Hash-Based Gaussian Mixture Model (HGMM) for Roadside LiDAR Smart Infrastructure Applications

Tianya Terry Zhang, Yi Ge, Anjiang Chen, Mina Sartipi, Senior Member, IEEE, and Peter J. Jin

Abstract—This paper proposes a point cloud background filtering method and explores applications that integrate LiDAR technology into roadside sensor networks, addressing challenges in handling large volumes of sparse and unstructured 3D data. Traditional methods, which classify background points based on descriptive statistics over many frames, are computationally inefficient for roadside LiDAR surveillance. This study introduces a novel approach using hash-based transformation combined with probabilistic Gaussian Mixture Model (GMM) techniques, enhancing the efficiency of high-dimensional multivariate modeling of LiDAR data. This Hash-based Gaussian Mixture Modeling (HGMM) optimizes feature selection and Gaussian components using AIC and BIC scores, mitigating computational burdens and parameter sensitivity. Unlike Cartesian coordinate-based techniques, HGMM processes LiDAR data in Spherical coordinates, preserving meaningful patterns and structures. For infrastructure LiDAR, object detection only pertains to a small amount of data in a fixed environment, allowing background reduction modeling to significantly enhance data chain efficiency by transitioning only a tiny portion of the foreground LiDAR point clouds. The method was tested on diverse LiDAR datasets, including the New Brunswick DataCity Testbed dataset, Transportation Forecasting Competition (TRANSFOR 24) Dataset, DAIR-V2X, and A9-Dataset, demonstrating its adaptability and efficiency across different scenarios. This approach enhances LiDAR sensors’ utility in supporting AI-enhanced decision-making processes and enriching the knowledge base of incorporating LiDAR sensors into current and prospective traffic management strategies.

Index Terms—Hashing algorithm, Gaussian mixture model, smart infrastructure application.

I. INTRODUCTION

LiDAR (Light Detection and Ranging) has excellent potential to be a beneficial unit of infrastructure sensor networks. For intersection analytics, where most crashes happen, a single LiDAR device could cover a wide area with accurate 3D distance measurement of the entire intersection. While it usually needs to have multiple cameras/radars to monitor one intersection from different angles. The LiDAR solution does not incur privacy concerns; therefore, it is suitable for privacy-sensitive applications, such as crowd management and vehicle tracking. LiDAR sensors can function in various environments, unlike the camera sensor, which relies on external illumination. The data collected through LiDAR could be used to optimize signal control and reduce delay and emission, as well as people counting for event planning. Combined with the Vehicle to Infrastructure (V2I) communication, the 3D data could be fed into many connected vehicle applications, such as Red-Light Violation Warnings and Queue Warnings (Q-WARN), to reduce congestion and prevent probable collisions. As a key perception component of autonomous vehicles, LiDAR technology evolves rapidly. Thanks to the industrial and academic communities’ enthusiasm to self-driving technology, it will soon become a worthy investment.

Difficulties arise, however, when applying roadside LiDAR for collecting high-resolution data. One of the biggest challenges is the large volume of 3D data that are too sparse and disordered to process. This leads to the non-negligible cost of data transmission and processing. Each LiDAR device produces millions of data points per second, a significant barrier to raw data handling and computing. For roadside applications, a more effective approach is to separate backgrounds and foregrounds and only keep the targeted objects. The existing roadside LiDAR background filtering methods were designed from engineering viewpoints according to aggregated descriptors, which are often sensitive to parameter settings, such as voxel size, number of channels, density, and distance thresholds. From the theoretical perspective, many well-developed dynamic background models for pixel images are undervalued and considered unsuitable for LiDAR point clouds. In previous studies, the LiDAR sensor was mainly used to collect rudimentary traffic parameters, such as speeds, volumes, positions, and object types. The advantages of the infrastructure LiDAR sensor has yet to be fully demonstrated.

This research resorts to hash-based transformation and successfully adopted the probabilistic Gaussian Mixture Model (GMM) methods to the roadside LiDAR application, showing that useful patterns and modalities were preserved. Many combinations of variables were investigated based on AIC and BIC scores to determine the optimal features and number of Gaussian components. The proposed Hash-based Gaussian Mixture Modeling (HGMM) allows high-dimensional multivariate models for processing point cloud data, making LiDAR background subtraction as efficient as the video data. This study adds to the growing body of research to demonstrate...
that the LiDAR sensor is not just an auxiliary device for information redundancies. Instead, it can upgrade traditional detectors in many smart infrastructure applications.

