{k'} = 1 \), then the PMBM filter reduces to a PMB filter [40].

C. Track Maintenance

For practical implementations of Bayesian MOT methods, an efficient track maintenance scheme is required for tracking a time-varying number of objects. In this paper, a track is defined as a sequence of local hypothesis densities at consecutive time steps that correspond to the same object. Track maintenance refers to the process where a track is initiated, sustained, and terminated. In what follows, we discuss the track maintenance scheme for vector-based and set-based MOT methods separately.

1) Vector-Based MOT Methods: Vector-based MOT methods can maintain track continuity by associating an object state estimate with a previous state estimate. However, they mainly rely on heuristic methods to consider the appearance/disappearance of new/existing objects. A commonly used rule for track maintenance is called M/N logic. Specifically, a tentative track is initiated when a measurement is not associated with existing tracks. This tentative track is confirmed if there are \( M \) measurement associations out of \( N \) consecutive time steps. The termination of tracks follows a similar procedure. Alternatively, one can use track-score-based logic for track maintenance by performing hypothesis tests.

2) Set-Based MOT Methods: The RFS formalism facilitates modeling the appearance/disappearance of new/existing objects in a Bayesian setting. For example, in the PMBM and PMB filters, we only extract object state estimates from Bernoulli components with existence probability above a certain threshold. However, in RFS-based MOT methods, time sequences of tracks cannot be constructed easily as the multi-object states are order-independent. One approach to maintaining track continuity is to add unique labels to the object states and form tracks by linking object state estimates with the same label [41]. A more appealing approach to solve the track-building problem is by computing multi-object densities on sets of trajectories [42], and a typical example is trajectory PMBM filter [43]. We note that the prediction and update in the PMBM filter can be seen as an efficient method for calculating the time marginals of the RFS of trajectories [43]. Therefore, track continuity in the PMBM and PMB filters is implicitly maintained and can be established using metadata. The detailed procedure is described in Section V-B2.

D. Approximations for Computational Tractability

Practical MOT implementations need efficient approximations to keep the computational complexity at a tractable level. The approximations methods can be categorized into local and global hypothesis reductions.

1) Local Hypothesis Reduction: The commonly used strategy to limit the number of local hypotheses is gating. Specifically, for a predicted local hypothesis density, only the associations of measurements inside its gated region are considered to reduce the number of updated local hypotheses. In addition, different MOT methods use diverse techniques to reduce the number of local hypotheses further after the update. For example, in MHT, the number of local hypotheses is limited by implementing \( N \)-scan pruning [44] or by pruning local hypotheses with low scores. In Gaussian implementations of PHD and CPHD filters, Gaussian mixture reduction is performed for the PPP intensity by pruning components with small weights and merging similar components. In PMBM and PMB filters, it is necessary to prune Gaussian components in the PPP intensity with small weights and Bernoulli components with small existence probabilities.

2) Global Hypothesis Reduction: The key to global hypothesis reduction relies on how to solve the data association problem efficiently. For cracking the multi-scan data association problem, typical solutions include Lagrangian relaxation [44] and Markov chain Monte Carlo sampling [45]. For MOT methods considering the single-scan data association problem, the most likely global hypothesis can be obtained by solving a 2D assignment problem using algorithms such as the Hungarian algorithm [46]. The \( H_{k'} \) best global hypotheses can be obtained using Murty’s algorithm [46]. The merging step in the JPDA filter and the track-oriented PMB (TO-PMB) filter uses the (approximate) marginal association probabilities. They can be either computed using the \( H_{k'} \) best global hypotheses obtained from Murty’s algorithm or directly obtained using loopy belief propagation (LBP) [47] without explicit enumeration of global hypotheses.
V. THE NEW GNN-PMB TRACKER

According to the analysis presented in [48] and [49], the PMBM filter results in the best performance in simulation, but it is computationally intensive. Thus, PMBM may not be suitable for the real-world scenario due to existence of many objects and the corresponding measurements per time step, as well as the management of massive hypotheses.

There are two common approaches to reducing the computational burden introduced in the hypothesis management part of the PMBM filter. The first approach is to merge different local hypotheses corresponding to the same Bernoulli component, e.g., using LBP, and the resulting filter is called TO-PMB [40]. The second approach is to only propagate the most likely global hypothesis. In this case, the PMBM recursion is reduced to the PM recursion with much lower computations. In this paper, we adopt the second approach and propose the GNN-PMB tracker for LiDAR-based 3D MOT.

In this section, we first explain the overall framework of the proposed GNN-PMB tracker, as illustrated in Fig. 3. Then, we elaborate on the details of two core modules, hypothesis management and track maintenance. At last, we discuss the rest of the modules and their corresponding parameters.

A. The Framework of GNN-PMB Tracker

The available information of the 3D bounding boxes provided by the LiDAR object detector includes \((x, y, z)\) coordinate in the global frame, bounding box size, orientation, velocity, detection score, and class type information. However, to keep the multi-object tracker as simple as possible, the proposed tracker only requires the \(x\) and \(y\) coordinates in the global Cartesian frame as input. Other information provided by the 3D LiDAR object detector is propagated without filtering. As shown in Fig. 3, the 2D bounding boxes in BEV, as the measurements of possible objects and clutters, would be utilized as input to the tracker. After the gating, the most likely global hypothesis is decided in hypothesis management procedure, by applying Hungarian algorithm to the association cost matrix, formed using input 2D bounding boxes and the predicted PPP and MB component. Then the updated local hypothesis densities are obtained by standard Kalman update according to the most likely global hypothesis. At last, pruning and state extraction from the MB are implemented afterward to refine the object states. In the last, the ID information of every object is given according to the meta information attached to every Bernoulli component to generate the trajectories.

