A Random Finite Set based Detection and Tracking using 3D LIDAR in Dynamic Environments

B.Kalyan, K.W.Lee, S.Wijesoma, D.Moratuwage
Dept. of Electrical and Electronic Engg.
Nanyang Technological University
50 Nanyang Avenue, Singapore 639798
bharath,leekw,eswwijesoma,mora0001{@ntu.edu.sg}

N.M. Patrikalakis
Dept. of Mechanical Engg., Center for Ocean Engg.
Massachusetts Institute of Technology
Cambridge, MA 02139
nmp@mit.edu

***Abstract*—In this paper we describe a fully integrated system for detecting and tracking pedestrians in a dynamic urban environment. The system can reliably detect and track pedestrians to a range of 100 m in highly cluttered environments. The system uses a highly accurate 3D LIDAR from Velodyne to segment the scene into regions of interest or blobs, from which the pedestrians are determined. The pedestrians are then tracked using probability hypothesis density (PHD) filter which is based on random finite set theoretic framework. In contrast to classical approaches, this random finite set framework does not require any explicit data associations. The PHD filter is implemented using a Gaussian Mixture technique. Experimental results obtained in dynamic urban settings demonstrate the efficacy and tracking performance of the proposed approach.**

***Index Terms*—Tracking, Velodyne, RFS, PHD filter**

I. Introduction

The ability of unmanned autonomous vehicles to detect and avoid colliding with objects in their area of operation is critical, especially in cases where they operate in close proximity to humans. In order to be operational and effective in dynamic urban environments, these vehicles must be able to detect people and track their motion in uneven terrain with varying degrees of clutter, occlusion and illumination. Laser scanners have proven to be efficient and less noisy in comparison with other ranging sensors such as radar and ultrasonic sensors that provide direct distance measurements. Most laser sensors reported in literature are restricted to two dimensions, that scan along a plane within a limited viewing angle. Mounted parallel to the ground plane, each scan acquires a sequence of range and bearing measurements. This allows easy detection of objects in the environment by applying straight forward signal processing methods. However, objects above or below the scanning plane cannot be detected. The limited number of measurements thereby complicates classification and tracking of objects. Additionally, in uneven terrains, the scanner might fail to detect objects. Over the last few years, fully three-dimensional laser scanners have been introduced. These 3D laser scanners use an array of beams organized in multiple planes to provide range, bearing and azimuth data of objects. This allows detection of many kinds of objects, with pedestrians in particular, even when the surrounding terrain is uneven.

Thus the pedestrian tracking problem using 3D LIDAR can be decomposed into two main tasks, namely the detection of the objects of interest, in this case the pedestrians, in each frame and tracking these detections over time. The major challenge is to estimate the state of an unknown number of pedestrians, based on the measurements corrupted by noise, in the presence of clutter. The classical approach for solving this problem is to use a stochastic filter such as Kalman filter or it's variants to each object and use a data association technique such as the nearest neighbor to assign the appropriate measurement to each filter and track each object separately [2], [9]. An alternative and a more elegant approach is to consider these unknown number of pedestrians as a multi-object set represented by a single meta object and the measurements received by the sensor as a single set of measurements [7] and modeling them as random finite sets (RFS). This allows estimating multiple objects in presence of clutter and with data association uncertainty to be cast in a Bayesian filtering framework.