II. RELATED WORK

A. LiDAR Object Detection

1) Supervised Learning Methods: Recent years have already seen many effective deep learning-based LiDAR object detection models developed for self-driving vehicles. Depending on the data representation methods, the deep learning models can be classified into the Spherical projection method (SqueezeSegv1/v2 [1], [2] and RangeNet++ [3]), Voxel-based method [5], [6], Point-based method [7], Frustum-based method [8], Pillar-based method [9], and 2D Bird-Eye-View (BEV) projection-Based Method [4], [10]. Another way to classify the LiDAR object detection model is based on whether it uses the two-stage region proposal network or a one-stage framework [11]. For example, the Yolov3d model builds a 3D object bounding box detection method on the Yolov2 image-based one-shot regression meta-architecture [14]. PointR-CNN [15] is a typical region proposal network that first generates a small number of high-quality 3D proposals to segment foreground objects and then refines the 3D proposal in the second stage to obtain final detection results. To address the costly labeling process of supervised learning, a weekly supervised framework [16] for 3D point cloud object detection and annotation is proposed. This weekly unsupervised framework only takes simple clicks to label the object center on BEV and then uses the BEV detection to reconstruct 3D parameters with a cascade network. Zhou et al. [17] developed a practical framework that outputs both 3D Instance Segmentation and Object Detection results with Spatial Embedding by considering both the global Bounding Box and local point information. By considering temporal information in consecutive frames, Graph Neural Network and Spatiotemporal Transformer Attention are applied for 3D object detection from point clouds [18]. Chen et al. [19] used range images generated from current and past 3D LiDAR scans together as inputs and outputs for labels in the current range image indicating moving objects or not. Li et al. [20] proposed a one-stage anchor-free 3D LiDAR vehicle detection framework consisting of a voxel/pillar feature extractor, backbone encoding layers, and detection head layers. In this paper [21], a modified convolutional architecture adds dense connections to the convolutional layers for better feature extraction. This allows the deep learning model for roadside LiDAR applications after training on an autonomous driving dataset.

2) Unsupervised Learning Methods: Given the stationary scenario of roadside LiDAR, many existing roadside LiDAR object detections are through background-filtering-based moving object detection. A coarse-fine triangle algorithm (CFTA) is devised to automatically identify the range threshold between moving foregrounds and distant backgrounds with histogram analysis [22]. A simulation-based research is conducted to quantitatively determine the optimal intallation angle of roadside LiDAR [23]. A 3D-density-statistic-filtering (3D-DSF) model [24] divides the 3D space into discrete cubes, which was selected as baseline model and implemented in the point-level evaluation section. An image-based method [25] projects LiDAR onto the X-o-Y plane and use the image-based background model to detect moving objects, which is at the cost of missing the Z-value of 3D measurements. Based on the operational principle of the LiDAR device, discrete horizontal and vertical angular values were considered as coordinates of pixels in the image matrix, and the mean and maximum distance for each azimuth were used to construct background references [26]. This paper [26] was similar to our approach and selected as one of the baseline models. In the reference [27], the authors developed a height azimuth background construction method by filtering out background points based on elevation changes. Another Ground plane removal method [28] was applied for vehicle detection and classification system. Their model includes preprocessing, outlier removal, ground plane segmentation, vehicle clustering, key points pair extracting oriented bounding box (OBB), and tracking. A vehicle speed estimation method [29] is explored through background removal, moving point clustering, and vehicle classification, which is based on the max-distance method, assuming that the static environment is the furthest point of each laser beam. An unsupervised point cloud pre-training framework ProposalContrast [30], is proposed that learns robust 3D representations by contrasting region proposals. This framework optimizes inter-cluster and inter-proposal separation to utilize the discriminativeness of proposal representations across semantic classes and object instances.

B. Problems

Many deep learning models have been developed in the context of self-driving scenarios, due to its great feature learning capability. Recently, the roadside LiDAR began to gain momentum as a new measure of traffic data acquisition for sensor fusion, safety analysis and connected vehicle applications. Compared to those models developed for a robot to explore and understand its ever-changing environment, the roadside LiDAR mainly detects moving objects in a fixed setting.

Understanding the unique characteristics of the stationary LiDAR sensor is vitally essential before diving into the methodology part. First, LiDAR points exhibit the radiate property, as decreasing point density with increasing distance from the LiDAR sensor. As the laser beams extend further, the number of points captured per unit area decreases, resulting in sparser point clouds. Therefore, voxel-based method leads to excessively memory space of empty 3D cubes and wasted computations. Second, The stationary background objects, despite their immobility, can appear to tremble or exhibit apparent movement in the LiDAR data. The angular values of the same laser beam between consecutive frames often experience drift. These drifts and offsets need to be addressed to accurately model and differentiate between stationary background objects and actual moving targets in LiDAR-based applications. Last but not the least, raw LiDAR data is unstructured, meaning that shuffling the 3D points does not change.
the data. In comparison, the image is structured where adjacent data points depend on each other, often modeled as the Markov Random Field (MRF). Since image-based background modeling requires structured input, previous studies could not apply the per-pixel Gaussian Mixture Modeling to each LiDAR point.

III. METHODOLOGY

LiDAR emits multiple laser beams as a modulated laser system, which can be viewed based on spherical coordinates. This section will transform the 3D LiDAR data into structured representations using hashing functions to deal with collisions when discretizing the angular values in 3D space. Then, the new LiDAR data will be fitted with Gaussian Mixture Models for multi-modal background modeling at each elevation-azimuth grid. Finally, we consider various combinations of variables to determine optimal features and numbers of Gaussian Mixture components based on information theory.