B. Hypothesis Management and Track Maintenance for GNN-PMB Tracker

1) Hypothesis Management: The hypothesis management procedure, as illustrated in Fig. 4, is one of the core modules of the proposed GNN-PMB tracker. At time \(k - 1\), the selected most likely global hypothesis consists of two local hypotheses, as indicated by the left side of the figure. The local hypothesis \(lh_{1}\) indicates that track \(T_1\) is associated with measurement \(m_1\), and the other local hypothesis \(lh_{2}\) stands for the association between track \(T_2\) and measurement \(m_2\). At time \(k\), a cost matrix is generated based on the most likely global hypothesis at time \(k - 1\) as prior information. Let \(c_{ij}\) denote the \(i\)-th row \(j\)-th column element in the cost matrix, given by the corresponding negative logarithm of the association likelihood \(w\). Such cost
Fig. 4. The hypothesis management procedure of GNN-PMB tracker. A cost matrix is generated based on the most likely global hypothesis at time $k - 1$ and then propagated to the current time $k$. At time $k$, the cost matrix is organized and the most likely global hypothesis is selected by solving a 2D assignment problem using this cost matrix. In this figure, $lb_i^m$ denotes the local hypothesis with measurement $m$ and object $t$. $gh$ stands for global hypothesis, $\lambda^c_a(\cdot)$ is the clutter intensity, and $w$ is the likelihood score of the local hypothesis.

The element for local hypothesis within the gated area of a track is set to be the negative logarithm of $w$. For the detected track, $w$ denotes the occurrence likelihood of a local hypothesis:

$$
T_{w_{m_i}} = \frac{1}{2\pi} |P|^{-\frac{1}{2}} \exp^{-\frac{1}{2}(m_i - T_j)^T P^{-1} (m_i - T_j)},
$$

where $m_i$ is the 2D position of the $i$-th measurement, $T_j$ is the 2D position of the $j$-th track, and $P$ is the covariance matrix of the Gaussian density. For instance, as indicated by the right side of Fig. 4, the association likelihood score of the local hypothesis where $T_{1j}$ is associated with $m_1$ is $u_1^1$. Therefore, the element of the cost matrix in the corresponding grid is $c_{11} = -\log(u_1^1)$. The costs of local hypotheses for existing tracks outside the gated area are set to be negative infinity. For instance, the $4$-th measurement lies outside the gating region of the $1$-st track, thus the corresponding cost is negative infinity. Besides, the $m_0$ in Fig. 4 denotes the missed detection hypothesis, which represents the hypothesis that the track is still alive but misdetected. Furthermore, any measurement can also be generated by a new track, so $T_{1new}$ is used to indicate the initial detection of a new track. For each measurement, if it is outside the gated area of the Poisson intensity of undetected objects, represented using a Gaussian mixture, the likelihood score of the corresponding new track is given by $\lambda^c_a(\cdot)$ which is the intensity of PPP represents the clutter measurement set $Z^c_a$. As each measurement initiates a single new track, the likelihood that $m_i$ creates new track $T_{1new}$, where $i \neq i_{new}$ is zero, and therefore its corresponding element in the cost matrix is infinity.

Every feasible global hypothesis at time step $k$ can be represented by a binary association matrix $A$. Finding the association matrix, that corresponds to the most likely global hypothesis, can be further formulated as a 2D assignment problem in the following form:

$$
a_{ij}^* = \arg\min_{a_{ij}} \sum_{i,j} \|c_{ij} a_{ij}\|
$$

s.t. $\sum_i a_{ij} = 1, \sum_j a_{ij} = 1,

$$
a_{ij} \in \{0, 1\},

$$
(11)

where $c_{ij}$ denotes the cost element in the aforementioned cost matrix; $a_{ij}$ is the selection term of the $i$-th row and $j$-th column of association matrix $A$. The selection term has a binary value, i.e. it is either 0 or 1. $a_{ij} = 1$ means the entry at the $i$-th row and $j$-th column is selected, thus the local hypothesis between the corresponding measurement and track is part of the output global hypothesis, and $a_{ij} = 0$ means the entry at the $i$-th row and $j$-th column is not selected. This 2D assignment problem can be solved by the Hungarian algorithm.

After obtaining the most likely global hypothesis, the local hypotheses consist of the selected global hypothesis with likelihood scores lower than a predetermined threshold need to be pruned. The final global hypothesis, which is formed by the remaining local hypotheses after pruning, will be propagated to time step $k + 1$.

2) Track Maintenance: As another critical module of GNN-PMB tracker, track maintenance is achieved by utilizing the meta-data information conveyed in the different Bernoulli components. Suppose a measurement is assigned to its corresponding new Bernoulli component instead of an existing one. In this case, the new Bernoulli component would be output as a valid object, and the object ID is the current maximum ID plus 1. The new ID is stored as metadata to the new Bernoulli component. For the situation that measurement is associated with an existing Bernoulli component in the most likely global hypothesis, the Bernoulli component would be output as an existing object, and the ID of the object remains the same as the one for the existing Bernoulli component.

C. Parameters for the GNN-PMB Tracker

1) Gating Distance: Gating, which prunes away all the detected bounding boxes with a distance to the centroid of predicted tracks smaller than a threshold, before organizing the possible local hypotheses for each track. The gating distance can be defined using different metrics, e.g., Euclidean distance, 2D IOU, 2D GIOU, 2D Mahalanobis distance, etc. In the proposed

$$
\text{g} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2},
$$

where $(x_1, y_1)$ and $(x_2, y_2)$ are the centroids of the predicted and detected tracks, respectively.
GNN-PMB, we choose the 2D Mahalanobis distance which incorporates the uncertainty information as the gating distance, and is easier to compute than the 2D IoU and the 2D GIoU. It is defined by

\[ d = \sqrt{(z_1 - z_2)^T \mathbf{P}^{-1} (z_1 - z_2)}, \]

(12)

where \( z_1 \) and \( z_2 \) are two points from the same Gaussian distribution with covariance matrix \( \mathbf{P} \).

2) Clutter Rate: The clutter generation process is a PPP with intensity \( \lambda_2(x) \), as discussed in Section III-C. To keep the computation simple, it is assumed that a constant expected number of clutter is generated uniformly across the field of view (FoV). Therefore, the clutter intensity is defined as the expected number of clutters over the area of FoV.

3) Survival Probability: As specified in Section III-C, the survival probability of an object, \( P^S(x_k) \), represents the possibility that an object remains at the next frame. Theoretically, \( P^S(x_k) \) should be defined in terms of the object position \( x_k \). E.g., the disappearance of an object probably occurs around the peripheries of the FoV. However, \( P^S(x_k) \) is set to a constant to keep the computation as simple as possible.

4) Probability of Detection: Also according to Section III-C, the probability of detection of a given object \( x \) is denoted by \( P^D(x_k) \). To better capture the likelihood of every detected bounding box in the dynamic environment, \( P^D(x_k) \) is specifically chosen as detection score of each detected bounding box provided by object detector, rather than being set as a constant value. For the Bernoulli component that has previously detected but misdetected at the current time step, the detection score at the previous time step is applied as \( P^D(x_k) \) at the current time step.

5) Pruning Threshold of Existence Probability: As explained in III-B, a Bernoulli process for existing object is characterised by an existence probability \( r \) and existence-conditioned PDF \( f(\cdot) \). To reduce computation, all Bernoulli components with existence probability less than the pruning threshold need to be eliminated. To this end, the threshold should be small enough to retain the Bernoulli components for multiple frames before it is discarded.

6) Detection Score Threshold: The detection score threshold is a predetermined threshold to prune the input of the tracker. For instance, a detection score threshold of 0.5 means only detected bounding boxes with a detection score higher than 0.5 would be used as input to the MOT tracker.

7) Non-maximum Suppression (NMS) Threshold: The object detector often creates multiple bounding boxes around the same object, but only one detection is required for each object. Non-max suppression is used to suppress the less likely bounding boxes. The NMS score is implemented as the 3D-IoU [23] score in GNN-PMB. If there are multiple bounding boxes where 3D-IoU score exceeds the specified NMS threshold, only the bounding boxes with the highest detection score would be kept as the tracker input.

8) Poisson Birth Density: The object birth process is also modeled by PPP, as introduced in Section III-C. The cardinality is distributed according to a Poisson distribution with intensity \( \lambda_1(\cdot) \). In the proposed GNN-PMB, \( \lambda_1(\cdot) \) is an unnormalized Gaussian mixture with identical weights, and the initial covariance matrix \( \mathbf{P}_0 \) is identical for all the Gaussian components. Therefore, the weight of the Gaussian distribution and the initial covariance matrix \( \mathbf{P}_0 \) need to be tuned to specify the Poisson birth density.

9) State Extraction Threshold: As described in III-C, the detected multi-object state is modeled by a MB for the proposed GNN-PMB tracker. Each Bernoulli RFS density has its corresponding existence probability. Only the Bernoulli components with an existence probability higher than the specified extraction threshold are considered valid objects for a given frame. Bernoulli RFS density with an existence probability lower than the specified extraction threshold would be silently maintained until the existence probability falls below the pruning threshold.

VI. EXPERIMENTS AND ANALYSIS

A. Dataset and Evaluation Metrics

There are four major LiDAR-based 3D detection and tracking benchmark datasets, namely KITTI [50], Waymo [51], nuScenes [21], and Argoverse [52]. In this work, we select the nuScenes benchmark dataset to conduct experiments as the popular nuScenes dataset provides the most reliable perception situation and a large amount of testing data in diverse scenarios for 3D MOT. To properly evaluate our proposed GNN-PMB, we use the following basic evaluation metrics: number of true positive (TP), number of false negative (FN), number of false positive (FP), number of ID switch (IDS), and number of times a trajectory is fragmented (FRAG). In addition, the number of mostly tracked (MT) tracks, which denotes the number of ground-truth trajectories that are tracked for at least 80% of their respective life span, together with the number of mostly lost (ML) tracks, which denotes the number of ground-truth trajectories that are tracked only for at most 20% of their respective life span, are also applied to provide overall evaluation. Moreover, more comprehensive metrics like the average multi-object tracking accuracy (AMOTA) and the corresponding accuracy measurement multi-object tracking precision (AMOTP) are also used in paper. The detailed information of all the metrics mentioned above can be found on the nuScenes 3D MOT challenge website.\(^1\)

B. Comparisons Among Different Bayesian Filters

In this section, the filter performance in terms of different MOT metrics is discussed, as shown in Table I. All the secondary metrics, including MT, ML, TP, FP, FN, IDS, FRAG, are evaluated under the recall with the highest MOTA.

1) Quality of Detection Input: The tracking performance of a tracker is affected by the performance of the attached detector. In the experiment, we have used three different object detectors, including the CenterPoint detector, the Megvii detector, and the PointPillar detector, with a descending order of their detection performance in the nuScenes LiDAR-based detection challenge.

\(^1\)https://www.nuscenes.org/tracking
TABLE I
COMPARISON OF LiDAR-BASED OBJECT DETECTORS AND BAYESIAN MULTI-OBJECT TRACKERS ON NuScenes VALIDATION SET

| Tracker  | Detector          | AMOTA†  | AMOTPDL  | MT†   | ML  | TP†  | FP†  | FN†  | IDS†  | PRA†  |
|----------|------------------|---------|----------|-------|-----|------|------|------|-------|-------|
| GNN      | PointPillars     | 0.251   | 1.403    | 1924  | 2158| 50631| 11776| 45334| 5932  | 2579  |
|          | Megvi            | 0.509   | 0.881    | 3323  | 1722| 67774| 10759| 29703| 4420  | 1209  |
|          | CenterPoint      | 0.603   | 0.735    | 3725  | 1447| 72192| 11767| 25076| 4629  | 1236  |
| PHD      | PointPillars     | 0.17    | 1.472    | 1443  | 2581| 35833| 42811| 52341| 14073 | 4728  |
|          | Megvi            | 0.268   | 1.201    | 478   | 4670| 17292| 6501 | 79652| 4947  | 3365  |
|          | CenterPoint      | 0.313   | 1.114    | 835   | 3403| 18856| 39724| 67996| 15045 | 6071  |
| TO-PMB   | PointPillars     | 0.185   | 1.380    | 2474  | 2232| 56833| 19096| 44444| 768   | 857   |
|          | Megvi            | 0.294   | 0.926    | 2429  | 3162| 49827| 18613| 51722| 348   | 403   |
|          | CenterPoint      | 0.324   | 0.812    | 2737  | 2955| 53982| 15382| 47433| 482   | 449   |
| PMBM     | PointPillars     | 0.269   | 1.260    | 2709  | 2311| 59250| 12060| 40800| 1847  | 843   |
|          | Megvi            | 0.577   | 0.739    | 4314  | 1528| 79094| 13536| 21829| 974   | 406   |
|          | CenterPoint      | 0.645   | 0.600    | 4591  | 1349| 82480| 14608| 18234| 1183  | 403   |
| GNN-PMB  | PointPillars     | 0.311   | 1.231    | 2734  | 2236| 60259| 9993 | 39945| 1023  | 769   |
|          | Megvi            | 0.619   | 0.716    | 4314  | 1552| 79434| 11710| 21955| 508   | 372   |
|          | CenterPoint      | 0.707   | 0.560    | 4608  | 1347| 83314| 12362| 18113| 650   | 345   |