There has been extensive research on pedestrian detection and tracking using optical stereo vision [1] and 3D LIDAR [8], [11], with specific interests towards unmanned autonomous vehicles. In order to reduce false alarm rates, which can be significant in cluttered urban landscapes, most of the algorithms are usually catered to detecting and tracking moving upright pedestrians [1], [8]. However, these methods have reported to perform reasonably well in partial occlusion, non-upright postures and static pedestrians. In this paper we describe a fully integrated system based on RFS for detecting and tracking pedestrians using a 3D LIDAR in a dynamic urban environment. It involves the application of the probability hypothesis density (PHD) filter, which is a recursion that propagates the first order statistical moment of the RFS of states in time, to track multiple objects of interest in presence of measurement uncertainty and false alarms without any explicit data association [6]. Due to its ability to handle non-linear and variable number of targets, it has been applied in various fields ranging from tracking multiple moving targets in uneven terrain [10] to detecting and tracking of underwater objects [4], [3], location of targets observed from multiple bi-static radars [12] and tracking human figures in digital video [14]. Our main contribution in this paper is the adaptation of the PHD filter based on finite set statistics (FISST) to the complex real-world pedestrian tracking scenarios using 3D scanning LIDAR, with particular emphasis on Velodyne HDL-64E[1]. We demonstrate its performance with the data obtained in a natural urban landscape.

The remainder of the paper is structured as follows. Section II details the data acquisition from Velodyne HDL along with pre-processing steps involved. Section III reviews the basics of PHD filtering followed by the process and measurement models as used in the filter. It also discusses the GM-PHD object tracking algorithm using Velodyne HDL. Results, based on the experiments conducted are presented in Section IV. Section V concludes the paper.

II. Velodyne Data Processing

Our method of object tracking is based on a scan-wise acquisition and processing which is performed in several steps: a scan acquisition from a Velodyne HDL which is a 3D-LIDAR, followed by a segmentation to determine the region of interest (ROI), and finally a GM-PHD filter based pedestrian tracker. The Velodyne HDL provides 3D range scans by rotating an array of 64 beams around its vertical axis at $5-15$ Hz (10 Hz in our application) and producing close to around 1.8 million points per rotation. In the horizontal direction, the array provides 360^o field of view (FOV) with an angular resolution of approximately 0.09^o. Vertically, the pitch angles range from -24.8^o to $+2^o$ with an angular resolution of 0.4^o. Its range measurement accuracy typically is within 10 cm. The sensor is mounted on top of the platform providing range scans with a full FOV in horizontal direction.

A. Segmentation and Blob Extraction

The 3D point cloud data from each scan is projected onto a cylinder whose axis is the rotational axis of the Velodyne HDL. This projection yields a range image, whose pixel intensity values correspond to the distance measurements as shown in fig. 2a. The bearing and azimuth index (u, v) in the range image is a direct mapping of the bearing and azimuth values (θ, ϕ) from the Velodyne HDL. Fig. 1 shows a 3D point cloud of the experimental environment as perceived by Velodyne HDL and its corresponding optical image. The range image is segmented using mean shift segmentation technique to determine the ROI in the range image. The segmentation process mainly comprises of two steps: mean shift filtering of the original range image data, followed by blob extraction of the filtered data points. A result of the blob extraction process on the range image is as shown in fig. 2. The mean of the range values r, in each blob along with its corresponding bearing θ, in the range image forms a single measurement $\mathrm{z} = \{r, \theta\}$. A collection of these measurements form a measurement set Z, that is used to update the PHD filter, the details of which are discussed in the following section.

III. Tracking Algorithm Description

This section describes the method for tracking multiple unknown number of pedestrians from the Velodyne HDL in presence of clutter. To achieve this, we use the PHD filter that is based on finite set statistics [6]. Modeling set valued states and measurements as RFS allows the problem of estimating multiple unknown of objects to be formulated in a Bayesian filtering framework. However, the propagation of the full posterior distribution using the optimal Bayesian approach is not practical due to computational complexity. A recursive Bayesian approach for approximating the first order statistical moment of the full posterior distribution known as the Probability Hypothesis Density (PHD) was proposed by [6] as a tractable alternative to the optimal Bayes filter. However, the realization of the PHD filter involves multiple integrals that have no tractable closed form expressions in general. Sequential Monte-Carlo (SMC) [6] and Gaussian mixture (GM) [13] approximation techniques were devised to implement the PHD filter. In this paper, we apply the Gaussian mixture variant to implement the PHD filter for reliable tracking of the unknown and varying number of pedestrians as observed by Velodyne HDL.

A. The PHD filter

Let the state of single object at time k be represented by $\mathrm{x}_k = \{x_k, \dot{x}_k, y_k, \dot{y}_k\} \in \mathcal{F}(\mathrm{x})$, where (x_k, y_k) are the object position and $(\dot{x}_k, \dot{y}_k)$ the object speed and $\mathcal{F}(\mathrm{x})$ is the single object space. Let the single object measurement at time k, which is as a result of segmentation and blob extraction from a single Velodyne HDL scan be represented by $\mathrm{z}_k = \{r_k, \theta_k\} \in \mathcal{F}(\mathrm{z})$. The corresponding multi-object states and the multi-object measurements are represented as finite sets $\mathrm{X}_k = \{\mathrm{x}_{k,1}, \dots, \mathrm{x}_{k,N_k}\}$ and $\mathrm{Z}_k = \{\mathrm{z}_{k,1}, \dots, \mathrm{z}_{k,l_k}\}$ which contain states of individual objects and measurements respectively [2]. The PHD filter recursion is a two step process:

- **PHD time update:** Given the process model, the predicted PHD,

$$D_{k|k-1}(\mathrm{x}_k|\mathrm{Z}^{(k-1)}) = \overbrace{\gamma_k(\mathrm{x}_k)}^{\text{new objects}} + \int \overbrace{p_S(\mathrm{x}_{k-1}).f_{k|k-1}(\mathrm{x}_k|\mathrm{x}_{k-1})}^{\text{existing objects}} .D_{k-1|k-1}(\mathrm{x}_{k-1}|\mathrm{Z}^{(k-1)})d\mathrm{x}_{k-1} \tag{1}$$

  where,
  - $\gamma_k(\mathrm{x}_k)$: PHD of the new incoming objects within the Velodyne HDL field of view (FOV)
  - $p_S(\mathrm{x}_{k-1})$: Probability of an object being re-observed
- **PHD data update:** Given a new set of measurements Z_k, the updated PHD,

$$D_{k|k}(\mathrm{x}_k|\mathrm{Z}^{(k)}) = (1-p_D)D_{k|k-1}(\mathrm{x}_k|\mathrm{Z}^{(k-1)}) + \sum_{\mathrm{z}_k \in \mathrm{Z}_k} \frac{p_D D_k(\mathrm{z}_k)}{\lambda_c c_k(\mathrm{z}_k) + p_D D_k(\mathrm{z}_k)} D_k(\mathrm{x}_k|\mathrm{z}_k) \tag{2}$$

  where,

$$D_k(\mathrm{z}_k) = \int f_k(\mathrm{z}_k|\mathrm{x}_k) D_{k|k-1}(\mathrm{x}_k|\mathrm{Z}^{(k-1)})d\mathrm{x}_k \tag{3}$$

$$D_k(\mathrm{x}_k|\mathrm{z}_k) = \frac{f_k(\mathrm{z}_k|\mathrm{x}_k)D_{k|k-1}(\mathrm{x}_k|\mathrm{Z}^{(k-1)})}{D_k(\mathrm{z}_k)} \tag{4}$$

[1]referred to as Velodyne high definition LIDAR (HDL)

[2]$\mathrm{x}_{k,i}$ and $\mathrm{z}_{k,i}$ are denoted as x_k and z_k for notational simplicity.

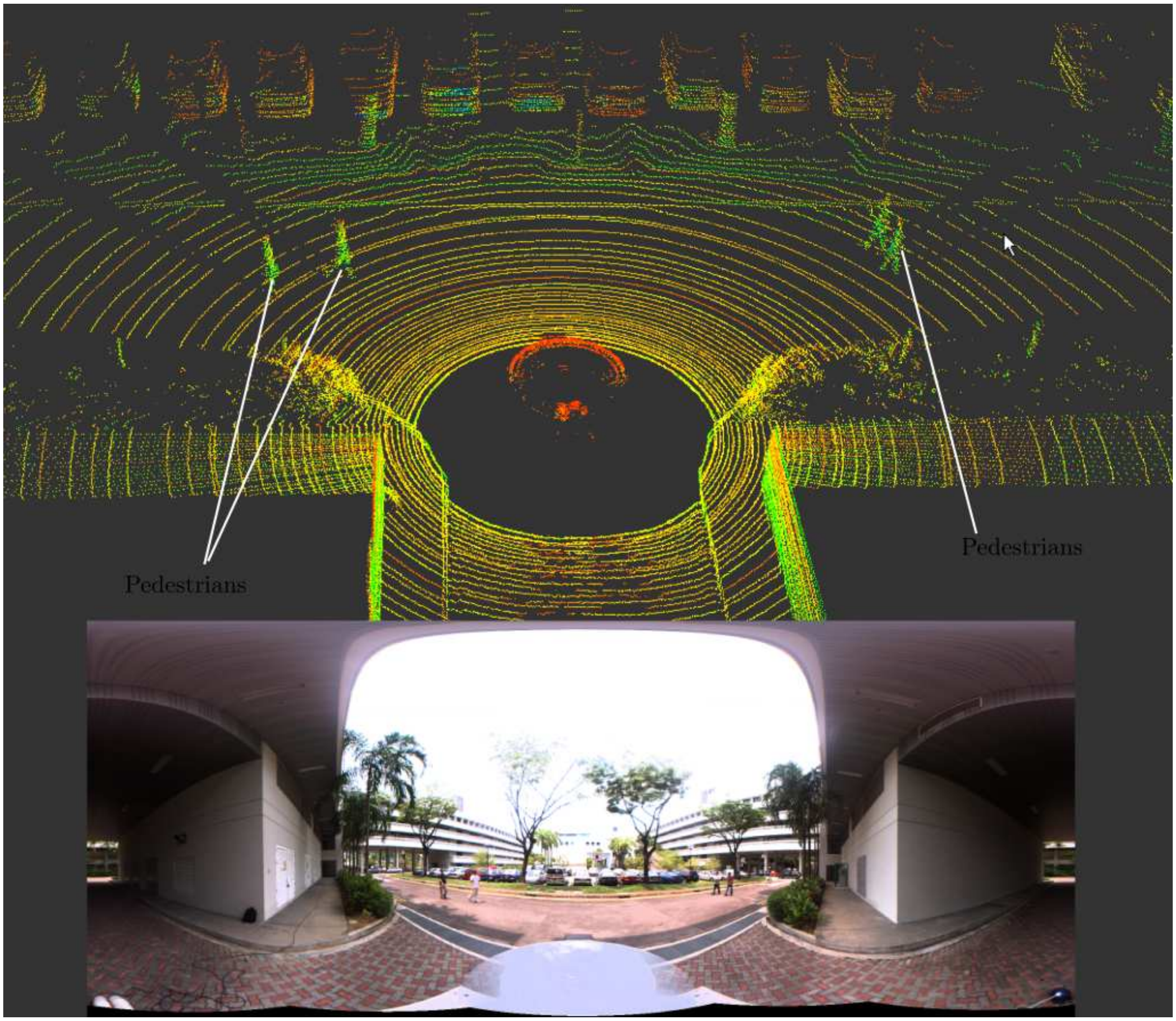
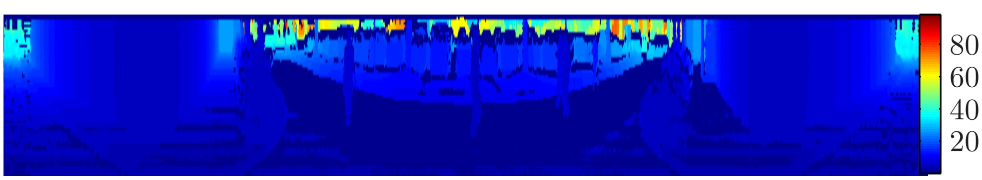
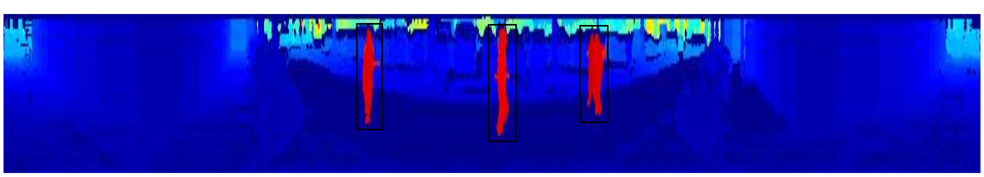


Fig. 1: 3D point cloud of the experimental environment as perceived by Velodyne HDL and its corresponding optical image (inset). The intensity of the signals are color-mapped with darker colors representing stronger intensity returns.



(a) Range image. Each pixel value correspond to a distance measurement as indicated by the colorbar.



(b) Segmented range image. The red blobs inside the rectangle indicate the output of the segmentation algorithm.

Fig. 2: Illustration of projection of point cloud from a scan in fig. 1 to obtain a range image. The result of mean-shift segmentation on the range image in (a) results in a segmented range image (b).

and,

- $f_k(\mathrm{z}_k|\mathrm{x}_k)$: is the sensor likelihood function $L_{\mathrm{z}}(\mathrm{x}_k)$
- λ_c: average number of false alarms per scan, which is assumed to be Poisson distributed
- $c_k(\mathrm{z}_k)$: distribution of each of the false alarms

B. Implementation of the GM-PHD filter tracker

In this work, we assume that each object moves according to the following Gaussian dynamics i.e.,

$$\mathrm{x}_k = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathrm{x}_{k-1} + \begin{bmatrix} \frac{T^2}{2} & 0 \\ T & 0 \\ 0 & \frac{T^2}{2} \\ 0 & T \end{bmatrix} \begin{bmatrix} \mathrm{v}_{1,k-1} \\ \mathrm{v}_{2,k-1} \end{bmatrix} \tag{5}$$

and the measurement model

$$\begin{aligned} r_k &= \left\| \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{matrix} \right\| \mathrm{x}_k + w_{1,k} \\ \theta_k &= arctan\left(\frac{\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \mathrm{x}_k}{\begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \mathrm{x}_k} \right) + w_{2,k} \end{aligned} \tag{6}$$

Thus, the state and the measurement process can be succinctly described as,

$$\mathrm{x}_k = F_k \mathrm{x}_{k-1} + G_k \mathrm{v}_{k-1} \tag{7}$$

$$\mathrm{z}_k = h_k(\mathrm{x}_k, \mathrm{w}_k) \tag{8}$$

where v_{k-1} and w_k are assumed to be zero mean Gaussian process noise and measurement noise with covariances Q_{k-1} and R_k, respectively. The implementation of the GM-PHD multi-object tracker is as proposed in [13]. For completeness, we summarize the key steps of the Velodyne HDL GM-PHD multi-object tracker in Table 1.

IV. Experiments and Results

The experiments were conducted in a natural outdoor setting at a car-park within the university. Fig. 1 shows a 3D point cloud of the experimental environment as perceived by Velodyne HDL and its corresponding optical image. The parameters used for the GM-PHD multi-object tracker are as follows. The number of Gaussian mixtures in the GM-PHD filter is limited to $J_k|k = 100$, with the Gaussian mixture pruning and merging parameters set to $\tau_p = 10^{-15}$ and $\tau_m = 4$ respectively. Two sequences demonstrating the performance of the multi-object tracker based on GM-PHD filter are shown fig. 3. The PHD filter does generate false estimates at times as observed at the end of the first sequence in fig. 3a. However, if the detections are not coherent and consecutive and the clutter is not persistent, then the PHD filter successfully manages to remove it.

V. Conclusion

In this paper we have presented a fully integrated system for detecting and tracking pedestrians in a dynamic urban environment using Velodyne HDL. We have presented a novel approach to segment the Velodyne HDL scans using a range image domain. A multi-object tracker based on GM-PHD

Algorithm 1 Velodyne HDL GM-PHD Multi-Object Tracker

• Initialize

At time $k = 0$, the PHD $D_{k|k}$ is initialized with a weighted sum of J_k Gaussians

$$D_{k|k}(\mathrm{x}|\mathrm{Z}_k) = \sum_{j=1}^{J_k} w_k^{[j]} \mathcal{N}(\mathrm{x}; \mu_k^{[j]}, \Sigma_k^{[j]})$$

These are distributed across the state space where each Gaussian term $\mathcal{N}(\mathrm{x}; \mu_k^{[j]}, \Sigma_k^{[j]})$ has a corresponding weight $w_k^{[j]}$, mean $\mu_k^{[j]}$, and variance $\Sigma_k^{[j]}$. At $k \geq 1$,

•Segmentation & Blob Extraction

The ROIs are extracted from the Velodyne HDL scan using the segmentation and blob extraction technique described in Section II. The mean of the range values r and their corresponding bearings θ of these blobs form the measurement set represented by Z_k at time k.

• PHD Time Update

The predicted PHD up to time k is a Gaussian mixture,

$$D_{k|k-1}(\mathrm{x}) = D_{S,k|k-1}(\mathrm{x}) + \gamma_k(\mathrm{x})$$

where, $D_{S,k|k-1}(\mathrm{x})$ is predicted intensity of the existing (survived) objects in the FOV of Velodyne HDL, given by,

$$D_{S,k|k-1}(\mathrm{x}) = p_S \sum_{j=1}^{J_{k-1}} w_{k-1}^{[j]} \mathcal{N}(\mathrm{x}; \mu_{S,k|k-1}^{[j]}, \Sigma_{S,k|k-1}^{[j]})$$

with,

$$w_{k|k-1}^{[j]} = p_S . w_{k-1}^{[j]}; \quad \mu_{S,k|k-1}^{[j]} = F_{k-1} \mu_{k-1}^{[j]};$$
$$\Sigma_{S,k|k-1}^{[j]} = Q_{k-1} + F_{k-1} \Sigma_{k-1}^{[j]} F_{k-1}^T$$

and, $\gamma_k(\mathrm{x})$ is the PHD representing the new incoming objects in the FOV of the sensor, given by,

$$\gamma_k(\mathrm{x}) = \sum_{j=1}^{J_{\gamma,k}} w_{\gamma,k}^{[j]} \mathcal{N}(\mathrm{x}; \mu_{\gamma,k}^{[j]}, \Sigma_{\gamma,k}^{[j]})$$

with,

$$w_{k|k-1}^{[j]} = w_{\gamma,k}^{[j]}; \quad \mu_{k|k-1}^{[j]} = \mu_{\gamma,k}^{[j]}; \quad \Sigma_{k|k-1}^{[j]} = \Sigma_{\gamma,k}^{[j]}$$

• PHD Data Update

The PHD measurement update is a Gaussian mixture given by,

$$D_{k|k}(\mathrm{x}) = (1 - p_D) D_{k|k-1}(\mathrm{x}) + \sum_{z \in \mathrm{Z}_k} D_{L,k}(z|\mathrm{x})$$

where,

$$D_{L,k}(z|\mathrm{x}) = \sum_{j=1}^{J_{k-1}+J_{\gamma,k}} w_{k|k}^{[j]} \mathcal{N}(\mathrm{x}; \mu_{k|k}^{[j]}, \Sigma_{k|k}^{[j]})$$

with,

$$w_{k|k}^{[j]} = \frac{p_D w_{k|k-1}^{[j]} f_k^{[j]}(z|\mathrm{x})}{\lambda_c c_k(z) + \sum_{l=1}^{J_{k-1}+J_{\gamma,k}} w_{k|k-1}^{[l]} f_k^{[l]}(z|\mathrm{x})}$$

$f_k^{[j]}(z|\mathrm{x}) = \mathcal{N}(z; h_k(\mu_{k|k-1}^{[j]}, 0), S_k^{[j]})$;

$\mu_{k|k}^{[j]} = \mu_{k|k-1}^{[j]} + K_k^{[j]}[z - h_k(\mu_{k|k-1}^{[j]}, 0)]; \Sigma_{k|k}^{[j]} = [I - K_k^{[j]} H_k^{[j]}] \Sigma_{k|k-1}^{[j]}$;

$K_k^{[j]} = \Sigma_{k|k-1}^{[j]} [H_k^{[j]}]^T [S_k^{[j]}]^{-1}; S_k^{[j]} = R_k + H_k^{[j]} \Sigma_{k|k-1}^{[j]} [H_k^{[j]}]^T$;

$H_k^{[j]} = \left. \frac{\partial h_k(\mathrm{x}_k, 0)}{\partial \mathrm{x}_k} \right|_{\mathrm{x}_k = \mu_{k|k-1}^{[j]}}$

Thus, there are $J_k = (1 + |\mathrm{Z}_k|)(J_{k-1} + J_{\gamma,k})$ Gaussian components in the updated PHD with $(1 + |\mathrm{Z}_k|)$ components for each prediction term at time k and the Gaussian mixture is of the form,

$$D_{k|k}(\mathrm{x}) = \sum_{j=1}^{J_k} w_{k|k}^{[j]} \mathcal{N}(\mathrm{x}; \mu_k^{[j]}, \Sigma_k^{[j]})$$

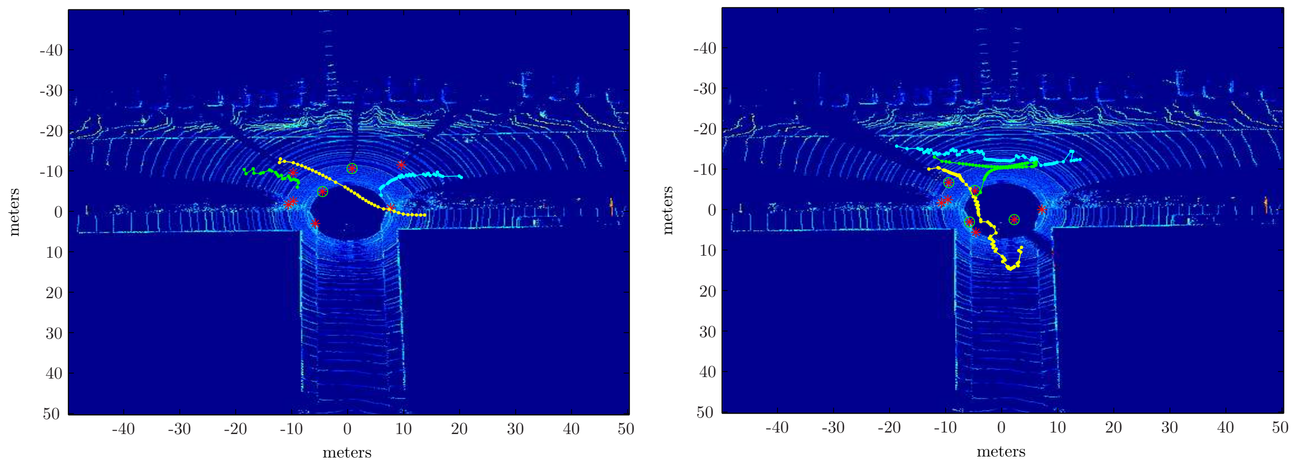
• Pruning & Merging

In the pruning stage, the Gaussians with weights below a pre-determined threshold τ_p, representing the updated PHD $D_{k|k}(\mathrm{x})$ are eliminated.

In the merging stage, the Gaussians whose distance between their means fall below a specific merging threshold τ_m, representing the updated PHD $D_{k|k}(\mathrm{x})$ are merged.

• Object State Estimation

The object states are obtained by selecting the Gaussians that are above a pre-determined threshold. In addition to these, the Gaussians that have already been classified as a valid object earlier are also included.



(a) (b)

Fig. 3: Tracks of pedestrians from the GM-PHD filter superimposed on the Velodyne HDL scan. Lines (in green, yellow and cyan) indicating pedestrian tracks with asterix (in red) indicating the clutter that filter failed to remove at the end of the sequence.

filter has been presented that effectively tracks pedestrians from noisy Velodyne HDL scans. The results demonstrate that the proposed algorithm successfully estimates and track the trajectories of the variable number of pedestrians in dynamic urban environments. The tracking case study presented here has a relatively good SNR ratio, however, it has been noted that under high cluttered environments and low SNR, PHD filter (as any other filter) performs poorly. To mitigate this problem, alternatives in form of cardinalized PHD (CPHD) filter [5] have been proposed. Future work will assess the feasibility of applying CPHD filter to track in environments with higher clutter.

Acknowledgment

The authors would like thank Dr. Ba- Ngu Vo for his valuable discussions and feedback on PHD filter.

This work was supported by National Research Foundation (NRF), Singapore and Center for Environmental Sensing And Modeling (CENSAM) under the auspices of the Singapore-MIT Alliance for Research and Technology (SMART).

References

[1] M. Bajracharya, B. Moghaddam, A. Howard, S. Brennan, and L. H. Matthies. A fast stereo-based system for detecting and tracking pedestrians from a moving vehicle. *The International Journal of Robotics Research*, page 0278364909341884, July 2009.

[2] Barshalom.Y and Fortman.T.E. *Tracking and Data Association*. Academic Press, 1988.

[3] D.E.Clark and J.Bell. Multi-target state estimation and track continuity for the particle PHD filter. *IEEE Transactions on Aerospace and Electronic Systems*, 43(3), July 2007.

[4] B. Kalyan, A. Balasuriya, and S. Wijesoma. Multiple target tracking in underwater sonar images using Particle-PHD filter. In *OCEANS 2006 - Asia Pacific*, pages 1–5, Singapore, 2006.

[5] R. Mahler. PHD filters of higher order in target number. *Aerospace and Electronic Systems, IEEE Transactions on*, 43(4):1523–1543, 2007.

[6] R. P. S. Mahler. Multi-target bayes filtering via first-order multi-target moments. *IEEE Transactions on Aerospace and Electronic Systems*, 39(4):1152–1178, 2003.

[7] R. P. S. Mahler. *Statistical Multisource-Multitarget Information Fusion*. Artech House Publishers, Feb. 2007.

[8] L. E. Navarro-Serment, C. Mertz, N. Vandapel, and M. Hebert. LADAR-based pedestrian detection and tracking. In *Proc. 1st. Workshop on Human Detection from Mobile Robot Platforms*, May 2008.

[9] G. Pulford. Taxonomy of multiple target tracking methods. *Radar, Sonar and Navigation, IEE Proceedings* -, 152(5):291–304, 2005.

[10] H. Sidenbladh and S. Wirkander. Tracking random sets of vehicles in terrain. In *Proc. 2003 IEEE Workshop on Multi-Object Tracking*, Madison WI, June 2003.

[11] S. M. Thornton, M. Hoffelder, and D. D. Morris. Multi-sensor detection and tracking of humans for safe operations with unmanned ground vehicles. In *Proceedings of 1st IEEE Workshop on Human Detection from Mobile Platforms*.

[12] M. Tobias and A. D. Lanterman. Probability hypothesis density-based multitarget tracking with bistatic range and doppler observations. *IEE Proc. Radar, Sonar, and Navigation*, 152(3), 2005.

[13] B. Vo and W. Ma. The gaussian mixture probability hypothesis density filter. *IEEE Transactions on Signal Processing*, 54(11):4091–4104, Nov. 2006.

[14] Y. Wang, J. Wu, A. A. Kassim, and W. Huang. Tracking a variable number of human groups in video using probability hypothesis density. In *18th International Conference on Pattern Recognition*, Hong Kong, Aug. 2006.