A. Hash-Based Point Cloud Transformation

Raw LiDAR data contains 3-dimensional measurements, and intensity values. Unstructured LiDAR point clouds need to be transformed into structured representations to enable convolutional or matrix operations. According to how to deliver the laser beam, LiDAR can be categorized as non-scanning LiDAR or scanning LiDAR. Non-scanning LiDAR is also called Flash LiDAR. Non-scanning LiDAR is constrained by the size and density of detector arrays and leads to a limited detection range. Scanning LiDAR contains rotary parts, classified as non-mechanical scanning (Optical Phased Arrays (OPA) scanners) and mechanical scanning LiDAR. (see Figure 1).

Due to factors like multiple returns from a single pulse, different LiDAR points could be mapped into the same spherical angular grid. Here we defined collision-aware hashing functions to convert LiDAR points from Cartesian coordinates into spherical coordinates. For a given LiDAR point with a spherical coordinate ($\alpha$, $\beta$, Range), we use the following Equation 1 and Equation 2 to map it onto the index of the elevation-azimuth grid ($H(\alpha)$, $H(\beta)$).

\[
H(\alpha) = \text{Round}\left(\frac{\text{FoV}_h}{\text{Azimuth Resolution}} + \alpha\right)
\]

\[
H(\beta) = \text{Round}\left(\frac{\text{FoV}_v}{\text{Elevation Resolution}} + \beta\right)
\]

where $\text{FoV}_h$ is the horizontal field of view. $\text{FoV}_v$ is the vertical field of view. $\text{FoV}_h$ and $\text{FoV}_v$ are used to shift the angular value to be positive. If two points collide into the same grid, we preserve the point with a smaller range, as the static backgrounds are often farther than the more informative foreground objects. In a uniform beam configuration, beams have equal angular spacing throughout the sensor’s vertical field of view. In contrast, a gradient beam configuration packs beams closely near the horizon for higher resolution, while beams towards the top and bottom are more spaced out. For gradient Velodyne LiDAR, the vertical angular is determined by the beam ID, which only need to calculate the azimuth grid using horizontal angular value. The default horizontal resolution are set as 0.2°. Using higher resolutions, like 0.18° or 0.1°, usually leads to fewer collisions and better background filtering results. This approach allows for a finer level of detail in representing the LiDAR data, reducing the likelihood of points colliding in the same grid cell. After applying the hashing algorithm, LiDAR points were stored in a multi-variable format to store X-Y-Z 3D measurement, range, and intensity for each LiDAR point. The final LiDAR data has the size of (Azimuth Grids * Elevation Grids * Number of Variables) for each data frame. (See Figure 2).

B. Multimodal Gaussian Mixture Model

The GMM models [31], [32], [33], [34] have been developed for decades and are implemented with a lot of skills and experience for satisfactory performance. Compared to the camera sensor, the LiDAR sensor does not experience moving shadows, camouflage (foregrounds have similar colors to the background), sudden illumination changes, and challenges like that. The GMM method is an excellent method for handling dynamic backgrounds. The probabilistic GMM model with K components that learns the background subcomponents is described as follows.

\[
p(x_i|\theta_1, \ldots, \theta_T) = \sum_{t=1}^{T} \pi_t \cdot N(x_i|\mu_t, S_t)
\]  

The index $i = 1, \ldots, n$ represents observations. The index $t = 1, \ldots, T$ represents components. The parameter $\theta_t$ for the $t^{th}$ component is defined as $\theta_t = (\mu_t, S_t, \pi_t)$. $\mu_t$ is the mean vector of the $t^{th}$ component. $S_t$ is the inverse covariance matrix of the $t^{th}$ component. $\pi_t$ is the mixing proportion for the $t^{th}$ component.

The probability of a given data point belonging to the background ($\theta$) is through marginalizing all mixture components.
components $T$:

$$P(x|B) = \sum_{t \in T} P(x|t, B)$$  \hfill (4)

In this equation, $P(x|B)$ represents the probability of a given data point $x$ belonging to the background (B). It is obtained by marginalizing over all mixture components $t$.

Then we can apply the Bayesian rule for background (B) /foreground (F) classification by computing the probability of sample $x$:

$$P(B|x) = \frac{P(x|B)P(B)}{P(x|B) + P(x|F)}$$  \hfill (5)

Based on the reference [35], $P(x|F) = 1$ is considered a uniform distribution, $P(x|B)$ is obtained from equation (4), and $P(B)$ is the predefined threshold value.

Bayesian non-parametric methods like Dirichlet Process Gaussian Mixture Models (DPGMM) are gaining popularity because of automatic model selection, but this method also come with increased computational costs. Given the real-time and continuous nature of LiDAR streaming datasets, the non-parametric model may not be suitable due to its processing inefficiencies and the necessity for instantaneous analysis in many LiDAR applications. The traditional parametric approach with explicit model selection and post-processing can achieve real-time processing speed with desirable accuracy.

C. Model Selection

As the LiDAR data contains rich spatial information, including 2D plane ($x$, $y$), elevation $z$, and intensity measurements ($Int$). The range information ($r$) can be easily derived from 3D measurements ($x$, $y$, $z$). This gives us five variables to choose from and provides opportunities to form different combinations. For instance, we can use ($x$, $y$, $z$) three variables, or only range ($r$) variables, or append intensity values to distance measurement to build two different models ($x$, $y$, $z$, $Int$) and ($r$, $Int$). This flexibility demands to be examined to balance the efficiency and accuracy of the number of variables. Some environments may naturally exhibit a higher level of complexity, requiring a larger number of components to capture the underlying distribution accurately. In contrast, simpler scenarios with fewer modes may require fewer components. The underlying structure and pattern of LiDAR data could be explored with different numbers of Gaussian components.

Two commonly used and well-known criteria, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) were applied to address the variable combinatorial issue and choose the optimal number of components. AIC and BIC metrics are based on the maximized log-likelihood value with a tradeoff of the number of estimators, expressed as Equations (6) and (7).

$$AIC = 2k - 2 \ln(\hat{L})$$  \hfill (6)

$$BIC = k \ln(n) - 2 \ln(\hat{L})$$  \hfill (7)

where $\hat{L}$ is the likelihood function of the candidate model, and $k$ is the number of model parameters, $n$ is the number of observations. In this case, the $n$ is the number of data frames used for model selection.

In our testing, the AIC and BIC scores are consistent on the model selection results. Figure 3 plots the BIC scores for different combinations of variables and the number of Gaussian components. The larger number of Gaussian components has a better fitness score. The elbow at Gaussian Component 2 indicates that the infrastructure LiDAR data demonstrates multimodal properties. The best model is chosen using the combination of X-Y-Z and range information. Adding the intensity of Lidar points could introduce ambiguities and requires more Gaussian components to capture the variances because the intensity values of point clouds are not as discriminative as the distance measurements. Using only range information gives lower performance, which explains why a single indicator/descriptor is not satisfactory for roadside LiDAR background modeling. After model evaluation, the XYZ + Range (XYZR) combination is selected as the optimal choice. As the HGMM model offers flexibility, Hybrid methods that combine optimal (XYZR) and suboptimal (XYZRI) will boost performances.

IV. EXPERIMENTAL SETUP

Two real-world Smart City LiDAR datasets in the US were selected to test and validate the proposed method. The first roadside LiDAR testing dataset is from eight sites of the planned 2.1-mile DataCity - Smart Mobility Testing Ground (SMTG) by Rutgers University’s Center for Advanced Infrastructure and Transportation (CAIT) to establish a living laboratory for connected and automated vehicle technologies in New Jersey. Eventually, multiple locations from two instrumented corridors will be equipped with high-resolution roadside sensors (128-beam Velodyne Prime LiDARs and 2.7k Cameras) and edge computing devices to enable innovative mobility applications. Those testing sites are selected with affinity to many restaurants, banks, and a university campus or monitoring a busy freeway. (See the urban arterial (Route 27) and state highway (NJ-18) settings in Figure 4).

The second testing dataset is from TRB 2024 Transportation Forecasting Competition (TRANSFOR 24) collected at City of...
Chattanooga’s Smart Corridor. The TRANSFOR 24 dataset is designed for transportation forecasting with a focus on the safety and protection of Vulnerable Road Users (VRUs), such as pedestrians and cyclists. It integrates LiDAR and video data, offering a comprehensive view of the environment. The LiDAR data, with 32 channels, are instrumented with the detection range of 0.05 to 120 meters. Additional video footage offers visual context, helping to interpret behaviors and validate LiDAR observations and develop and test advanced sensors and algorithms. (See the Chattanooga’s Intersection layout in figure 4).

V. MODEL VALIDATION

This section will completely evaluate the HGMM model’s efficacy using manually processed ground-truth data at point level. This evaluation aims to quantify the accuracy of the background/foreground segmentation for each LiDAR point. Each point from background filtering results and ground truth data can be classified as a true positive, true negative, false positive, or false negative. Three performance metrics were used, including Precision, Recall, and F1 score, which were defined in the following equations.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (9)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (10)
\]

The Weighted F1 Score, which takes into account the number of samples in each class, is defined as:

\[
\text{Weighted F1 Score} = \sum_{i=1}^{N} w_i \times \text{F1 Score}_i \quad (11)
\]

\[
w_i = \frac{\text{No. of samples in class } i}{\text{Total number of samples}} \quad (12)
\]

A. Comparison to Baselines

Three types of roadside LiDAR background filtering methods were implemented. Their segmented foreground points after background subtraction are presented in Figure 5.

1) Statistical Baselines: This type of method uses aggregated frames to construct the background reference by analyzing statistics such as mean, variance, and frequencies. The Coarse-Fine Triangle Algorithm (CFTA) [22] employs a histogram-based triangle threshold to automatically determine the dividing value between foreground targets and background objects. The second statistical baseline model, named as Mean-Max model [26], utilizes a two-step approach to establish the background range for each elevation-azimuth value, using the mean and maximum range values calculated from continuous frames. Both baseline methods are designed to automatically find the thresholds that distinguish backgrounds from foregrounds based on aggregated frames.

2) Change Detection Baselines: The second type of baselines focuses on point cloud change detection, using a
TABLE I
PERFORMANCE METRICS FOR DIFFERENT METHODS ACROSS DIFFERENT TIME PERIODS

| Method         | AM Peak     |           |           |           |           |
|----------------|-------------|-----------|-----------|-----------|-----------|
|                | Accuracy    | Precision | Recall    | F1 Score  |           |
| CPTA           | 99.05%      | 97.23%    | 98.70%    | 97.81%    |           |
| HGMM           | 99.63%      | 99.60%    | 98.54%    | 99.01%    |           |
| MeanMax        | 96.54%      | 93.62%    | 95.44%    | 93.74%    |           |
| K-D Tree       | 99.26%      | 97.87%    | 98.92%    | 98.26%    |           |
| Octree         | 99.46%      | 99.04%    | 98.19%    | 98.57%    |           |
| PointSeg       | 95.48%      | 91.32%    | 91.64%    | 91.18%    |           |
| SqueezeSegV2   | 96.89%      | 90.55%    | 90.35%    | 90.01%    |           |
| PM Peak        |             |           |           |           |           |
|                | Accuracy    | Precision | Recall    | F1 Score  |           |
| CPTA           | 98.58%      | 97.17%    | 98.36%    | 97.61%    |           |
| HGMM           | 99.01%      | 98.95%    | 97.34%    | 97.93%    |           |
| MeanMax        | 95.92%      | 94.31%    | 95.61%    | 94.16%    |           |
| K-D Tree       | 98.69%      | 97.48%    | 98.38%    | 97.80%    |           |
| Octree         | 99.07%      | 98.63%    | 97.89%    | 98.21%    |           |
| PointSeg       | 95.18%      | 91.64%    | 91.72%    | 91.63%    |           |
| SqueezeSegV2   | 96.06%      | 90.82%    | 90.52%    | 89.93%    |           |

pre-recorded LiDAR frame captured at midnight with no traffic to filter out background points. To enhance algorithm efficiency, K-D tree and Octree data structures are utilized. A K-D tree is a binary search tree that organizes points in k-dimensional space to efficiently support range searches and nearest neighbor searches. The Octree change detection method voxelizes the point cloud and background into two separate Octree structures, which allows for the efficient searching of points stored in the point cloud Octree that do not exist in the background Octree.

3) Deep Learning Baselines: Two well-known semantic segmentation models, SqueezeSegV2 [2] and PointSeg [41], were implemented. Both deep learning models use the same data transformation strategy, converting point clouds in spherical coordinates and concatenating x-y-z-range-intensity channels into one multidimensional array. The training data preparation process involves applying and merging the above-mentioned unsupervised background filtering methods to obtain initial foreground labels, followed by a manual process to remove residual noise. A total of 1800 LiDAR frames were labeled using data from 9 LiDARs at different locations (7 LiDARs from New Jersey DataCity SMTG and 2 LiDARs from Chattanooga). Additionally, 20 frames in AM Peak, 20 frames in Mid-Day and 20 frames in PM Peak were prepared as testing dataset.

As shown in TABLE I, our model achieved the best Accuracy and F1-Score performances among all the implemented methods at AM Peak and PM Peak hour. During midday with less traffic, the proposed HGMM is surpassed by change detection baselines but shows better results than statistical and deep learning baselines. All models’ precision is very high because the background and foreground point clouds are highly imbalanced and the majority of background points can be filtered out. The background objects account for over 90% of the total point clouds (Figure 6). Compared to statistical methods, our approach requires significantly fewer frames (less than 300 frames) to learn the background. It also offers flexibility with different combinations of features, resulting in improved accuracy and performance measures. In comparison to change detection baselines, our proposed HGMM model, despite retaining more background points, better at preserving small objects with fewer points. Our method is adaptable and automatic, unlike change detection methods that necessitate manual background selection. Additionally, LiDAR data often drifts over time, altering backgrounds in ways that can undermine the effectiveness of change detection methods. When compared to deep learning baselines, our approach not only achieves better accuracy but also operates more efficiently as an unsupervised learning method, requiring no training data. It demands fewer computational resources, making it easier to align with additional sensors, such as cameras or radars, for additional fusion approach. Our method demonstrates low complexity, robustness, and explainability, working well with fixed-position Lidar datasets, as evidenced by quantitative assessments conducted under different time-of-day scenarios.

B. Discussion

The HGMM method offers stable and rapid adaptation to dynamic environments for foreground segmentation using infrastructure LiDAR. Unlike existing angular value projection methods, our approach uses hash functions that consider collision probability. Traditional spherical angular projection methods fix the angular resolution, while in HGMM, angular resolutions are adjustable parameters that can be calibrated for finer representation and improved performance. Since LiDAR data comes with varying densities and channels, pre-trained deep learning models, which typically require input data with fixed dimensions, often struggle to generalize across different settings. Particularly, LiDAR sensors vary in range, number of beams, and resolution. HGMM addresses this issue by providing a plug-and-play solution that works with any resolution and brand of LiDAR, offering superior flexibility and adaptability. The HGMM method does not rely on sensor specifications, significantly reducing the complexity of integrating LiDAR with existing infrastructure.

LiDAR stands out in capturing spatial features due to its precision in measuring distances and creating detailed 3D models of its surroundings. However, it traditionally struggles with semantic features because LiDAR data alone does not tell what these objects are; it only shows where they are and their shape. For example, LiDAR can tell that there’s an object at a certain location and describe its dimensions, but it won’t directly indicate whether that object is a person, a street sign, or a vehicle. When deep learning models process LiDAR data, they face challenges in differentiating between objects of interest (like pedestrians or vehicles) and less relevant objects (like buildings and trees) with only distances and dimensions. This lack of additional semantic information can undermine the deep learning model’s performance, as the model may
struggle to interpret what each object is based solely on the point cloud measurement data.

VI. SMART INFRASTRUCTURE APPLICATION

This segment explores the various applications of LiDAR sensors and shows how the proposed HGMM method facilitates an assortment of smart infrastructure applications.

A. Object Detection and Tracking

After background filtering, a density-based bounding box estimator was used to obtain instances from the foreground points. In Figure 7, TRB TRNASFOR 24 LiDAR data is used to show the foreground points clustering and classification results, including human (in cyan), car (in red) and truck/bus (in blue).

The trajectory data can be obtained with bounding box detections for continuous LiDAR frames. Each detected object was encoded into a state space with 3-dimensional coordinates, angles, and speeds. The object classification is based on detection bounding box 3D measurements, height-to-length ratio, and traveling speeds. Then the JPDA-IMM-UKF algorithm [36], [37] was implemented to obtain path-level information for the multi-object tracking task. This JPDA-IMM-UKF algorithm is derived from the Unscented Kalman Filter (UKF) and resolves the data association issue by combing it with the Joint Probabilistic Data Association (JPDA). The algorithm can efficiently overcome the hit-and-miss problem when tracked vehicles are temporarily obstructed.

This trajectory-level outcome will help integrate the infrastructure LiDAR into connected intersection applications. Figure 8 plots the time-space diagram with high-resolution (0.1s) vehicle trajectory for the westbound approach from the
Fig. 7. Object detection and classification with TRANSFOR24 LiDAR data (Cyan: Human; Red: Car; Blue: Bus/Truck).

Route-27 & George intersection from New Brunswick Testbed to show the cycle-by-cycle traffic flow patterns. We defined the entry location of westbound lanes as the reference location and calculated the relative distance of vehicle position at each timestamp. The time-space diagram is an essential analytic tool for understanding the traffic flow, especially useful for shockwave discussion, queuing formation, and dissipation.

The challenges in point cloud object detection and tracking algorithms are predominantly due to blind zones, occlusions caused by external factors like buildings or other vehicles, and signal miss issues arising from the material properties or reflection angles of objects [23], [38]. However, potential solutions such as sensor fusion and point cloud completion techniques can mitigate these issues. Future advancements in sensor technology, such as the integration of multiple LiDAR sensors, along with ongoing research in estimating and filling in missing data points, hold promise for further improving object detection and tracking accuracy in challenging scenarios with occlusions and missing data.

Fig. 8. Cycle-by-cycle time-space diagram from roadside LiDAR vehicle trajectory.

Fig. 9. Using LiDAR as stopbar detectors for signalized intersection.

Fig. 10. Roadside LiDAR for traffic signal performance measurements (Phase 2: Northbound; Phase 6: Southbound; Phase 3: Westbound; Phase 4: Eastbound).

B. Signal Performance Measurements

LiDAR sensor is not only an add-on to current traffic detectors but could also become an alternative to upgrade those legacy detection systems. Figure 9 shows that the LiDAR...
sensor replaces the traditional stopbar detectors (Video or Inductive loop) for lane-by-lane stopbar detection from all approaches at a signalized intersection. During weekday rush hour, roadside detector data, CCTV video and signal phase and timing data (SPaT) were collected. With the moving object trajectory and bird-eye-view (BEV) projection, we can properly draw stopbar detection zones on a 2D-plane image and utilize LiDAR input for vehicle presence detection. Vehicle presence events can be recognized based on the segmentation of foreground cloud points within the detection zone. The On-Off detector state will be recorded, as well as the corresponding detector ID and timestamp for that event.

In the Figure 10, a novel signal performance metric Rutgers Stopbar Coordination Diagram (RCD) is created by combing SPaT data with stop bar detection from the LiDAR sensor. The RCD diagram visualizes important signal performance metrics, such as vehicle departure headway, startup delay, phase duration, volume, occupancy, etc. From the new coordination diagram, the split failure issue can be quickly spotted and diagnosed using time headway. A significant time gap usually implies that queuing vehicles were cleared. Therefore, vehicles arriving at green do not need to stop during that cycle. A split failure was considered if all gaps within one cycle were very small and almost constant. Namely, queuing vehicles for that approach are unable to clear during the green interval.

C. V2X Application

Smart infrastructure and V2X (Vehicle-to-Everything) applications utilize various sensors to compensate for individual limitations, offering a comprehensive view of surroundings and ensuring system reliability. Such detailed and diverse data is not only essential for effective V2X communication but also crucial for training advanced AI models in self-driving cars. The HGMM background filtering method can greatly improve the sensor data acquisition process and build the foundation for precision-timestamped, multi-modal data for V2X self-driving systems. In this section, we applied our algorithm on two large scale datasets that contains diverse road scenarios and were captured by cooperative multi-modal sensors.

- The DAIR-V2X cooperative driving dataset [39], collected in Beijing, China, utilizes infrastructural LiDARs with a spectrum that encompasses 300 beams and a capture frequency of 10Hz. It features a horizontal field of view (FOV) of 100 degrees and a vertical FOV ranging from -30 degrees to 10 degrees. The LiDAR can detect objects at a distance of up to 280 meters with an accuracy of $\pm 3\text{cm}$.
- The A9-Dataset [40], collected in Munich, Germany, utilizes the Ouster OS1-64 (gen 2). This LiDAR features 64 vertical layers with a vertical angular resolution of $0.26^\circ$ to $0.52^\circ$ and a horizontal angular resolution ranging from $0.18^\circ$ to $0.7^\circ$. It has a vertical field of view (FOV) of 33.1$^\circ$ and a full 360$^\circ$ horizontal FOV. The system can detect objects at a distance of up to 120 meters with an accuracy varying between 1.5cm to 10cm.

By applying the HGMM background filtering method to the DAIR-V2X and A9 datasets, substantial annotation errors were discovered for different classes of objects. Table II reveals these errors in detail. Despite these datasets being designed...
Unlike traditional point-wise or voxel-wise Cartesian ex-hibit distinct properties that impact their characteristics and ground modeling. 3D point clouds generated by LiDAR scans have not been effectively utilized for roadside LiDAR back-data distribution. However, these well-established techniques employed in various machine learning algorithms to model functioning, which rely heavily on accurate, real-world data for optimal advancement and safety of smart infrastructure technologies, V2X applications. Ensuring data integrity is crucial for the quality assurance issue when developing safety-critical applications, the HGMM managed to learn fairly accurate background frames. Despite these challenging conditions and limited samples, the HGMM managed to learn fairly accurate background models. Figure 11 demonstrates how the HGMM roadside LiDAR method can improve the data quality of existing V2X sensor datasets. The figure shows a visual comparison between a camera’s field of view and the corresponding point cloud model of an intersection scene for both DAIR-V2X (first row) and A9 (second row) datasets. The misalignment between the camera’s field of view and the LiDAR data can lead to a significant number of missed labels in the LiDAR dataset.

Validating large-scale labeled images and LiDAR point clouds of multiple road segments from different angles is very laborious and costly. Using HGMM can greatly speed up the data acquisition and labeling process and address the quality assurance issue when developing safety-critical V2X applications. Ensuring data integrity is crucial for the advancement and safety of smart infrastructure technologies, which rely heavily on accurate, real-world data for optimal functioning.

| TABLE II | Annotation Errors From DAIR-V2X and A9 Infrastructure LiDAR |
|----------|----------------------------------------------------------|
| DAIR-V2X-I Example Infrastructure LiDAR | | |
| Labeled | Missed | Total | Error Rate |
| Car | 1036 | 235 | 1281 | 18.20% |
| Motorcyclist | 118 | 0 | 118 | 0.00% |
| Pedestrian | 329 | 86 | 415 | 20.72% |
| Bus | 32 | 6 | 38 | 15.79% |
| Cyclist | 220 | 22 | 242 | 9.09% |
| Truck | 16 | 4 | 20 | 20.00% |
| Van | 52 | 3 | 55 | 7.14% |
| A9 KO_S4 Infrastructure LiDAR | | |
| Labeled | Missed | Total | Error Rate |
| Car | 680 | 555 | 1235 | 44.94% |
| Trailer | 355 | 171 | 526 | 32.51% |
| Van | 113 | 12 | 125 | 9.60% |
| Truck | 8 | 19 | 27 | 70.37% |

to serve as ground truth for detection and tracking algorithm development, a comparative analysis against the proposed background filtering output indicated notable omissions of moving traffic components. These inaccuracies highlight a significant quality issue within these datasets and emphasize the necessity for enhanced validation and annotation practices. As V2X infrastructure LiDAR datasets, DAIR-V2X provides around 150 frames per location, while A9 datasets provides around 200 frames per location. Noticeable labeling errors were found in almost every frame sample, and there are also obvious sensor movements in parts of their provided frames. Despite these challenging conditions and limited samples, the HGMM managed to learn fairly accurate background models. Figure 11 demonstrates how the HGMM roadside LiDAR method can improve the data quality of existing V2X sensor datasets. The figure shows a visual comparison between a camera’s field of view and the corresponding point cloud model of an intersection scene for both DAIR-V2X (first row) and A9 (second row) datasets. The misalignment between the camera’s field of view and the LiDAR data can lead to a significant number of missed labels in the LiDAR dataset.

Validating large-scale labeled images and LiDAR point clouds of multiple road segments from different angles is very laborious and costly. Using HGMM can greatly speed up the data acquisition and labeling process and address the quality assurance issue when developing safety-critical V2X applications. Ensuring data integrity is crucial for the advancement and safety of smart infrastructure technologies, which rely heavily on accurate, real-world data for optimal functioning.

VII. CONCLUSION AND OUTLOOK

Multivariate probabilistic methods have been widely employed in various machine learning algorithms to model data distribution. However, these well-established techniques have not been effectively utilized for roadside LiDAR background modeling. 3D point clouds generated by LiDAR scans exhibit distinct properties that impact their characteristics and analysis. Unlike traditional point-wise or voxel-wise Cartesian coordinate-based methods, which may suffer from the computational burden and sensitivity to parameter settings, this Spherical angular-based approach allows for efficient background modeling at the individual LiDAR point level, similar to processing pixels in an image. The proposed method retains the full information of the LiDAR data while avoiding the loss of information associated with 2D plane projections, ultimately improving computational efficiency and accuracy in LiDAR modeling.

In this study, HGMM was implemented on 128-beam LiDAR (New Brunswick DataCity dataset), 64-beam LiDAR (A9 dataset), 32-beam LiDAR (TRB TRANSFOR 24) and 300-beam LiDAR (DAIR-V2X). The foreground/background segmentation evaluation results demonstrate that the Hash-based transformation successfully preserves meaningful patterns and structures within the LiDAR data, enabling effective application of these probabilistic models for roadside LiDAR background modeling. This paper also applied Information Theory criteria to assess different combinations of measurement variables, which helps to understand each LiDAR feature’s effect. Several novel applications have been discussed to explore the versatility of LiDAR as a roadside detector, showing the advantages of LiDAR sensors compared to conventional traffic detectors.

Applying the proposed model could greatly improve the efficiency of point cloud processing. For infrastructure LiDAR, the object detection task only pertains to a small amount of data in a fixed environment. Therefore, background modeling can significantly improve the data chain efficiency by only transitioning a tiny portion of the foreground LiDAR point clouds. Raw data volume of 128-beam LiDAR is around 24 GB/hour and standard file compression tools such as 7zip and WinRAR could only have around 40% to 55% data reduction rate, resulting in 250 GB/day/LiDAR. In our testing experiment, 31.7 GB of raw point cloud data were compressed into 760 MB foregrounds and a background reference, which reduced over 90% redundancy in roadside LiDAR data. A compression rate of 90% from proposed background filtering algorithm will make our 1.5 PB storage node lasting more than 10 years for all instrumented sites.

This paper touched on critical issues of applying roadside LiDAR for smart infrastructure applications, ranging from low-level working principles & data representation to medium-level feature selection & modeling and, eventually, high-level application cases. These findings demonstrate how to integrate LiDAR detection into daily traffic operation strategies with moving object detection and data compression, movement-related information, and signal performance diagnosis. The algorithm of the HGMM model can be readily deployed to different scenarios and help transportation agencies upgrade traffic detection devices to empower the data-driven decision-making process. Future work could incorporate advanced instance segment and tracking to further refine the roadside LiDAR sensing performance [42], [43].

ACKNOWLEDGMENT

The views and conclusions contained in this document are those of the authors and should not be interpreted as...
Tianya Terry Zhang is currently a Research Professor with the University of Tennessee at Chattanooga, working in the multidisciplinary area of combining machine learning and sensors with connected and automated technology to build safe, green, and smart mobility applications. His first-author articles have appeared in leading journals, such as *Journal of Transportation Engineering—Part A: Systems*, *Transportation Research—Part C: Emerging Technologies*, *IEEE Transactions on Intelligent Transportation Systems*, and *Transportation Research Record*. His research interests include connected automated vehicle technologies, computer vision, traffic operations, traffic flow theory, and data sciences. In his previous research, a new framework named segmentation-is-tracking was devised for high-resolution vehicle trajectory extraction that can be used for traffic flow study and highway monitoring.

Yi Ge received the M.S. degree in transportation engineering from the Department of Civil and Environmental Engineering, Rutgers University, in 2020, where he is currently pursuing the Ph.D. degree in transportation engineering. He won the ITSNJ 2020 Outstanding Graduate Student Award and the ITSNJ 2021 Future of ITSNJ Award.

Mina Sartipi (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2001, and the M.S. and Ph.D. degrees in electrical and computer engineering from Georgia Tech in 2003 and 2006, respectively. She is currently the Founding Director of the Center for Urban Informatics and Progress (CUIP), University of Tennessee at Chattanooga (UTC), where she is also a Guerry Professor with the Computer Science and Engineering Department. Her research was funded by NSF, NIH, DOE, State of Tennessee, Lyndhurst Foundation, and other industry organizations, and focuses on data-driven approaches to tackle real-world challenges in smart city applications focused on mobility, energy, and health.

Anjiang Chen received the B.S. degree from the Department of Transportation, Beijing University of Technology, Beijing, China, in 2018, and the M.S. degree in civil and environmental engineering from Northwestern University, Evanston, USA, in 2020. He is currently pursuing the Ph.D. degree with the Department of Civil and Environmental Engineering, Rutgers, The State University of New. His research interests include traffic trajectory analysis, simulation, and traffic flow analysis.

Peter J. Jin received the B.S. degree from the Department of Automation, Tsinghua University, Beijing, China, and the M.S. and Ph.D. degrees from the University of Wisconsin-Madison, in 2007 and 2009 respectively. He is an Associate Professor at the Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey. He led the effort of building the 2.2-mile DataCity Smart Mobility Testing Ground in New Brunswick, NJ, USA. He has published over 60 journal papers and over 70 conference papers. His research interests include connected and automated vehicle technologies, transportation big data analytics, and unmanned aerial vehicle applications.