† The upper arrow indicates that better performance is registered with higher score, same for the following tables.
‡ The lower arrow indicates that better performance is registered with lower score, same for the following tables.

Note: JPDA and CPHD trackers have also been implemented, but cannot be executed in real-time due that massive number of detection provided by nuScenes dataset strain JPDA and CPHD computation.

The CenterPoint detection result is provided by its authors, and the nuScenes 3D tracking challenge organizer provides the results for PointPillars and Megvi. As shown in Table I, the tracking performances are commensurate to the quality of the detection input for all the evaluated trackers. The best performance is achieved with the CenterPoint detector input for each tested tracker, whereas the worst performance is achieved with the PointPillars detector.

1) Performance of GNN, PHD, JPDA, and CPHD: As shown in Table I, GNN outperforms PHD by a large margin despite using only simple hypothesis management for data association and heuristic rules for track maintenance, since PHD does not use the proper approach for data association, which may not be suitable for the situation in the real-world LiDAR-based 3D MOT.

In principle, the JPDA and CPHD filters may boost performance compared to GNN and PHD, respectively. However, the marginal association probability computation in JPDA [36] is NP-hard, and the computational complexity of the cardinality distribution in CPHD increases exponentially with the number of objects. These computation requirements cannot be reached due to the enormous number of detection inputs offered by nuScenes dataset. Due to memory limits, both filters would be automatically terminated during the tracking process.

3) Performance of TO-PMB: Table I also shows that the performance of TO-PMB in our experiments is insufficient. One of the reasons is that the TO-PMB suffers from the coalescence problem, which refers to the phenomenon that multiple tracked objects are merged into one tracked object when multiple objects move in close proximity. This effect is demonstrated by Fig. 5, which is taken from a scene where the vehicle goes through a parking lot with tightly parked vehicles. In this particular frame, there should have 8 parked cars in the studied area, but the TO-PMB filter only tracked seven parked cars as it merged two cars into one.

4) Performance of PMBM: The PMBM filter achieves the second best performance in Table I, outperforming all other filters but with the AMOTA score 8.769% lower than that of the GNN-PMB with the CenterPoint detector. The performance disparity can be attributed to the fact that incorrect local hypotheses would persist for the PMBM filter. In contrast, only the most likely global hypothesis is propagated in the GNN-PMB filter, thereby decreasing the probability of incorrect local hypotheses being persist. Fig. 6 compares the tracking result between the

---

2https://bit.ly/3bWHWSWA
3https://www.nuScenes.org/data/detection-pointpillars.zip
4https://www.nuScenes.org/data/detection-megvi.zip

Fig. 5. The coalescence of the TP-PMB filter. This figure demonstrates the coalescence problem of the TO-PMB filter by comparing the tracking results of the same time step from both the GNN-PMB tracker and the TO-PMB tracker, where the grey dots are the LiDAR point cloud, the blue bounding boxes are the TP object, and the green bounding box with dashed line is the FN object, i.e., the object indeed exists but has not been tracked. In the highlighted red rectangle area, GNN-PMB tracks all the objects correctly, but TO-PMB incorrectly merges two object estimates and as a result misses one object.
TABLE II
Ablation Study of Filter Parameters. This Table Indicates How the Tracking Result Changes With Different Parameters

| Parameters                  | Value | AMOTA↑ | AMOTP↓ | RBACK↑ | MT↑  | ML↑  | TP↑  | FP↑  | FN↑  | IDS↑ | PRAG↓ |
|-----------------------------|-------|--------|--------|--------|------|------|------|------|------|------|-------|
| Detection Score Threshold   | 0.1   | 0.707  | 0.559  | 0.738  | 4606 | 1353 | 83088| 12300| 18167| 642  | 341   |
| NMS Threshold               | 0.1   | 0.707  | 0.560  | 0.715  | 4611 | 1329 | 83057| 12491| 18082| 758  | 335   |
| Poisson Birth Density       | 0.0001| 0.626  | 0.764  | 0.653  | 4415 | 1588 | 81039| 14055| 20254| 604  | 429   |
| State Extraction Covariance | 15    | 0.699  | 0.566  | 0.730  | 4494 | 1518 | 81942| 13036| 19386| 569  | 291   |
| Initial Covariance          | 100   | 0.625  | 0.761  | 0.636  | 4363 | 1646 | 80024| 13538| 21169| 704  | 366   |

Fig. 6. Cascading failure of the PMBM filter. This figure compares the tracking results between the PMBM filter and the GNN-PMB filter. In (a), bus_{10} in dash bounding box is a false alarm track. The track of bus_{10} is sustained by the PMBM filter, leading to an ID switch where the ID of the valid track changes from bus_{1} to bus_{10} in the later time step illustrated by (b). These two errors are a series of cascading failures because the PMBM filter propagates K best global hypotheses. In the GNN-PMB filter, however, only the valid track for bus_{1} is registered, and it correctly maintains the same track ID for bus_{1} in the following time step, as illustrated by (c) and (d).

PMBM filter and the proposed GNN-PMB filter. The ground truth contains only one track, bus_{1}. However, the PMBM filter erroneously predicted the existence of another track bus_{10}. The PMBM filter made a mistake in the first frame, and the error actually persisted to the next frames and resulted in an ID switch.

C. Ablation Study of Parameters for the GNN-PMB Tracker

Since the performance of the proposed GNN-PMB tracker can be tuned by parameters described in Section V-C, a set of carefully selected parameters can let the GNN-PMB achieves a reasonably good performance. In this section, we present the parameter tuning process of the proposed GNN-PMB tracker, for the bus class type as an example. According to our experiments, some of the settings have minor influence on the performance of the GNN-PMB tracker hence their values are supplied directly as follows:

- The gating threshold is $\sqrt{40}$ in 2D Mahalanobis distance.
- The clutter rate is set to 0.001 over the area of FoV.
- The survival probability is set to 0.7.
- The pruning threshold of existence probability is $10^{-6}$.

Changing the values of the rest parameters, on the other hand, causes significant variations in performance, as shown in Table II. We elaborate on the details below:

1) Detection Score Threshold: The first row of Table II indicates that discarding the bounding box with a predetermined detection score threshold would result in a lower AMOTA score. When the detection score threshold is set to be 0.1, 0.2, 0.3, the AMOTA score are 0.707, 0.672 and 0.625 respectively. It should also be noted that the recall decreases, and the number of fragmented trajectories increase as the detection score threshold increases.

The TP score of the three detection thresholds are 83088, 80971, and 80601, respectively. It means that there are 2117 valid detection with detection score between 0.1 and 0.2, and that there are 370 valid detection with detection score between 0.2 and 0.3. As the change of detection score threshold, the number of fragmented trajectories varies as well. Specifically, the number of fragmented trajectories increased from 328 to 500, when the detection score threshold increased from 0.2 to 0.3. On the contrary, the number of fragmented trajectories decreased from 341 to 328 when the threshold is set to be 0.1. Therefore, the decrease in the number of fragmented trajectory is an artifact that fewer mostly tracked trajectories exist.

Our experiment results suggest that crucial information for trajectory continuity is contained in detection with a detection score lower than 0.3. Trajectory continuity is essential to later modules such as trajectory prediction and motion planning. Our observation shows that rather than applying the law of parsimony and removing the detection with a low detection score, the output of the LiDAR 3D object detector should be provided to the MOT tracker in its entirety.
TABLE III
TRACKING RESULTS OF PROPOSED METHOD AND DIFFERENT MODEL-BASED TRACKERS USING LiDAR ON nuSCENES

| Method             | AMOTA↑ | AMOTP↓ | MT↑   | ML↓   | TP↑   | FP↓   | FN↓   | IDS↓ | FRAG↓ |
|--------------------|--------|--------|-------|-------|-------|-------|-------|------|-------|
| AB3D-MOT (BROS 2020)* [11] | 0.151  | 1.501  | 1086  | 4428  | 34808 | 15088 | 75730 | 9027 | 2557  |
| Stanford-NPR-TRI (NeurIPS Workshop 2019)* [12] | 0.550  | 0.798  | 4294  | 2184  | 85399 | 17533 | 33216 | 950  | 776   |
| RFS-M3 (ICRA 2021)* [27] | 0.619  | 0.752  | 5107  | 1878  | 90872 | 16728 | 27168 | 1525 | 856   |
| CBMOT-LiDAR (BROS 2021)* [26] | 0.649  | 0.592  | 5319  | 1966  | 94916 | 16469 | 24092 | 557  | 450   |
| SimpleTrack (Arxiv 2021)* [13] | 0.668  | 0.550  | 5476  | 1780  | 95539 | 17514 | 23451 | 575  | 591   |
| BPTracker (Proceedings of the IEEE 2018)* [28] | 0.664  | 0.606  | 5186  | 2259  | 95053 | 18581 | 24358 | 154  | 221   |
| ImmortalTracker (Arxiv 2021)* [29] | 0.677  | 0.599  | 5565  | 1669  | 96584 | 18012 | 21661 | 320  | 477   |
| GNN-PMB (Out*) | 0.678  | 0.560  | 5698  | 1622  | 97274 | 17071 | 21521 | 770  | 451   |
| PF-MOT tracker (ICRA 2022)** [53] | 0.682  | N/A    | N/A   | N/A   | N/A   | N/A   | N/A   | N/A | N/A   |
| GNN-PMB (Ours)** | 0.707  | 0.560  | 4608  | 1347  | 83134 | 12362 | 18113 | 650  | 345   |

* The metrics are reported on the nuScenes test set.
** The metrics are reported on the nuScenes validation set.

2) NMS Threshold: We set the NMS threshold to 0.1, which allows us to discard the majority of overlapping detected bounding boxes even if only a small overlap occurs, yielding an AMOTA score of 0.707 with the CenterPoint object detector. In contrast, by setting the NMS threshold to 0.98, almost all the overlapped detected bounding boxes are kept as input to the GNN-PMB tracker, resulting a lower AMOTA score of 0.68. Our observation that a basic NMS pre-process operation can improve the AMOTA score by 0.27 denotes that only the detected bounding boxes with the least overlap with each other should be used as input to the tracker.

3) Poisson Birth Density: As discussed in section V-C, proper initialization of Poisson birth density is critical before tracking the newborn object. When the initial covariance of the Gaussian component increases from 15 to 100, the AMOTA score is reduced from 0.699 to 0.625, and the related AMOTP score increases from 0.566 to 0.761. This is reasonable since, as the covariance grows, the estimation uncertainty would increase, leading to performance degradation.

The weight of Gaussian components in the Poisson intensity also plays a crucial role in tracking newborn objects. When the weights increase, measurement becomes more likely to be construed as originating from a new object than an existing one since the cost of being associated with a new object would decrease. On one extreme, when the weight is too large, every measurement would be interpreted as the initial detection of a new track. On the other extreme, where the weight is too small, the measurements would always be associated with existing tracks, and PPP would initiate no new tracks. According to the ablation study, the AMOTA decreases as the weight decreases, while the number of ID switches varies at a different pace as the weight varies. Consequently, such observation demonstrates that finding a suitable value for the weight of the Gaussian component to fit the statistics of object birth in the nuScenes dataset is essential.

4) State Extraction Threshold: Since the probability of the Bernoulli component signifies how probable an object exists, precisely selecting a reasonable extraction threshold becomes important for the final estimation of object states. According to our ablation investigation, raising the extraction threshold increases AMOTA, MT, and TP while lowering ML, FP, and FN. This result reveals that preserving the Bernoulli components with a high enough existence probability would result in a more accurate track estimation.

D. Comparison with Other State-of-the-art Methods

1) Performance Comparison With State-of-the-art Model-based LiDAR Trackers Using LiDAR Only: In the nuScenes test dataset, our proposed LiDAR only tracker, GNN-PMB, is compared against various model-based LiDAR only trackers, and the results are reported in Table III. Among the model-based trackers, the proposed GNN-PMB tracker receives the highest AMOTA score of 0.678. In terms of AMOTP, the GNN-PMB tracker achieves a score of 0.560, which is only 0.01 lower than that of SimpleTrack, and it is the second-best AMOTP score amongst all compared model-based trackers. However, it should be noted that all the other trackers use all the 3D information provided by the object detector as input of the tracker, whereas the GNN-PMB tracker only utilizes the x and y coordinates in the global frame as input.

In addition, our proposed method also outperforms all the model-based trackers in MT, ML, TP, and FN, with the FP score coming in second only to CBMOT-LiDAR, and notably, achieving an MT score of 5698. Even with lower frequency input (at 2Hz), the GNN-PMB tracker still managed to track 4.05% more tracks than SimpleTrack, which employed detection at 10Hz as the input. Moreover, the GNN-PMB tracker is even superior to ImmortalTracker, which refines the trajectory with more accumulated information in the future through post-processing, though it is impractical in the real-world system for online tracking. Another recently proposed LiDAR-only model-based tracker, the PF-MOT tracker, only reported performances partially in the nuScenes validation dataset excluding the test dataset, limiting the comparison on the validation dataset. The result shows that the GNN-PMB tracker achieves an AMOTA of 0.707, which is higher than the 0.682 AMOTA achieved by the PF-MOT tracker.

The drawbacks of the GNN-PMB tracker are the higher IDS and FRAG compared to BPTracker, which has only 154 ID switches and 221 fragments. However, considering relatively low values in MT, the low ID switches achieved by the BP-Tracker could be partially due to an artifact from having less mostly tracked tracks.
TABLE IV

| Method               | AMOTA↑ | AMOTP↓ | MT↑  | ML↓  | TP↑  | FP↑  | FN↑  | IDS↓ | FRAG↓ |
|----------------------|--------|--------|------|------|------|------|------|------|-------|
| SimTrack (ICCV 2021)*[29] | 0.645  | 0.681  | 5063 | 1986 | 9209 | 17443| 26430| 1042 | 472   |
| OGR3MOT (IEEE RAAL 2022)*[30] | 0.656  | 0.620  | 5278 | 2094 | 9526 | 17877| 24013| 1088 | 371   |
| NEBP (Arxiv 2022)*[31]     | 0.673  | 0.586  | 5380 | 2126 | 97023| 19535| 22380| 162  | 256   |
| GNN-PMB (Ours)*            | 0.678  | 0.560  | 5698 | 1622 | 97274| 17071| 21521| 770  | 431   |
| TransMOT (IEEE TV 2022)**[54] | 0.674  | 0.754  | 2096 | N/A  | N/A  | 9499 | 14071| 1403 | N/A   |
| GNN-PMB (Ours)**          | 0.849  | 0.387  | 2762 | 668  | 49182| 6140 | 8791 | 344  | 170   |

* The metrics are reported on the nuScenes test set.
** Our tracker is based on LiDAR only.

2) Performance Comparison With State-of-the-art Data-Driven Trackers Using LiDAR Only: We also compare our proposed GNN-PMB with data-driven state-of-the-art trackers. As shown by Table IV, our proposed GNN-PMB tracker achieves the best performance in all evaluation metrics other than IDS and FRAG, when it is compared with the data-driven trackers. In particular, our proposed GNN-PMB tracker achieves an AMOTP score of 0.56, which is 0.026 better than the NEBP tracker, the second-best tracker among all the learning-based trackers. Furthermore, the AMOTP score indicating that the GNN-PMB tracker can provide more accurate position information than the NEBP tracker.

3) Performance Comparison With State-of-the-art Trackers Using LiDAR and Camera Fusion: To further demonstrate the advantages of our proposed GNN-PMB tracker, we present the comparison between the proposed GNN-PMB and the state-of-the-art trackers using LiDAR and camera fusion. It is worth pointing out that our proposed GNN-PMB tracker using LiDAR only, still can obtain comparable tracking performance on AMOTA and AMOTP and it is even superior on MT, TP, FP, and FN, as shown in Table V. Such results show an enormous potential to increase performance even further when the GNN-PMB tracking framework is extended into settings with the fusion of LiDAR and camera.

VII. CONCLUSION AND FUTURE WORK

A systematic comparison is provided among different random vector-based Bayesian and RFS filters on the nuScenes dataset. Based on the analysis, a simple but effective online multi-object tracker GNN-PMB is proposed. Due to the simple structure of the RFS framework, the proposed GNN-PMB tracker requires no additional module, heuristic trick, or data association and track maintenance rule. Thus, it is simple to tune, and it achieves the state-of-the-art performance on nuScenes dataset. Its performance can be further improved with simple modifications, such as providing 3D state information as input, using both LiDAR and camera as sensor modalities, and including a meta-learning-based parameter auto-tuning module, etc. In our future work, downstream tasks such as trajectory prediction and motion planning will be incorporated to evaluate the multi-object trackers, such that a more robust multi-object tracker in an end-to-end manner can be designed.

REFERENCES

[1] S. Gao, Z. Peng, L. Liu, D. Wang, and Q.-L. Han, “Fixed-time resilient edge-triggered estimation and control of surface vehicles for cooperative target tracking under attacks,” IEEE Trans. Intell. Veh., early access, Jun. 2022, doi: 10.1109/TIV.2022.3184076.
[2] Y. Du, B. Qin, C. Zhao, Y. Zhu, J. Cao, and Y. Ji, “A novel spatio-temporal synchronization method of roadside asynchronous MMW radar-camera for sensor fusion,” IEEE Trans. Intell. Transp. Syst., early access, Oct. 2021, doi: 10.1109/TITS.2021.3119079.
[3] J. Nidamanuri, C. Nibhanupudi, R. Assfalg, and H. Venkataraman, “A progressive review-emerging technologies for ADAS driven solutions,” IEEE Trans. Intell. Veh., vol. 7, no. 2, pp. 326–341, Jun. 2022.
[4] T. Gao, H. Pan, and H. Gao, “Monocular 3D object detection with sequential feature association and depth hint augmentation,” IEEE Trans. Intell. Veh., vol. 7, no. 2, pp. 240–250, Jun. 2022.
[5] T. Yin, X. Zhou, and P. Krahenbuhl, “Center-based 3D object detection and tracking,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 11784–11793.
[6] J. Liu et al., “Deep instance segmentation with automotive radar detection points,” IEEE Trans. Intell. Veh., early access, Apr. 2022, doi: 10.1109/TIV.2022.3168899.
[7] W. Xiong, J. Liu, Y. Xia, T. Huang, B. Zhu, and W. Xiang, “Contrastive learning for automotive mmWave radar detection points based instance segmentation,” in Proc. IEEE Int. Conf. Intell. Transp. Syst., 2022, arXiv:2203.06553.
[8] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, and H. Chen, “A survey on trajectory-prediction methods for autonomous driving,” IEEE Trans. Intell. Veh., vol. 7, no. 3, pp. 652–674, Sep. 2022, doi: 10.1109/TIV.2021.3167103.
[9] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, “FairMOT: On the fairness of detection and re-identification in multiple object tracking,” Int. J. Comput. Vis., vol. 129, no. 11, pp. 3069–3087, Nov. 2021.
[10] Y. Zhang et al., “Bytetrack: Multi-object tracking by associating every detection box,” 2021, arXiv:2110.06864.
[11] X. Weng, J. Wang, D. Held, and K. Kitani, “3D multi-object tracking: A baseline and new evaluation metrics,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2020, pp. 10359–10366.
[12] H.-K. Chiu, A. Priolleti, J. Li, and J. Bohg, “Probabilistic 3D multi-object tracking for autonomous driving,” in Proc. IEEE Int. Conf. Robot. Automat., 2021, pp. 14227–14233, doi: 10.1109/ICRA4806.2021.9561754.

[13] Z. Pang, Z. Li, and N. Wang, “SimpleTrack: Understanding and rethinking 3D multi-object tracking,” 2021, arXiv:2111.09621.

[14] H.-K. Chiu, J. Li, R. Ambruș, and J. Bohg, “Probabilistic 3D multi-modal, multi-object tracking for autonomous driving,” in Proc. IEEE Int. Conf. Robot. Automat., 2021, pp. 14227–14233.

[15] A. Kim, A. Osep, and L. Leal-Taixé, “EagerMOT: 3D multi-object tracking via sensor fusion,” in Proc. IEEE Int. Conf. Robot. Automat., 2021, pp. 11315–11321.

[16] S. J. Davey, M. G. Ruten, and B. Cheung, “A comparison of detection performance for several track-before-detect algorithms,” in Proc. 11th Int. Conf. Inf. Fusion, 2008, pp. 1–8.

[17] Z. Lu, V. Rathod, R. Votel, and J. Huang, “RetinaTrack: Online single stage joint detection and tracking,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 14656–14666.

[18] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “Exploiting the circulant structure of tracking-by-detection with kernels,” in Proc. Eur. Conf. Comput. Vis., 2012, pp. 702–715.

[19] Y. Zeng, C. Ma, M. Zhu, Z. Fan, and X. Yang, “Cross-modal 3D object detection and tracking for auto-driving,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2021, pp. 3850–3857.

[20] H. Wen, W. Han, C. Wen, Y. Li, and C. Wang, “3D multi-object tracking in point clouds based on prediction confidence-guided data association,” IEEE Trans. Intell. Transp. Syst., vol. 23, no. 6, pp. 5668–5677, Jun. 2022.

[21] H. Caesar et al., “nuScenes: A multimodal dataset for autonomous driving,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 11621–11630.

[22] X. Wang, C. Fu, Z. Li, Y. Lai, and I. He, “DeepFusionMOT: A 3D multi-object tracking framework based on camera-LiDAR fusion with deep association,” IEEE Robot. Automat. Lett., vol. 7, no. 3, pp. 8260–8267, Jul. 2022.

[23] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU loss: Faster and better learning for bounding box regression,” in Proc. AAAI Conf. Artif. Intell., vol. 34, 2020, pp. 12993–13000.

[24] G. Guo and S. Zhao, “3D multi-object tracking with adaptive cubature Kalman filter for autonomous driving,” IEEE Trans. Intell. Veh., early access, Mar. 2022, doi: 10.1109/TIV.2022.3158419.

[25] Q. Wang, Y. Chen, Z. Pang, N. Wang, and Z. Zhang, “Immortal tracker: Tracklet never dies,” 2021, arXiv:2111.13672.

[26] N. Benburska, J. Schröder, and A. Zell, “Score refinement for confidence-based 3D multi-object tracking,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., 2021, pp. 8083–8090.

[27] S. Pang, D. Morris, and H. Radha, “3D multi-object tracking using random finite set-based multiple measurement models filtering (RFS-M3) for autonomous vehicles,” in Proc. IEEE Int. Conf. Robot. Automat., 2021, pp. 13701–13707.

[28] F. Meyer et al., “Message passing algorithms for scalable multitarget tracking,” Proc. IEEE, vol. 106, no. 2, pp. 221–259, Feb. 2018.

[29] C. Luo, X. Yang, and A. Yuille, “Exploring simple 3D multi-object tracking,” 2021, arXiv:2111.09621.

[30] M.-F. Chang et al., “Argoverse: 3D tracking and forecasting with rich sensor fusion,” in Proc. IEEE Intell. Veh. Symp. (IV), 2022, pp. 852–859.

[31] Jianian Liu received the B.Eng. degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2007, the M.Eng. degree in telecommunication engineering from the University of Melbourne, Melbourne, VIC, Australia, and the M.Sc. degree in communication systems from Lund University, Sweden, in 2009 and 2012, respectively.

[32] He has more than ten years of experience in software and algorithm design and development. He has held Senior R and D roles with the AI consulting, automotive, and telecommunication industries.

[33] His research interests include applying statistical signal processing and deep learning for medical image processing, wireless communications, IoT networks, indoor sensing, and outdoor perception using a variety of sensor modalities include radar, camera, LiDAR, WiFi, etc.

[34] Liping Bai received the M.Eng. degree in instrument engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2021. She is currently working toward the Ph.D. degree in automation and control engineering with Beihang University, Beijing, China. Her research interests include intersection between control theory and machine learning.

[35] J. L. Williams, “Marginal multi-Bernoulli filters: RFS derivation of MHT, JIPDA, and association-based member,” IEEE Trans. Aerosp. Electron. Syst., vol. 51, no. 3, pp. 1664–1687, Jul. 2015.

[36] E. H. Aoki, P. K. Mandel, L. Svensson, Y. Boers, and A. Bagchi, “Labeling uncertainty in multi-target tracking,” IEEE Trans. Aerosp. Electron. Syst., vol. 52, no. 3, pp. 1006–1020, Jul. 2016.

[37] Á. F. García-Fernández and L. Svensson, “Trajectory PHD and CPHD filters,” IEEE Trans. Signal Process., vol. 67, no. 22, pp. 5702–5714, Nov. 2019.
Yuxuan Xia (Member, IEEE) received the M.Sc. degree in communication engineering and the Ph.D. degree in signal processing from the Chalmers University of Technology, Gothenburg, Sweden, in 2017 and 2022, respectively. He is currently a Postdoctoral Researcher with the Department of Electrical Engineering, Chalmers University of Technology. His main research interests include multi-object tracking and sensor fusion, especially for extended objects. He has co-organized tutorials on multi-object tracking at the Fusion 2020, 2021 and 2022 conferences.

Bing Zhu (Member, IEEE) was born in Wuhan, China. He received the B.S and Ph.D. degrees in control theory and applications from Beihang University, Beijing, China, in 2007 and 2013, respectively. He was with the University of Pretoria, Pretoria, South Africa, as a Postdoctoral Fellow supported by Vice-Chancellor Postdoctoral Fellowship from 2013 to 2015. He was with Nanyang Technological University, Singapore, as a Research Fellow from 2015 to 2016. In 2016, he joined Beihang University, as an Associate Professor. His research interests include model predictive control, smart sensing for UAV and UGV, and demand-side management for new energy systems.

Dr. Zhu is an Associate Editor for Acta Automatica Sinica.

Qing-Long Han (Fellow, IEEE) received the B.Sc. degree in mathematics from Shandong Normal University, Jinan, China, in 1983, and the M.Sc. and Ph.D. degrees in control engineering from the East China University of Science and Technology, Shanghai, China, in 1992 and 1997, respectively.

Professor Han is currently a Pro Vice-Chancellor (Research Quality) and a Distinguished Professor with the Swinburne University of Technology, Melbourne, VIC, Australia. He held various academic and management positions with Griffith University, Brisbane, QLD, Australia, and Central Queensland University, Rockhampton, QLD, Australia. His research interests include networked control systems, multi-agent systems, time-delay systems, smart grids, unmanned surface vehicles, and neural networks.

Professor Han was the recipient of the 2021 Norbert Wiener Award (the Highest Award in systems science and engineering, and cybernetics) and The 2021 M. A. Sargent Medal (the Highest Award of the Electrical College Board of Engineers Australia). He was also the recipient of The 2022 IEEE Systems, Man, and Cybernetics (SMC) Society Andrew P. Sage Best Transactions Paper Award, The 2021 IEEE/CAA Journal of Automatica Sinica Norbert Wiener Review Award, The 2020 IEEE SMC Society Andrew P. Sage Best Transactions Paper Award, The 2020 IEEE Transactions on Industrial Informatics Outstanding Paper Award, and The 2019 IEEE SMC Society Andrew P. Sage Best Transactions Paper Award.

Professor Han is a Member of the Academy Europea (The Academy of Europe). He is a Fellow of The International Federation of Automatic Control (IFAC) and a Fellow of The Institution of Engineers Australia (IEAust). He is a Highly Cited Researcher in both Engineering and Computer Science (Clarivate Analytics). He was an AdCom Member of IEEE Industrial Electronics Society (IES), a Member of IEEE IES Fellows Committee, and Chair of IEEE IES Technical Committee on Networked Control Systems. He is the Co-Editor-in-Chief of IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, Deputy Editor-in-Chief of IEEE/CAA JOURNAL OF AUTOMATIC SINICA, Co-Editor of Australian Journal of Electrical and Electronic Engineering, an Associate Editor for 12 international journals, including the IEEE TRANSACTIONS ON CYBERNETICS, IEEE Industrial Electronics Magazine, Control Engineering Practice, and Information Sciences, and a Guest Editor for 14 Special Issues.

He was an Endeavour Australia Cheung Kong Research Fellow, a Visiting Scholar with The Chinese University of Hong Kong, Hong Kong, a Research Associate with the University of New South Wales, Sydney NSW, Australia, and a Postdoctoral Research Fellow with James Cook University. Before academia, Dr. Huang held various positions in the industry, such as Senior Software Engineer, Senior Data Scientist, Project Team Lead, and Technical Authority. He has coauthored a Best Paper Award from the 2011 IEEE WCNC, Cancun, Mexico. His research interests include deep learning, smart sensing, computer vision, pattern recognition, wireless communications, and IoT security. He is a co-inventor of one patent on MIMO systems. He was with several international conferences as the TPC Chair, Track Chair, Program Vice Chair, and Local Chair. He is currently the MTT-S/Com Vice-Chair and Young Professionals Affinity Group Chair for the IEEE Northern Australia Section.