Order Statistic Estimation with Application to Tracking in Autonomous Driving

Journal:	*IEEE Transactions on Aerospace and Electronic Systems*
Manuscript ID	TAES-2022-0375.R1
Manuscript Type:	Regular Paper
Date Submitted by the Author:	27-Aug-2022
Complete List of Authors:	Yang, Kaipei; University of Connecticut, ECE Bar-Shalom, Yaakov; University of Connecticut, ECE Willett, Peter; University of Connecticut Hunt, Shawn; DENSO International America Inc Southfield
Keywords:	Tracking, Estimation, Lidar Data, Shape measurement, Vehicles

Order Statistic Estimation with Application to Tracking in Autonomous Driving

Kaipei Yang[1], Yaakov Bar-Shalom[1], Peter Willett[1], Shawn Hunt[2]
[1]Dept. of ECE, University of Connecticut, Storrs, CT 06269-4157
[2]DENSO International America Inc.
Email: [1]{kaipei.yang, peter.willett, yaakov.bar-shalom}@uconn.edu, [2]shawn.hunt@na.denso.com

***Abstract*—In this work, we study the order statistic estimation and provide a simple solution to lidar bounding box (BB) centroid estimation for an autonomous vehicle (AV). The estimated BB centroid and its uncertainty will be used in object tracking as measurement and measurement noise variance; the latter is commonly not available from the sensor manufacturer, and is needed for data association and target state estimation. The detected and clustered sensor observations for a single target, on one axis, are assumed to have i) a triangular density, or ii) a uniform (rectangular) density. These densities constitute the likelihood function (LF) of the corresponding support (the interval where they are nonzero) boundaries. Estimators are proposed for the LF support boundary parameter estimation and the centroid coordinates are given by the centers of the boundaries. Experiments using real data are carried out to show the performance of the proposed methods for autonomous driving applications. A comparison with the Max-Min average approach shows the superiority of the proposed algorithm.**

I. INTRODUCTION

Object detection provides information needed for target tracking and plays a core role in autonomous driving. The observation uncertainties (standard deviation of measurement noise)[1] are crucial in achieving high motion state estimation accuracy and consistency[2]. Therefore, capturing reliably the uncertainty in object tracking is indispensable for autonomous driving safety — the major concern. Lidar, as one the most used onboard sensors, can provide color and depth information of the foreground and background. Object detection algorithms using lidar observations have been widely studied. In this work, we study the order statistic estimation with an application of bounding box centroid (position and its uncertainty) estimation detected by a lidar.

While the uncertainty of bounding box labeling has been studied in the literature using machine learning (ML) and neural networks (NN) with deep learning, little work has addressed the connection between the label uncertainty and measurement uncertainty, which is needed for object tracking. The concept of epistemic uncertainty (caused by process noise) and aleatoric uncertainty (caused by measurement noise) was introduced in [10], which proposed a Bayesian Deep Learning algorithm for obtaining the aforementioned uncertainties. Meyer et al. [13] proposed an improved lidar-based object detector by combing the benefits of the Kullback-Leibler Divergence loss and lidar label uncertainty. The authors proposed in [18] a generative model to estimate bounding box label uncertainties from lidar point clouds, and defined a new representation of the probabilistic bounding box through spatial distribution. Based on [10], Meyer et al. [14] introduced an efficient probabilistic 3D object detector using lidar. The mean and variance of the bounding box are discussed with the assumption that each point from the bounding box shares the same variance. A novel form of the loss function (via neural networks) was proposed in [16] to improve the performance of lidar-based object detection and discussed the uncertainty of an individual measurement for prediction. A sampling-free approach for computing the epistemic uncertainty of a neural network was presented in [17]. A review of uncertainty quantification in deep learning is given in [1]. Center-based object tracking via machine learning using 3D bounding box is introduced in [5]. Bai te al. [4] proposed TransFusion for lidar-camera fusion where a feed-forward network (FFN) is used to predicts the bounding box center offset and shape.

Model-based extended object tracking can provide target shape (including the centroid) information, however, it is done in the process of filtering along with motion state estimation [5]. Granström el al. [7] proposed a method to solve the problem of tracking a rectangle target using laser measurements by using sample covariance eigenvalues and the reference point, i.e., the nearest angle of the rectangular to the laser range sensor. A Gaussian mixture probability hypothesis density (GM-PHD) filter is used for the estimation of an extended state vector including target motion and shape parameters. More details are given in [8] for extended object tracking with various approaches for shape estimation, which is coupled with motion state estimation. Turning to real-world sensor fusion systems, the application for extended object tracking was introduced in [21], in which the authors presented a tracker using an interacting multiple model (IMM) estimator for kinematic information and a truncated Gaussian scheme for shape (width/length/orientation) estimation, and a hierarchical association method according to both kinematic and shape information. In [11], the authors considered ellipsoidal object extensions, which are defined by symmetric, positive definite random matrices. Note that the above literature assumed known measurement noise statistics.

In some AV applications, the representative kinematic be-

[1]The measurement noises are commonly assumed to be white zero-mean Gaussian.

[2]Consistency means the filter calculated uncertainties (standard deviations) associated with the state estimates match statistically the actual errors.

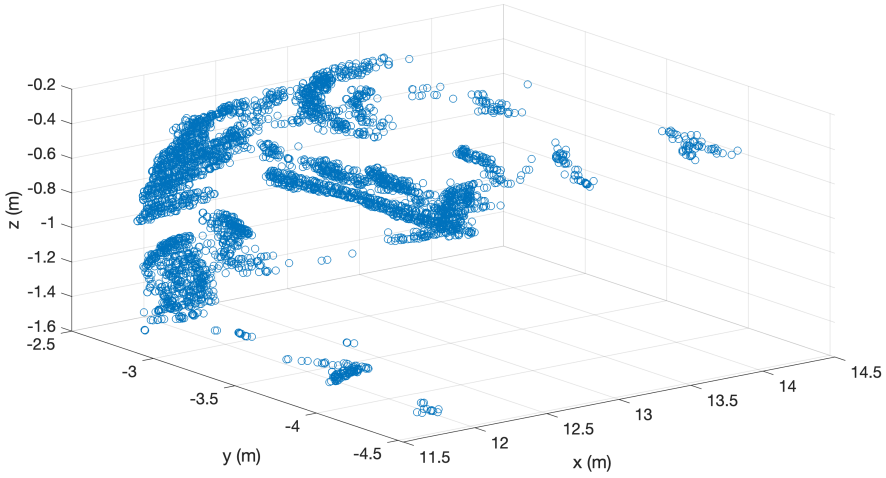


Fig. 1: Detected and clustered position observations for a single target from lidar

haviors of the object are described by the state of a reference point, which is commonly chosen to be the nearest observable corner to the ego vehicle. However, the shape estimation suffers from problems such as an erratic reference point as the object-sensor geometry changes, and insufficient object kinematic information, i.e., the reference point cannot fully represent the object behaviors (e.g., a smoothly turning object might be observed as static if the reference point is the rotation pivot point). In consequence, it is necessary to estimate the bounding box centroid and its uncertainty.

In this work, we investigate the order statistic estimation for a general triangular density and, focusing on a rectangular AV bounding box, proposed a simple solution for centroid estimation, which is decoupled from state estimation (this is not extended target tracking). Centroid estimation is a pre-process of measurement-to-object association and target motion tracking. Fig. 1 shows the detected and clustered observations for a target of interest. The observations, along each axis, are assumed to have: 1) a triangular density or 2) a uniform density. Empirically, the uniform is a questionable assumption compared to the triangular one, but also provides significant improvement in centroid estimation compared to the Max-Min average method. The high resolution lidar measurements from an oncoming vehicle are more likely to have a triangular pdf, i.e., there will be more measurements around its corner nearest to the lidar than the farthest — this motivated the use of a triangular pdf. We also considered a "generalized triangular" pdf[3] where the hypotenuse is not a straight line but a (convex "U") parabola or a cubic. The distribution of sensor observations depends on sensor-object geometry. An unbiased estimator is then used to obtain the support boundary parameters and its uncertainty via order statistics estimation for the triangular density and Maximum Likelihood estimation for the rectangular density. For AV applications, the approach is used for lidar bounding box centroid estimation based on detected and clustered lidar point clouds. We assume that object detection is a preprocessing of the bounding box centroid estimation and it is available in the detector. The bounding box estimation results will then be used in target motion tracking.

Sec. II formulates the problem and introduces order statistics for a generalized triangular density. The order statistic estimation is introduced in Sec. II-A and the support boundary estimation is shown in Sec. II-B. Generalized triangular pdf based centroid estimation is introduced in Sec. III. The detailed discussion for uniform pdf based Maximum Likelihood bounding box centroid estimation is in Sec. IV. The scenario setup and numerical results can be found in Sec. V. Conclusions are given in Sec. VI.

II. ORDER STATISTICS FOR GENERALIZED TRIANGULAR DENSITY

In this section, we discuss the order statistic estimation for a generalized triangular density. Assuming the lidar measurement (clustered) in a generic coordinate x has a density with Probability Density Function (pdf) and Cumulative Distribution Function (cdf)

$$f(x) = \begin{cases} \frac{(p+1)(b-x)^p}{(b-a)^{p+1}} & ; \quad a \le x \le b \\ 0 & \text{else} \end{cases} \tag{1}$$

$$F(x) = \begin{cases} 0 & x < a \\ 1 - \left(\frac{b-x}{b-a}\right)^{p+1} & a \le x \le b \\ 1 & x > b \end{cases} \tag{2}$$

within the support boundaries a, b; $p = 1$ is a right triangle, $p = 2$ is a "parabolic triangle" (convex "U"), $p = 3$ is a "cubic triangle". The clustered measurements are assumed to be i.i.d. and the total number is n. Specially, for target i we have $n = n^i$ (27). The superscript is dropped here for simplicity. Define $x_{(k)}$ as the k^{th} order statistic, i.e, the k^{th} element when they are sorted in an ascending order in the coordinate considered.

[3]We admit that the term "generalized triangular" can be misleading but there does not seem to be one that is better.

In general, for a population of i.i.d. random variables having pdf $f(\cdot)$ and cdf $F(\cdot)$, the pdf of the k^{th} order statistic is

$$f_k(x_{(k)}) = \left(\begin{array}{c} n \\ k \end{array} \right) k f(x_{(k)}) F(x_{(k)})^{k-1} (1 - F(x_{(k)}))^{n-k} \quad (3)$$

A. Order Statistic Estimation

For the variable defined in (1) and (2), the k^{th} order statistic has the expected value

$$\mathcal{E}\{x_{(k)}\} = \int_a^b x \left(\begin{array}{c} n \\ k \end{array} \right) k f(x_{(k)}) F(x_{(k)})^{k-1} (1 - F(x_{(k)}))^{n-k} dx \quad (4)$$

$$= \int_a^b x \left(\begin{array}{c} n \\ k \end{array} \right) k \frac{(p+1)(b-x)^p}{(b-a)^{p+1}} \left(1 - \left(\frac{b-x}{b-a} \right)^{p+1} \right)^{k-1} \times \left(\left(\frac{b-x}{b-a} \right)^{p+1} \right)^{n-k} dx \quad (5)$$

$$= b - \int_a^b \left(\begin{array}{c} n \\ k \end{array} \right) k \frac{(p+1)(b-x)^{p+1}}{(b-a)^{p+1}} \left(1 - \left(\frac{b-x}{b-a} \right)^{p+1} \right)^{k-1} \times \left(\left(\frac{b-x}{b-a} \right)^{p+1} \right)^{n-k} dx \quad (6)$$

Letting

$$m = \left(\frac{b-x}{b-a} \right)^{p+1} \quad (7)$$

we have

$$dx = -\frac{b-a}{p+1} m^{\frac{-p}{p+1}} dm \quad (8)$$

and

$$\mathcal{E}\{x_{(k)}\} = b - \int_0^1 \left(\begin{array}{c} n \\ k \end{array} \right) k(p+1) m (1-m)^{k-1} \cdot m^{n-k} \frac{b-a}{p+1} m^{\frac{-p}{p+1}} dm \quad (9)$$

$$= b - \left(\begin{array}{c} n \\ k \end{array} \right) k(b-a) \int_0^1 m^{\alpha_1 - 1} (1-m)^{\beta_1 - 1} \quad (10)$$

$$= b - \left(\begin{array}{c} n \\ k \end{array} \right) k(b-a) B(\alpha_1, \beta_1) \cdot \int_0^1 \frac{1}{B(\alpha_1, \beta_1)} m^{\alpha_1 - 1} (1-m)^{\beta_1 - 1} \quad (11)$$

$$= b - \left(\begin{array}{c} n \\ k \end{array} \right) k B(\alpha_1, \beta_1)(b-a) \quad (12)$$

where

$$\alpha_1 = n - k + \frac{1}{p+1} + 1 \quad \beta_1 = k \quad (13)$$

and $B(\alpha_1, \beta_1) = \frac{\Gamma(\alpha_1)\Gamma(\beta_1)}{\Gamma(\alpha_1 + \beta_1)}$ is the beta function. Note in (11), we have an integral of a beta distribution pdf.

The second moment of the k^{th}order statistic is

$$\mathcal{E}\{x_{(k)}^2\} = \int_a^b x^2 \left(\begin{array}{c} n \\ k \end{array} \right) k f(x_{(k)}) F(x_{(k)})^{k-1} \cdot (1 - F(x_{(k)}))^{n-k} dx \quad (14)$$

$$= \int_a^b \left[(b-x)^2 + 2xb - b^2 \right] \left(\begin{array}{c} n \\ k \end{array} \right) k \frac{(p+1)(b-x)^p}{(b-a)^{p+1}} \times \left(1 - \left(\frac{b-x}{b-a} \right)^{p+1} \right)^{k-1} \left(\left(\frac{b-x}{b-a} \right)^{p+1} \right)^{n-k} dx \quad (15)$$

$$= \int_a^b \left(\begin{array}{c} n \\ k \end{array} \right) k(p+1)(b-a) \frac{(b-x)^{p+2}}{(b-a)^{p+2}} \times \left(1 - \left(\frac{b-x}{b-a} \right)^{p+1} \right)^{k-1} \left(\left(\frac{b-x}{b-a} \right)^{p+1} \right)^{n-k} dx + 2b\mathcal{E}\{x_{(k)}\} - b^2 \quad (16)$$

Again, we use (7) and the above becomes

$$\mathcal{E}\{x_{(k)}^2\} = \int_0^1 \left(\begin{array}{c} n \\ k \end{array} \right) k(p+1)(b-a) m^{\frac{p+2}{p+1}} \times (1-m)^{k-1} m^{n-k} \frac{b-a}{p+1} m^{\frac{-p}{p+1}} dm + 2b\mathcal{E}\{x_{(k)}\} - b^2 \quad (17)$$

$$= \int_0^1 \left(\begin{array}{c} n \\ k \end{array} \right) k(b-a)^2 m^{\alpha_2 - 1} (1-m)^{\beta_2 - 1} dm + 2b\mathcal{E}\{x_{(k)}\} - b^2 \quad (18)$$

$$= \left(\begin{array}{c} n \\ k \end{array} \right) k(b-a)^2 B(\alpha_2, \beta_2) + 2b\mathcal{E}\{x_{(k)}\} - b^2 \quad (19)$$

where

$$\alpha_2 = n - k + \frac{2}{p+1} + 1 \quad \beta_2 = k \quad (20)$$

B. PDF Support Boundary Estimation

Define

$$\gamma_{n,k} = \left(\begin{array}{c} n \\ k \end{array} \right) k B(\alpha_1, \beta_1) \quad (21)$$

then the expected value of the k^{th}order statistic (12) becomes

$$\mathcal{E}\{x_{(k)}\} = \gamma_{n,k} a + (1 - \gamma_{n,k}) b \quad (22)$$

We propose the following estimators, based on [19]

$$\begin{aligned} \hat{a} &\equiv \mu_k^a x_{(k)} + \mu_l^a x_{(l)} \\ \hat{b} &\equiv \mu_k^b x_{(k)} + \mu_l^b x_{(l)} \end{aligned} \quad (23)$$

for arbitrary order statistics k and l. In order that these be unbiased estimators of a and b we require

$$\left(\begin{array}{cc} (1-\gamma_{n,k}) & (1-\gamma_{n,l}) \\ \gamma_{n,k} & \gamma_{n,l} \end{array} \right) \left(\begin{array}{c} \mu_k^a \\ \mu_l^a \end{array} \right) = \left(\begin{array}{c} 0 \\ 1 \end{array} \right) \quad (24)$$

$$\left(\begin{array}{cc} (1-\gamma_{n,k}) & (1-\gamma_{n,l}) \\ \gamma_{n,k} & \gamma_{n,l} \end{array} \right) \left(\begin{array}{c} \mu_k^b \\ \mu_l^b \end{array} \right) = \left(\begin{array}{c} 1 \\ 0 \end{array} \right) \quad (25)$$

III. TRIANGULAR PDF BASED BOUNDING BOX CENTROID ESTIMATION

In object clustering (by discretizing the top-down view into bins [14]), a set of lidar points S_i whose centroids – absolute box centers – fall into the same bin (cluster) i form a single bounding box for object i. Define the bin indicator

$$\alpha^i(k) = \begin{cases} 1 & k \in S_i \ (\mathbf{c}(k) \text{ falls in bin } i) \\ 0 & \text{otherwise} \end{cases} \qquad k = 1, \ldots, N \tag{26}$$

where N is the total number of detections in the scan of interest.

The total number of points in the set S_i, with membership indicator $\alpha^i(k)$, is

$$n^i = \sum_{k=1}^{N} \alpha^i(k) \tag{27}$$

After clustering, the position observation set for target i is defined as

$$\mathbf{O}^i = \mathrm{col}\big[[x(k)\ y(k)\ z(k)]'\big]_{k \in S_i} \tag{28}$$

which is a stacked vector of all the clustered 3D position vectors.

The estimation for the centroid will be done independently across coordinates. Thus, as an illustration, the following analysis is only for the x-axis, where the results can be used for any axis. A subset of (28), only considering the x-axis, is defined as

$$\mathbf{O}_x^i \triangleq \mathrm{col}\,[x(k)]_{k \in S_i} \tag{29}$$

which contains n^i, given in (27), position measurements. Assuming the observations are i.i.d. random variables having pdf (1) and cdf (2), the support boundary estimate is $(\hat{a}^i, \hat{b}^i)$. The centroid of the bounding box can be obtained by

$$\hat{c}_{x,T}^i = \frac{1}{2}(\hat{a}^i + \hat{b}^i) \tag{30}$$

The subscript T indicates triangular pdf based estimation.

The variance of the centroid is

$$\sigma_{x,T}^2 \triangleq P(\hat{c}_{x,T}^i) = \frac{1}{4}[P(a^i) + 2P(a^i, b^i) + P(b^i)] \tag{31}$$

From Sec. II, using the 1st and nth order statistics in (23), i.e., $l = 1$ and $k = n$, we have

$$\begin{aligned} \hat{a} &= \mu_{(n^i)}^a x_{(n^i)} + \mu_1^a x_{(1)} \\ \hat{b} &= \mu_{(n^i)}^b x_{(n^i)} + \mu_1^b x_{(1)} \end{aligned} \tag{32}$$

where

$$\mu_{(n^i)}^a = \frac{B(n^i + \frac{1}{p+1}, 1) - 1}{n^i \left[B(n^i + \frac{1}{p+1}, 1) - B(1 + \frac{1}{p+1}, n^i)\right]} \tag{33}$$

$$\mu_1^a = \frac{1 - B(1 + \frac{1}{p+1}, n^i)}{n^i \left[B(n^i + \frac{1}{p+1}, 1) - B(1 + \frac{1}{p+1}, n^i)\right]} \tag{34}$$

$$\mu_{(n^i)}^b = \frac{B(n^i + \frac{1}{p+1}, 1)}{n^i \left[B(n^i + \frac{1}{p+1}, 1) - B(1 + \frac{1}{p+1}, n^i)\right]} \tag{35}$$

$$\mu_1^b = \frac{-B(1 + \frac{1}{p+1}, n^i)}{n^i \left[B(n^i + \frac{1}{p+1}, 1) - B(1 + \frac{1}{p+1}, n^i)\right]} \tag{36}$$

Similarly, the variance of the centroid can be obtained using (31) with

$$\begin{aligned} P(a^i) &= \mathrm{Var}(\mu_{\mathrm{n}}^{\mathrm{a^i}} \mathrm{x}_{(\mathrm{n^i})} + \mu_1^{\mathrm{a^i}} \mathrm{x}_{(1)}) \\ &\approx (\mu_{n^i}^{a^i})^2 \mathrm{Var}(\mathrm{x}_{(\mathrm{n^i})}) + (\mu_1^{\mathrm{a^i}})^2 \mathrm{Var}(\mathrm{x}_{(1)}) \qquad (37) \\ P(b^i) &= \mathrm{Var}(\mu_{\mathrm{n}}^{\mathrm{b}} \mathrm{x}_{(\mathrm{n^i})} + \mu_1^{\mathrm{b^i}} \mathrm{x}_{(1)}) \\ &\approx (\mu_n^{b^i})^2 \mathrm{Var}(\mathrm{x}_{(\mathrm{n^i})}) + (\mu_1^{\mathrm{b^i}})^2 \mathrm{Var}(\mathrm{x}_{(1)}) \qquad (38) \\ P(a^i, b^i) &= \mathrm{Cov}[(\mu_{\mathrm{n}}^{\mathrm{a^i}} \mathrm{x}_{(\mathrm{n^i})} + \mu_1^{\mathrm{a^i}} \mathrm{x}_{(1)}), (\mu_{\mathrm{n}}^{\mathrm{b^i}} \mathrm{x}_{(\mathrm{n^i})} + \mu_1^{\mathrm{b^i}} \mathrm{x}_{(1)})] \\ &\approx \mu_{n^i}^{a^i} \mu_{n^i}^{b^i} \mathrm{Var}(\mathrm{x}_{(\mathrm{n^i})}) + \mu_1^{\mathrm{a^i}} \mu_1^{\mathrm{b^i}} \mathrm{Var}(\mathrm{x}_{(1)}) \qquad (39) \end{aligned}$$

assuming independence between $x_{(1)}$ and $x_{(n^i)}$. We can use the first and second moments to obtain the variance

$$\mathrm{Var}(x_{(1)}) = \mathcal{E}\{x_{(1)}^2\} - \mathcal{E}\{x_{(1)}\}^2 \tag{40}$$

$$\mathrm{Var}(x_{(n^i)}) = \mathcal{E}\{x_{(n^i)}^2\} - \mathcal{E}\{x_{(n^i)}\}^2 \tag{41}$$

Note that since the true boundary $[a\ b]$ are not available, boundary estimates $[\hat{a}\ \hat{b}]$ should be used to obtain the above centroid variance with

$$\mathcal{E}\{x_{(1)}\} = \hat{b} - n^i B(n^i + \frac{1}{p+1}, 1)(\hat{b} - \hat{a}) \tag{42}$$

$$\mathcal{E}x_{(1)}^2 = n^i(\hat{b} - \hat{a})^2 B(n^i + \frac{2}{p+1}, 1) + 2\hat{b}\mathcal{E}\{x_{(1)}\} - \hat{b}^2 \tag{43}$$

$$\mathcal{E}\{x_{(n^i)}\} = \hat{b} - n^i B(1 + \frac{1}{p+1}, n^i)(\hat{b} - \hat{a}) \tag{44}$$

$$\mathcal{E}\{x_{(n^i)}^2\} = n^i(\hat{b} - \hat{a})^2 B(1 + \frac{2}{p+1}, n^i) + 2\hat{b}\mathcal{E}\{x_{(n^i)}\} - \hat{b}^2 \tag{45}$$

Thus, one can use (30) and (31) as centroid measurement and measurement uncertainty, respectively, in object tracking.

IV. UNIFORM PDF BASED ESTIMATION

Assume the observation set (29) is a set of independent variables drawn from a uniform distribution $\mathcal{U}(x_L^i, x_U^i)$, i.e., a rectangular pdf, and

$$\mathbf{x}^i = [x_L^i\ x_U^i]' \tag{46}$$

is the nonrandom parameter vector that defines the lower (L) and upper (U) boundaries of the support of the likelihood function of $\mathbf{x}$ based on $\mathbf{O}^i$ (the pdf of $\mathbf{O}^i$, with the subscript dropped, conditioned on x)

$$\Lambda(\mathbf{x}^i; \mathbf{O}^i) = \prod_{j=1}^{n^i} p(x(\ell_j^i) | x_L^i, x_U^i) = \frac{1}{(x_U^i - x_L^i)^{n^i}} \tag{47}$$

The unbiased ML (maximum likelihood) estimator for (46) is [19]

$$\hat{x}_L^i = \frac{n^i o_m^i - o_M^i}{n^i - 1}, \qquad \hat{x}_U^i = \frac{n^i o_M^i - o_m^i}{n^i - 1} \tag{48}$$

where

$$o_m^i \triangleq \min\{o_x^i(\ell_1^i), \ldots, o_x^i(\ell_{n^i}^i)\} \tag{49}$$

$$o_M^i \triangleq \max\{o_x^i(\ell_1^i), \ldots, o_x^i(\ell_{n^i}^i)\} \tag{50}$$

The corresponding covariance matrix is

$$P^i(\mathbf{x}) = E\left[(\hat{\mathbf{x}}^i - \mathbf{x}^i)(\hat{\mathbf{x}}^i - \mathbf{x}^i)'\right] = \begin{bmatrix} P^i_{11} & P^i_{12} \\ P^i_{21} & P^i_{22} \end{bmatrix} \quad (51)$$

where

$$P^i_{11} = P^i_{22} = \frac{n^i(x^i_U - x^i_L)^2}{(n^i-1)(n^i+1)(n^i+2)} \quad (52)$$

and

$$P^i_{12} = P^i_{21} = \frac{-(x^i_U - x^i_L)^2}{(n^i-1)(n^i+1)(n^i+2)} \quad (53)$$

The centroid of the bounding on x-axis with the boundary of $[\hat{x}^i_L \quad \hat{x}^i_U]$ can be obtained by

$$\hat{c}^i_{x,R} = \frac{1}{2}(\hat{x}^i_L + \hat{x}^i_U) \quad (54)$$

with corresponding variance

$$\begin{aligned} \sigma^2_{x,R} =& E\left[(\hat{c}^i_{x,U} - \frac{1}{2}(x^i_l + x^i_u))^2\right] \\ =& \frac{1}{4}P^i_{11} + \frac{1}{2}P^i_{12} + \frac{1}{4}P^i_{22} \\ =& \frac{(x^i_U - x^i_L)^2}{2(n^i+1)(n^i+2)} \end{aligned} \quad (55)$$

Note that, to implement the above, one needs to use the estimates $(\hat{x}^i_L, \hat{x}^i_U)$ instead of (x^i_l, x^i_u) for approximation since the true values are not available. Following the same procedure, the centroid of the bounding on y-axis and z-axis can be obtained using $\mathbf{O}^i_y$ and $\mathbf{O}^i_z$, respectively.

V. NUMERICAL RESULTS

A. Simulation Results

In this subsection, performance (sensitivity) of the proposed centroid estimator is tested by using various generalized triangular densities, i.e., different p in (1). Random samples (300 samples in simulation) are drawn from the true support boundaries $[5 \ 9]$ with a given triangular density for each experiment. The centroid (center of the support boundaries) estimation Root Mean Square Error (RMSE) from 100 Monte-Carlo runs are given in Table I. