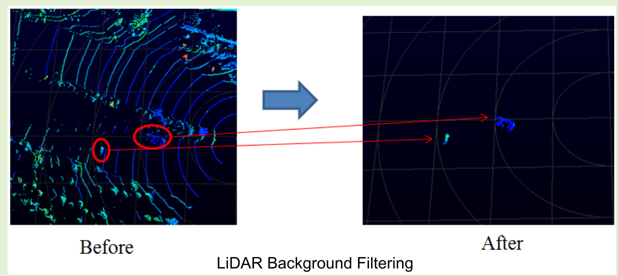
IEEE Sensors Council

An Automatic Background Filtering Method for Detection of Road Users in Heavy Traffics Using Roadside 3-D LiDAR Sensors With Noises

Yongsheng Zhang, Hao Xu, and Jianqing Wu

Abstract—**Background filtering for roadside Light Detecting and Ranging (LiDAR) data could benefit detection of road users with high frequency and accuracy. But existing methods are unsatisfactory under traffic congestions and neglect noises of LiDAR sensors. This paper attempts to develop background identification and filtering methods for heavy traffics and LiDAR noises. In particular, a 3-D space-time diagram of the LiDAR's detection space over frames is divided into plenty of subspace-frames, where one subspace-frame refers to one subspace from the isometric division of detection space in one frame. For subspace-frames at the same position, an unsupervised clustering method with unknown number of clusters is applied for classifying subspace-frames into different groups and the background group is identified based on background characteristics. The proposed clustering method is a *t*-distribution mixture model with constraints, where weights follow a Dirichlet distribution so that some weights have the possibility to be 0. A Bayesian estimation framework is developed to carry out the clustering. Then each background subspace-frame is divided into cubes with equal lengths. By aggregating points over frames in the same cube, the cube with aggregated points is regarded as background cube. For background filtering, points in background cubes are background points and excluded. Finally, examined by data from a Velodyne VLP-32 LiDAR sensor, the proposed approach has better filtering results no matter under free-flow conditions or congestions. Using background filtering results for road users' detection is introduced as well.**



Index Terms—**Roadside LiDAR sensor, background filtering, detection of road users, unsupervised clustering, Bayesian estimation framework.**

I. Introduction

CONNECTED vehicle (CV) system is expected to be the solution to significantly improve traffic efficiency and safety. Since it will eventually come true, we should be ready for it. Generally, in a whole CV system, various sensors and communication technologies will be integrated to connect all road users on roads, including on-board, vehicle-to-vehicle, vehicle-to-infrastructure, vehicle-to-internet, etc. However, currently, connected or autonomous vehicles (AVs) are extremely less compared to unconnected vehicles. Even in the future several decades, there might still be mixed traffics with unconnected and connected vehicles on roads [1]. Thus, it is a good solution to utilize road-infrastructure-based sensors to actively detect unconnected vehicles with high resolution and accuracy and broadcast related information to all road users.

The Light Detection and Ranging (LiDAR) sensor which has been verified by AVs [2] has attracted more and more attention in recent years to roadside applications due to its perfect performances in providing 3-D point clouds of road users with high resolution and accuracy in the night as

Manuscript received January 5, 2020; revised February 17, 2020; accepted February 20, 2020. Date of publication February 28, 2020; date of current version May 15, 2020. This work was supported by the Young Scholar Program of Shandong University. The associate editor coordinating the review of this article and approving it for publication was Dr. Yen Kheng Tan. *(Corresponding author: Jianqing Wu.)*

Yongsheng Zhang is with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China (e-mail: zys8303@gmail.com).

Hao Xu is with the Department of Civil and Environmental Engineering, University of Nevada at Reno, Reno, NV 89557 USA (e-mail: haox@unr.edu).

Jianqing Wu is with the School of Qilu Transportation, Shandong University, Jinan 250100, China (e-mail: jianqingwusdu@sdu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2020.2976663

perfectly as during the daytime. But the LiDAR is usually more expensive than commonly used sensors, like camera and radar. For roadside applications, cost-efficient LiDAR sensor with fewer channels than on AVs is usually used. Though sparse data points will be obtained from cost-efficient LiDAR sensors, the density of data points is enough for detecting and tracking all road users. Each type of sensors has its own advantages and disadvantages. An integration of multiple sensors will be expected in the future [3]. But this paper just focuses on LiDAR sensors, since related research is relatively less, compared to vision-based research.

Roadside LiDAR means the sensor is deployed in a place to monitor the region of interest for a long time, including fixed and portable deployments [4]. Fixed (or stationary) deployment [5] means LiDAR is installed roadside permanently, such as installing on a pole. Portable deployment [6], [7] means the sensor is temporarily installed on such as tripods, parked vehicles, etc. Fixed deployment is more stable for long-term traffics, whereas portable deployment is more flexible for observing short-term traffics. Both deployments can collect LiDAR-point based road users mixed with backgrounds. In order to obtain road users, it is essential to separate road users and backgrounds. That is the so-called background filtering or road users' detection. Since the roadside LiDAR sensor can detect the same place for a long time so that amounts of data over time can be gotten, the background filtering method involving learning background-related information from huge frames of historical data is direct and attractive. However, some challenges still remain with requirements of high-accurate and automatic background filtering results.

Background points include dynamic and static points. Dynamic points refer to the points possibly waving in different timestamps due to winds, such as tree branches, bushes, grasses, etc. Usually, the waving range is not too large. Those dynamic backgrounds sometimes are very near to pedestrians or fully/partially block pedestrians on sidewalks, leading to great influences on pedestrians' detection. Static points refer to points relatively fixed over time, such as buildings, ground lines, etc. However, due to the vibration of LiDAR sensor, even for static points, they will fluctuate in a very small range over frames. Noises caused by dynamic background points and the vibration of sensor will deeply affect the performance of separating backgrounds and road users. Especially for portable deployment, relatively unfirmed installation will induce more noises. Existing manual background filtering methods [6]–[8] which manually select several frames without road users as backgrounds will be affected heavily by noises.

Another challenge of background filtering for roadside LiDAR data is that misidentifications of vehicle points as backgrounds will increase via existing automatic background filtering methods [9] under heavy traffic conditions or at busy intersections. Existing automatic methods mainly utilize differences of point density between the space passed by moving vehicles and spaces containing backgrounds to separate road users and backgrounds. But under heavy traffic conditions or at busy intersections, the difference may not be significant enough, resulting in bad performance of background filtering.

Apparently, automatic methods which don't require selected frames or pre-defined region of interests as inputs are more attractive compared with manual methods. This paper pays attention to developing an automatic background filtering method for detecting all road users within the detection range of roadside LiDAR sensor. In particular, by mining amounts of frames of LiDAR data, a general automatic background identification model based on an unsupervised clustering method is developed for the whole detection range during any time period and under any traffic condition. Then a real-time automatic background filtering method is proposed based on background identification results to detect road users frame by frame. Finally, the proposed approach is evaluated based on data collected at intersections in Reno, NV, United States. The contributions are highlighted as follows.

1) A new framework of automatic background filtering for roadside 3-D LiDAR data with noises is proposed, where an unsupervised clustering method is integrated for background identification under any traffic condition without manual inputs.

2) An automatic background identification method is proposed for roadside 3-D LiDAR data by developing a t-distribution mixture model (TMM) with constraints for unsupervised clustering, where the weights follow a Dirichlet distribution for unknown number of clusters.

3) A Bayesian framework is developed for estimating the proposed clustering model, where Lagrangian multiplier method is utilized for integrating TMM and constraints and Metropolis-Hastings algorithm is utilized for sampling.

The following sections are organized as follows. Section 2 reviews existing research. In Sections 3 and 4, the methodology and algorithm of background identification and filtering are exhibited in details. Then, numerical results are analyzed in Section 5. At last, Section 6 concludes this study.

II. Literature Review

The future CV system expects that all road users, infrastructures and related departments are able to have real-time, highly efficient and accurate communications with each other. But there are great gaps between current situations and future expectations. One of the gaps is that not all road users will be equipped with related sensors and connected to the system in the near future, leading to mixed traffics on roads, that is connected and unconnected road users [1]. Even though there are onboard sensors which can be mounted on vehicles, such as autonomous vehicles, to detect unconnected road users, they can be seldom found on roads currently. To bridge this gap to a great extent, utilizing roadside infrastructures which can actively detect all road users and broadcast traffic information through wireless communication technologies to all road users and departments will be a better solution in the near future [10]. Thus, in order to get real-time high-resolution micro traffic data of all road users (i.e. connected and unconnected road users) at the same time, roadside sensors are expected to be employed. Among various traffic sensors, LiDAR sensor which is similar to radar, but can detect smaller objects with higher accuracy and frequency will be a good option by installing it roadside. Moreover, compared to camera data for road users detection and tracking [11], [12], LiDAR data has the advantage of not being affected by daylights and providing more accurate 3-D points instead of 2-D pixel-based data [13], [14].

A. Research Based on Different Sensors

In the literature, much background filtering or road users detection related work has been done for airborne (or side-fire) LiDAR, onboard (or mobile) LiDAR and non-LiDAR data respectively. For the airborne LiDAR data, the background usually consists of evenly-distributed high-density ground surface points [15]. Related background filtering methods include slope-based methods [16], surface-based methods [17], segment-based methods [18], and others. But those methods don't consider other background points like buildings and vegetation which have different characteristics to ground surfaces. In terms of onboard LiDAR data, backgrounds change with the movements of sensors [19]. Compared to roadside LiDAR, onboard LiDAR is required to have higher density of points. Higher density of points usually means higher price per LiDAR unit. Moreover, onboard LiDAR has a smaller detection range and is easier to be blocked by other vehicles, leading to the difficulties in detecting vehicles far away and pedestrians on sidewalks. Commonly used onboard LiDAR-based methods for detection of road users include patch segmentation and classification method [20], deep-learning method [21], voxel-based method [22], etc. Those methods require much more computational efforts, but have lower accuracy than roadside LiDAR-based methods. The roadside LiDAR sensor is expected to be cheaper and have wider detection range to detect no matter vehicles on lanes or pedestrians on sidewalks. The cheaper sensor certainly means fewer laser beams, which will result in sparse LiDAR points of objects, especially when the object is far from the sensor. Thus, the special characteristics of roadside LiDAR data different from mobile LiDAR data should be utilized for background filtering. With respect to non-LiDAR data, many methods can be found, especially for video or image data [23], [24]. But vision-based detection methods cannot be directly used for 3-D point clouds generated by LiDAR sensor, since vision-based methods usually utilized the grayscale variation over a 2-D image for detection, such as HOG-based methods [25], mixture of Gaussian method [24], etc. And the 2-D image projected from a 3-D scene will lose sophisticated spatial structure information, leading to a poor performance of 3-D object detection in pixel-based image. Although RGB-D cameras can be used to collect 2-D images with depth information and some related research have been done [26]–[28], differences between pixel-based image and point clouds still exist. Therefore, the method of background filtering for roadside LiDAR data should be studied specially.

B. Research Based on Roadside LiDAR

Roadside LiDAR sensor can scan the surface of all road users and surrounding backgrounds by providing 3-D points clouds so that plenty of frames in the same place can be obtained. Huge amounts of frames in the same place will help highlight the property that backgrounds are more stable across frames, compared with road users. The property is able to prompt specific background filtering methods for roadside LiDAR data. Ibisch *et al.* [29] developed a grid-based approach of background filtering for roadside LiDAR data to detect vehicles in a parking garage, where the whole detection space was divided into cells. Based on the LiDAR rays, the cells occupied by background points can be identified and excluded. Tarko *et al.* [30] utilized the height information to filter the points of ground surface. Lee and Coifman [6] manually selected a frame without road users as a reference frame and points in the current frame had the same positions to those in the reference frame were regarded as background points. Zhang *et al.* [8] developed a data association method for filtering backgrounds of roadside LiDAR data. But the method still needed to select manually an initial frame. Lv *et al.* [4] proposed a raster method for filtering backgrounds within pre-defined region of interests. Above methods require manual selection of region of interest or empty frames for references. Meanwhile, above-mentioned noises are paid little attention to.

The sample size for manual methods is relatively small due to large human costs. Small sample size is limited in dealing with dynamic background points and above-mentioned noises. Thus, the automatic methods are more attractive. Automatic methods are able to keep road users' points and delete background points in the detection range of LiDAR without requirements of empty frames and region of interests. Wu *et al.* [31] proposed an automatic density-based background filtering algorithm (3D-DSF) which doesn't need to manually select suitable frames. In the method, the whole detection range was divided into amounts of small cubes with equal lengths and the point density of each cube in each frame was calculated. Then by frames aggregation, the sum of point density over all frames of each cube can be gotten. The assumption of their study is that the sum of point density of background cube will be much smaller than that of the cube with road users. Later, they provided a method of automatically learning thresholds from history data to separate background cubes from other cubes [8]. The 3D-DSF method works well under free-flow conditions. But under congestions, the accuracy will decline dramatically by misrecognizing road users' points as background points, possibly because the sum of point density of some cubes with road users may be further away from that of other cubes with road users and closer to that of background cubes.

Via automatic methods, backgrounds and road users can be separated automatically without any inputs of human resources. And large-sized samples are guaranteed for training automatic models, as any frame can be utilized. Especially for portable deployment, automatic methods will be more suitable, since the short-term data collection might not have empty frames. However, the background identification model in existing automatic methods still require frames with low traffic volumes to better learn backgrounds information. If frames used for training existing

TABLE I
THE SYMBOLS AND DEFINITIONS

Symbol	Definition	Symbol	Definition
z_i	Features of element i	μ_j	Centroid of group j
ω_j	Weight	M	Sample size
$f(\cdot)$	Probability density function	J	Number of groups
$F(\cdot)$	Posterior distribution	$Dir(\cdot)$	Dirichlet distribution

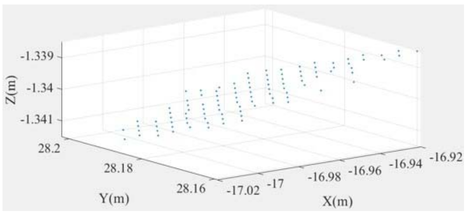


Fig. 1. An illustration of the vibrations of a LiDAR sensor.

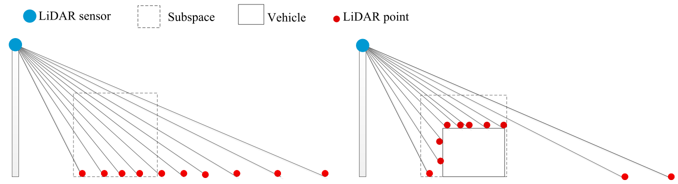


Fig. 2. An illustration of a subspace without (left) or with (right) a vehicle.

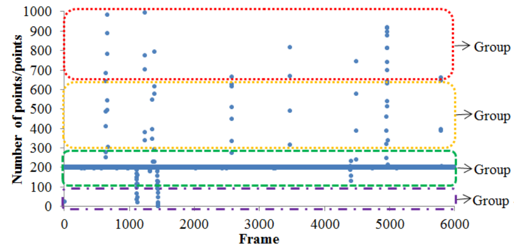


Fig. 3. An example of frames classification of one subspace.

background identification models have heavy traffics or road users stopped at intersections at red lights, background filtering performance will dramatically decline.

Therefore, this paper makes full use of huge amounts of frames' data to develop automatic background identification and filtering methods for roadside LiDAR data with noises under any traffic condition.

III. Background Filtering Methodology

A. Notation

Here are some symbols and definitions for readers.

B. Roadside LiDAR Data

Compared with on-board LiDAR, roadside LiDAR sensor is installed roadside to monitor the traffics in the same place all the time. In this study, a Velodyne 32-channel 360-degree rotating LiDAR sensor is used. It launches 32 laser beams to scan the road users and surrounding environments by rotating to output 3-D point clouds with a 360-degree horizontal and ± 15-degree vertical field of view. The rotating frequency is 10Hz. The effective detection radius is about 200m (656ft). Each LiDAR point includes XYZ coordinates, distance to the sensor, timestamp, laser ID, intensity, azimuth, vertical angle, etc.

Though the LiDAR sensor can obtain data with high resolution and accuracy, points will fluctuate over frames due to the vibration of LiDAR sensor, even for static background points. As shown in Fig.1, the static background point at the same location fluctuates over randomly selected frames. Thus, manual methods which need to manually select one or several frames without road users are limited in handling the noises.

Moreover, it is hard work to manually select enough number of frames where there are no road users in the whole space, especially for some intersections with heavy traffics all the day and the portable investigations. This paper expects to find automatic background filtering method for all types of traffic conditions (i.e., free-flow or congestion).

From the roadside LiDAR data, we can observe that compared to the locations of backgrounds, those of road users may change more greatly frame by frame due to the movements. Thus, it will be a good option to identify backgrounds if the frames can be classified into different groups where the frames in one group only contain backgrounds. After background identification, background filtering will be easily carried out.

In this paper, the whole space (i.e. detection range) is divided into subspaces with equal lengths (e.g. 5m). From a perspective of space-time diagram, a subspace in a frame is represented by a subspace-frame. A subspace over frames corresponds to amounts of subspace-frames, whereas subspaces in a frame can be represented by subspaces-frame. Via subspace background identification, for each subspace, background subspace-frames without road users can be identified by clustering methods. As shown in Fig.2, features (e.g., number of points, centroid, etc.) of the subspace will differ a lot by comparing two frames where one frame has a vehicle and another frame only has background points. Namely, the two subspace-frames with/without road users differ a lot in terms of number of points and centroid of points, etc. For example, the number of points changes from 4 to 8 points after a vehicle moves into this subspace.

In practice, one subspace may be occupied by only parts of one road user and the occupied rate may vary over frames, leading to significant changes of features over frames. And sometimes, even for subspace-frames without road users, related features may change a lot, since background points in one subspace may be blocked by road users in other subspaces. Thus more than two groups of subspace-frames will be classified according to features. For instance, if only using number of points as the feature for classification as shown in Fig.3, four groups may be better: Group 1 is subspace-frames with more road users' points; Group 2 is subspace-frames with fewer road users' points; Group 3 is subspace-frames without road users; Group 4 is subspace-frames where background points are blocked. Actually, the number of groups is a random value. The unknown number of groups for each subspace will be estimated together with clustering in this paper.

After obtaining the background subspace-frames of each subspace, in order to save computational time and take the noises into account, the subspace can be cut into plenty of small cubes with equal lengths (e.g., 0.1m). The cube with points in the background subspace-frames will be identified as a background cube. Then the background cubes can be used for real-time background filtering. Namely, the points located within background cubes will be deleted. And the left points can be further used for road users detection. More detailed information can be found in the following subsections.

C. Background Identification

1) Subspace Partition: In order to better identify the subspace without road users in a frame, side length of subspace should be determined reasonably, otherwise, too long side will weaken some specific features and reduce the possibility of finding it, whereas too short side can't capture stable and significant features of the subspace in each frame as well as increasing the computational time with the explosive increase of number of subspaces. Considering that the side length of a vehicle is usually 3-7m, the side length of subspace is recommended as 3-7m by this paper. In case study, we will compare results based on different side lengths to see the differences.

2) Clustering With an Unknown Number of Clusters: Usually, there may be multiple objects in one frame, leading to multiple centroids. Given an object with features, its distance to each centroid is assumed to follow a mixture t-distribution [32]. So a t-distribution mixture model (TMM) with constraints is proposed. The mixed probability density function is

$$f(z_i;\boldsymbol{\mu},\boldsymbol{\omega}) = \sum_{j=1}^{J^{\max}} \omega_j \frac{\Gamma(\frac{\alpha+1}{2})}{\sqrt{\alpha\pi}\,\Gamma(\frac{\alpha}{2})}\left(1+\left\|z_i-\mu_j\right\|^2\big/\alpha\right)^{\frac{-(\alpha+1)}{2}}, \quad j=1,\ldots,J,\ldots J^{\max} \tag{1}$$

$$\begin{cases} \sum_{j=1}^{J}\omega_j = 1 \quad \sum_{j=J+1}^{J^{\max}}\omega_j = 0 \\ \omega_j \geq 0 \quad J \leq J^{\max} \end{cases} \tag{2}$$

where α is usually set as 1; J is unknown; $J^{\max}$ is a maximum number of groups. The reason to involve $J^{\max}$ is related to a method of automatically determining the number of groups.

The weight is related to the proportion of the population belonging to one group. But it cannot tell the probability that one object i belongs to group j. Usually, the probability will be different for each object and each group. It is essential in this paper to know the probability of each object belonging to each group to identify whether one subspace in one frame is background or not. We propose the following probability equation to measure the probability [32], which can also be regarded as the similarity between one object and centroid.

$$q_{ij} = \frac{\left(1+\left\|z_i-\mu_j\right\|^2\big/\alpha\right)^{-\frac{\alpha+1}{2}}}{\sum_{j'=1}^{J^{\max}}\left(1+\left\|z_i-\mu_{j'}\right\|^2\big/\alpha\right)^{-\frac{\alpha+1}{2}}}, \quad j=1,\ldots,J,\ldots J^{\max} \tag{3}$$

Above formulation is a soft assignment which means one object has different probabilities to be assigned to different groups. So there is a relationship between object-based probability q_{ij} and population-based weight ω_j, as follows.

$$\frac{\sum_{i=1}^{M} q_{ij}}{M} \overset{M\to\infty}{=} \omega_j \quad, \quad j=1,\cdots,J,\cdots J^{\max} \tag{4}$$

How to determine the number of groups is a great challenge for clustering. Usually, the number is given in advance, such as K-means, supported vector machine, neural network, and random forest, etc. This paper proposes an automatic method of determining the number of groups.

Firstly, we introduce $J^{\max}$ into above formulations. It is not easy to give the actual number of groups. But it's easy to provide a maximum value, e.g., 100 for simplicity. Given the maximum value, a zero weight means the corresponding group doesn't exist. We call it empty-group in this paper. So the number of non-zero weights will be the number of groups.

Secondly, weights $\boldsymbol{\omega}$ are assumed as random variables. The prior distribution is assumed as a Dirichlet distribution, that is $\boldsymbol{\omega} \sim Dir(\boldsymbol{\rho})$. The sum of weights sampled from this distribution each time keeps 1. And Dirichlet distribution benefits that it has a large possibility to sample extreme values, such as 0. So it will benefit the searching for a suitable number of groups.

Finally, for one subspace, several groups may be generated. The group with the most frames will be background group since there will be longer time of no road users in the subspace if the number of frames is large enough (e.g., 3000-12000frames).

D. Background Filtering

After generating the background group of each subspace, we can know the frames without road users in the subspace. Such frames may be just parts of all frames for some subspaces. So deeper processing should be done for more efficient real-time background filtering.

For generated background frames of one subspace, all points in background frames belong to background points. If we cut the subspace into smaller cubes, cubes with background points can be regarded as background cubes all the time. It means when doing background filtering for new data, points located in background cubes will be excluded immediately.

Suppose the XYZ range of a background cube b is $x_1^b \sim x_2^b$, $y_1^b \sim y_2^b$, $z_1^b \sim z_2^b$. Then a new background point (x, y, z) should satisfy: $x_1^b \leq x < x_2^b$, $y_1^b \leq y < y_2^b$, $z_1^b \leq z < z_2^b$.

IV. Algorithm

The core part of background identification and filtering method is the background learning. After learning background information, it is very easy to do background filtering following above descriptions. Thus, this section will only introduce the clustering algorithm for background learning. The goal of the estimation is equivalent to optimize the likelihood function $L'(\boldsymbol{\mu},\boldsymbol{\omega};\mathbf{z})$ constrained by (4). The objective function is

$$\max \quad L'(\boldsymbol{\mu},\boldsymbol{\omega};\mathbf{z}) = \prod_i f(z_i;\boldsymbol{\mu},\boldsymbol{\omega}) \tag{5}$$

Fig. 4. The location of data collection (left) and the LiDAR sensor (right).

Suppose the sample size used for estimation is large enough. Via the Lagrangian multiplier method, the objective function and constraints can be integrated by adding Lagrangian multipliers $\boldsymbol{\lambda}$. For simplicity, if $\sum_{i=1}^{M} q_{ij} - M\omega_j \geq 0$, $\lambda_j = -1$; otherwise, $\lambda_j = 1$. By this means, only weights and centroids are unknown parameters required to be estimated.

$$\max\ L(\boldsymbol{\mu}, \boldsymbol{\omega}; \mathbf{z}) = \prod_i f(z_i; \boldsymbol{\mu}, \boldsymbol{\omega}) + \sum_{j=1}^{J^{\max}} \lambda_j \Big(\sum_{i=1}^{M} q_{ij} - M\omega_j\Big) \tag{6}$$

Since all parameters are random variables and their prior distributions can be made full use of, a Bayesian framework can be derived, given the observations $\mathbf{z}$. As shown below, the mean value of each parameter from accepted samples of the Bayesian framework will be equivalent to the optimization of (6).

$$F(\boldsymbol{\mu}, \boldsymbol{\omega}; \mathbf{z}) \propto \left[\prod_i f(z_i; \boldsymbol{\mu}, \boldsymbol{\omega}) + C(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\omega})\right] f(\boldsymbol{\mu}) f(\boldsymbol{\omega}) \tag{7}$$

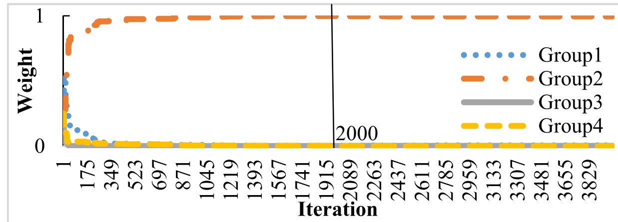
where $C(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\omega}) = \sum_{j=1}^{J^{\max}} \lambda_j (\sum_{i=1}^{M} q_{ij} - M\omega_j)$.

Background identification will be implemented subspace by subspace. To obtain posterior distributions of the centroid and weights of each subspace, Metropolis-Hastings (MH) sampling algorithm is employed to estimate the Bayesian framework. Based on the random walk process, candidate samples of centroids are generated from a conditional normal distribution $\boldsymbol{\mu}^{t+1}|\boldsymbol{\mu}^{t} \sim N(\boldsymbol{\mu}^{t}, 1)$ and weights are generated from a conditional Dirichlet distribution $\boldsymbol{\omega}^{t+1}|\boldsymbol{\omega}^{t} \sim Dir(\boldsymbol{\rho}^{t+1})$.

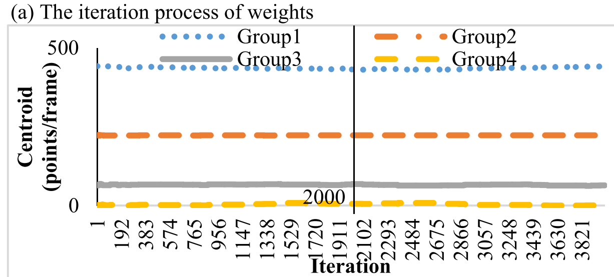
V. Numerical Results

A. Data

A Velodyne 32-channel LiDAR sensor (10Hz) is installed at the intersection of McCarran @ Evans, Reno, Nevada, United States to collect data, as shown in Fig.4. The effective detection range is 200m in radius. The data used in this paper was collected from 17:00 to 17:30 (during peak hour) on March 31st and from 6:00am to 6:30am (during the off-peak hour) on April 1st in the year of 2019. There were 18000 frames collected during peak and off-peak hours respectively, where 1-12000 frames were selected as training data for machine learning to identify backgrounds and 12001-18000 frames were used as test data for testing the background filtering performance. The average numbers of points in one frame of the two half-an-hours are 37902 points and 38504 points respectively. The ground truths of the following evaluated results are all obtained by manually observing LiDAR data from Veloview 3.5.0 (the software developed by Velodyne for watching LiDAR data).



(a) The iteration process of weights



(b) The iteration process of centroids

Fig. 5. The iteration process of weights (a) and centroids (b).

B. Results and Discussions

1) Model Estimations: Background learning is the core part of the whole process of background filtering. Based on the collected data, the proposed background learning approach is run on a desktop equipped with a processor of Intel (R) Core (TM) i7-7700 CPU @ 3.60GHz and an installed memory (RAM) of 32.0GB. The number of iterations is set as 4000, where the results are derived from the last 2000 iterations, and the fore 2000 iterations are for the burn-in test. In this example, the side length of the subspace is 5m and the side length of the small cube is 0.1m. The feature used for background identification and filtering is the number of points in one frame of the subspace. We set the maximum number of groups as 4, including the group with more road users' points, the group with fewer road users' points, the group with only backgrounds and no blocks, and the group with only backgrounds but parts of them are blocked. The initial values of weights are 0.25, 0.25, 0.25 and 0.25 respectively. The initial values of centroids are 0, the minimum number of points, the average number of points and the maximum number of points.

Since the number of subspace is very large, we just randomly select one subspace to show the iteration process associated with the weights and centroids in Fig.5. We can see that the weights will differ significantly after amounts of iterations, even though they have the same initial values. And each parameter will fluctuate slightly around a value after amounts of iterations. The sum of weights is always equal to be 1.

All estimated values and corresponding Bayesian confidence intervals (BCI) are displayed in Table II, where we can see that all estimations can be accepted with more than 95%

TABLE II
THE ESTIMATIONS AND ACTUAL VALUES OF WEIGHTS AND CENTROIDS

G.	Weight		Centroid	
	Estimations (95%BCI)	Actual values	Estimations (95%BCI)	Actual values
1	0.0042 (0.004192, 0.004233)	0.0048	435.1868 (435.0452, 435.3285)	420.0000
2	0.9949 (0.994898, 0.994955)	0.9940	222.6643 (222.6637, 222.6648)	222.8140
3	0.00086 (0.000851, 0.000868)	0.0012	65.0855 (65.0248, 65.1463)	69.5000
4	$1.78*10^{-6}$ $(1.55, 2.02)*10^{-6}$	0	3.9597 (3.8305, 4.0888)	0

confidence level. It can be seen that the estimated values are very close to the actual values. The average absolute errors of the weights and centroids are 0.00045 and 5.9277 points respectively.

Moreover, as shown in Table II, the weight of group 4(G.4) tends to be 0, indicating that there are only 3 effective groups for the subspace. And the weight of group 2(G.2) is the largest, indicating that frames in group 2 are background frames for the subspace. Possibly group 3(G.3) contains only backgrounds but some of them are blocked and group 1(G.1) contains fewer road users' points. Then using the background frames, small background cubes for background filtering can be generated.

C. Model Evaluations

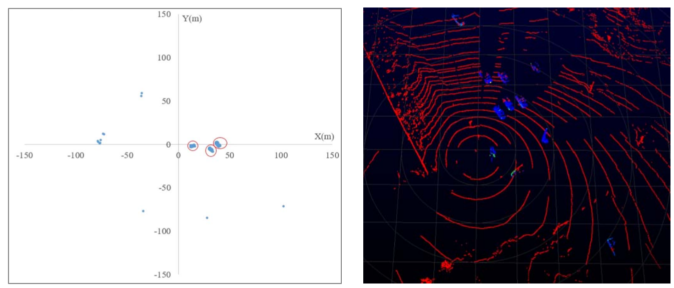
In addition to the above statistical test of estimations, this paper exhibits more direct evaluations by calculating two types of errors, compared to the 3D-DSF, data association (DA), and raster methods. The 3D-DSF method is an automatic method. The DA and raster methods are manual methods. Error Type 1 counts the number of points misrecognized as road users' points, whereas Error Type 2 counts the number of points misrecognized as background points.

The only way to calculate the two types of errors is to count the mismatched points in each frame manually. As the number of frames is a little large, every 500 frames are selected for evaluation. Training data and test data are evaluated separately, since the evaluation of training data represents the goodness-of-fit of the model to the data, and the evaluation on test data represents the performance of prediction. The average errors under two typical traffic states (i.e., free-flow and congestion) are displayed in Table III.

In Table III, the proposed method has smaller errors for both Error Types under free-flow condition or congestions, compared to 3D-DSF, DA and raster methods. Especially under congestions, the proposed method is much better than those methods. The accuracies of the proposed method for training and test data are all higher than those of the three methods. Above results show the superiority of the proposed method over the three methods. The perfect performance of the proposed method under free-flow and congested conditions can be clearly seen in Fig.6 as an example. Under both conditions, all road users have most/all LiDAR points left and background points are mostly deleted (red points are background points).

TABLE III
EVALUATIONS ON THE PROPOSED METHOD

	Condition	Method	Error Type1 (points/frame)	Error Type2 (points/frame)
Train	Free-flow (06:00-06:30)	3D-DSF	21.52	31.96
		DA	20.11	29.35
		Raster	20.98	29.06
		Proposed	9.04	12.92
	Congestion (17:00-17:30)	3D-DSF	48.14	926.50
		DA	38.56	508.55
		Raster	29.36	586.63
		Proposed	11.00	258.57
Test	Free-flow (06:30-07:00)	3D-DSF	33.00	28.72
		DA	25.99	28.05
		Raster	23.54	27.32
		Proposed	15.66	20.28
	Congestion (17:30-17:00)	3D-DSF	63.42	273.75
		DA	42.22	198.65
		Raster	35.03	209.11
		Proposed	25.08	172.83



(a) Under free-flow condition (b) Under congestions

Fig. 6. Background filtering performance based on the proposed method.

D. The Impact of the Side Length of Subspace

This part is to investigate impacts of side length of subspace on background filtering performance by comparing two types of errors, as shown in Table IV. The data for investigation is the frames 1-12000 in 17:00-17:30, March 31st, 2019.

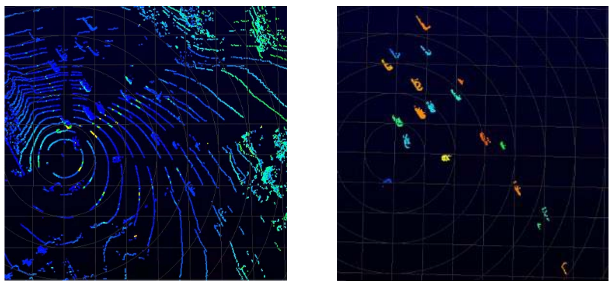
From the table, we can see that too long or too short side length will cause larger errors especially for Error Type 2, namely, more road users will be misrecognized as background points. Though 1m side length has the smallest Error Type 1, the Error Type 2 is much larger. By comprehensive comparison, 5m side length is more attractive, since the 5m side length has the smallest Error Type 2 and similar Error Type 1 with 3m, 7m, and 9m side lengths. We'd like to recommend 3-7m in practice.

E. Applications in Road Users' Detection

Though most of background points can be excluded by the proposed background filtering method, there are still some background points left within the detection range. What's more, it is still a problem of identifying one point belonging to which road user. By object clustering, the left background points can be excluded and road user points can be classified into different groups. The points in one group form one road user. Then by object identification, each road user will be identified as a vehicle, a bicycle or a pedestrian.

TABLE IV
THE IMPACT OF THE LENGTH OF SUBSPACE

Subspace side length (m)	Error Type 1 (points/frame)	Error Type 2 (points/frame)
1	3.72	427.28
3	8.29	288.14
5	11.00	258.57
7	9.57	280.79
9	10.57	288.00



(a) raw data (b) data after object detection

Fig. 7. Road users detection.

For object clustering, the density-based spatial clustering of applications with noise (DBSCAN) method is very attractive for point clustering. There are two major parameters, including searching radius and minimum number of points for the traditional DBSCAN method. The searching radius determines the searching range within which another point will be included as the neighbor points of one point. The minimum number of points is a threshold, where if the number of neighbor points of one point exceeds the threshold, the point and its neighbor points will be included into a cluster; otherwise, the neighbor points are not included. Since the point density of roadside LiDAR decreases with the distance to the sensor, the traditional DBSCAN method should be revised by using varying searching radius and minimum number of points [5].

For object identification, a classifier based on the selected features is usually trained. Since the LiDAR data based object is formed by point cloud, the features which are easily computed from the point cloud are usually selected for an automatic identification procedure. The commonly used features include (1) number of points in one cluster, (2) nearest Distance between the point in one cluster and the LiDAR sensor, (3) object length, (4) object width, (5) object height, (6) difference between object length and object height [33], [34]. Among various classification methods, such as Naïve Bayes, support vector machine, K-nearest neighbor, and random forest, the random forest method which has been evaluated to have the highest accuracy [35] is used in this paper.

After object clustering and identification, all road users within the detection range are automatically detected and identified frame by frame. We evaluated the detection accuracy at the intersection of McCarran and Evans, Reno, from 17:00 to 17:30 (during peak hour) on March 31st, 2019. The processing and detection results in one frame are shown in Fig.7 for an example. In this frame, all road users are vehicles, the same to the detection results. After background filtering, object clustering and identification, each color represents one road user as shown in Fig.7 (b). In the 18000 frames, manually counted road users are regarded as the ground truth. By comparing to the detected results based on the proposed background filtering method, the average accuracy is 98.18%. By checking the data, the errors are found to be mainly caused by occlusions. For example, one vehicle cannot be detected because most of the vehicle is blocked. The few left points may be filtered by background filtering or object clustering.

VI. CONCLUSION

Roadside LiDAR data is a mixture of road users and backgrounds. In order to better identify and track all road users, background filtering is firstly carried out. Existing methods were limited heavily by noises and if the LiDAR sensor is installed at a busy intersection or used for portable investigations. This paper proposes a clustering-based framework of background learning and filtering for roadside LiDAR data even with heavy traffics. The proposed method can be utilized in any traffic scene and any time period of a day to get more accurate background filtering results.

In theory, innovative framework, model and algorithm are proposed. In particular, a clustering-based framework of background identification and filtering for roadside LiDAR data is proposed. A t-distribution mixture model with constraints of the relationship between object-based probability and population-based weight, Dirichlet distribution based weights and an unknown number of groups are proposed for clustering. The model is estimated based on a Bayesian method.

In application, the background filtering accuracy is highly improved no matter under free-flow conditions or congestions. And the proposed method has a high transferability which allows it to be used in any time and location. By combining background filtering, object clustering and identification, road users are detected for real-time application.

Some limitations of this paper still exist. For example, the proposed method is not validated in special weather conditions, such as rain, fog, snow, etc. Another example is that the used detection method of road users is applicable for identifying vehicles and pedestrians, but may not be suitable for more detailed classifications, such as skateboarder, wildlife, etc.

In the future, it is expected to have the following, but not limited to, research directions. One is to evaluate and improve background identification and filtering in different weather conditions. Another one is to use more advanced methods, like deep-learning method, to identify more road users based on the LiDAR-points-based objects generated by this paper.

ACKNOWLEDGMENT

On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

[1] *National Connected Vehicle Field Infrastructure Footprint Analysis*, U.S. Dept. Transp., Washington, DC, USA, 2013.

[2] M. Wu and B. Coifman, "Quantifying what goes unseen in instrumented and autonomous vehicle perception sensor data—A case study," *Transp. Res. C, Emerg. Technol.*, vol. 107, pp. 105–119, Oct. 2019.

[3] X. Zhao, P. Sun, Z. Xu, H. Min, and H. Yu, "Fusion of 3D LIDAR and camera data for object detection in autonomous vehicle applications," *IEEE Sensors J.*, to be published.

[4] B. Lv, H. Xu, J. Wu, Y. Tian, and C. Yuan, "Raster-based background filtering for roadside LiDAR data," *IEEE Access*, vol. 7, pp. 76779–76788, 2019.

[5] J. Zhao, H. Xu, H. Liu, J. Wu, Y. Zheng, and D. Wu, "Detection and tracking of pedestrians and vehicles using roadside LiDAR sensors," *Transp. Res. C, Emerg. Technol.*, vol. 100, pp. 68–87, Mar. 2019.

[6] H. Lee and B. Coifman, "Side-fire lidar-based vehicle classification," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2308, no. 1, pp. 173–183, Jan. 2012.

[7] A. P. Tarko, K. B. Ariyur, M. A. Romero, V. K. Bandaru, and C. G. Lizarazo, "TScan: Stationary LiDAR for traffic and safety studies-Object detection and tracking," Purdue University, West Lafayette, IN, USA, Tech. Rep. FHWA/IN/JTRP-2016/24, 2016, pp. 5–6.

[8] Z. Zhang, J. Zheng, X. Wang, and X. Fan, "Background filtering and vehicle detection with roadside lidar based on point association," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 7938–7943.

[9] J. Wu, H. Xu, Y. Sun, J. Zheng, and R. Yue, "Automatic background filtering method for roadside LiDAR data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2672, no. 45, pp. 106–114, Jun. 2018.

[10] Y. Cui, H. Xu, J. Wu, Y. Sun, and J. Zhao, "Automatic vehicle tracking with roadside LiDAR data for the connected-vehicles system," *IEEE Intell. Syst.*, vol. 34, no. 3, pp. 44–51, May 2019.

[11] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[12] Y. Wan, Y. Huang, and B. Buckles, "Camera calibration and vehicle tracking: Highway traffic video analytics," *Transp. Res. C, Emerg. Technol.*, vol. 44, pp. 202–213, Jul. 2014.

[13] N. Csanyi and C. K. Toth, "Improvement of lidar data accuracy using lidar-specific ground targets," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 4, pp. 385–396, Apr. 2007.

[14] A. Mukhtar, L. Xia, and T. B. Tang, "Vehicle detection techniques for collision avoidance systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2318–2338, Oct. 2015.

[15] B. Yang, Z. Dong, G. Zhao, and W. Dai, "Hierarchical extraction of urban objects from mobile laser scanning data," *ISPRS J. Photogram. Remote Sens.*, vol. 99, pp. 45–57, Jan. 2015.

[16] G. Vosselman, "Slope based filtering of laser altimetry data," in *Proc. ISPRS*, Jul. 2000, pp. 935–942.

[17] M. Elmqvist, "Ground surface estimation of laser scanner radar data using active shape models," in *Proc. Commission. III Symp., Photogramm. Comp. Vis. (ISPRS)*, Graz, Austria, 2002, pp. 114–118.

[18] B. Yang, R. Huang, Z. Dong, Y. Zang, and J. Li, "Two-step adaptive extraction method for ground points and breaklines from lidar point clouds," *ISPRS J. Photogram. Remote Sens.*, vol. 119, pp. 373–389, Sep. 2016.

[19] M. Brackstone, M. McDonald, and B. Sultan, "Dynamic behavioral data collection using an instrumented vehicle," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1689, no. 1, pp. 9–16, Jan. 1999.

[20] D. Zeng Wang, I. Posner, and P. Newman, "What could move? Finding cars, pedestrians and bicyclists in 3D laser data," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 4038–4044.

[21] B. Nagy and C. Benedek, "3D CNN-based semantic labeling approach for mobile laser scanning data," *IEEE Sensors J.*, vol. 19, no. 21, pp. 10034–10045, Nov. 2019.

[22] A. Asvadi, C. Premebida, P. Peixoto, and U. Nunes, "3D lidar-based static and moving obstacle detection in driving environments: An approach based on voxels and multi-region ground planes," *Robot. Auto. Syst.*, vol. 83, pp. 299–311, Sep. 2016.

[23] J. Zheng, Y. Wang, N. L. Nihan, and M. E. Hallenbeck, "Extracting roadway background image: Mode-based approach," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1944, no. 1, pp. 82–88, Jan. 2006.

[24] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[25] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893.

[26] Y. Sun, M. Liu, and M. Q.-H. Meng, "Active perception for foreground segmentation: An RGB-D data-based background modeling method," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1596–1609, Oct. 2019.

[27] Y. Sun, M. Liu, and M. Q.-H. Meng, "Motion removal for reliable RGB-D SLAM in dynamic environments," *Robot. Auto. Syst.*, vol. 108, pp. 115–128, Oct. 2018.

[28] Y. Sun, M. Liu, and M. Q.-H. Meng, "Improving RGB-D SLAM in dynamic environments: A motion removal approach," *Robot. Auto. Syst.*, vol. 89, pp. 110–122, Mar. 2017.

[29] A. Ibisch *et al.*, "Towards autonomous driving in a parking garage: Vehicle localization and tracking using environment-embedded LIDAR sensors," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2013, pp. 829–834.

[30] A. Tarko, M. Romero, K. Ariyur, V. Bandaru, and C. Lizarazo, "Detecting and tracking vehicles, pedestrians, and bicyclists at intersections with a stationary LiDAR," in *Proc. 18th Int. Conf. Road Safety Five Continents (RSC)*, Jeju Island, South Korea, May 2018, pp. 16–18.

[31] J. Wu, H. Xu, and J. Zheng, "Automatic background filtering and lane identification with roadside LiDAR data," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.

[32] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[33] J. Shackleton, B. Vanvoorst, and J. Hesch, "Tracking people with a 360-degree lidar," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug. 2010, pp. 420–426.

[34] H. Zhao *et al.*, "Detection and tracking of moving objects at intersections using a network of laser scanners," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 655–670, Jun. 2012.

[35] J. Wu, H. Xu, Y. Zheng, Y. Zhang, B. Lv, and Z. Tian, "Automatic vehicle classification using roadside LiDAR data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 6, pp. 153–164, May 2019.

Yongsheng Zhang received the bachelor's degree from the Department of Traffic and Transportation, North China Polytechnic University of China, in 2011, and the Ph.D. degree in traffic and transportation planning and management from the Beijing Jiaotong University of China in 2017. From 2017 to 2018, he was a Postdoctoral Research Associate with The Hong Kong Polytechnic University, Hong Kong. Since 2018, he has been a Postdoctoral Scholar with the University of Nevada Reno, USA. His research focuses are in the areas of connected-vehicle applications with roadside LiDAR sensors, big data mining, travel behavior analysis, and travel demand prediction.

Hao Xu received the bachelor's and master's degrees from the Department of Automation, University of Science and Technology of China in 2004 and 2007, respectively, and the master's and Ph.D. degrees in civil engineering from Texas Tech University in 2009 and 2010, respectively. From 2010 to 2013, he was a Postdoctoral Research Associate with Texas Tech University. Since 2018, he has been an Associate Professor with the Department of Civil and Environmental Engineering, University of Nevada Reno. His research focuses in the areas of connected-vehicle applications with roadside LiDAR sensors, driving behavior analysis, and traffic safety. His research team has set up several LiDAR-based connected intersections in Nevada and is performing research on LiDAR-based infrastructures.

Jianqing Wu received the B.S. and M.S. degrees in civil engineering from Shandong University, Jinan, China, in 2012 and 2015, respectively, and the Ph.D. degree from the Department of Civil and Environmental Engineering, University of Nevada Reno, USA, in 2018. He was a Postdoctoral Research Associate with the University of Nevada Reno in 2019. Dr. Wu is a Professor with Shandong University. His research interests include driver behavior analysis, intelligent transportation systems, traffic safety, and big data.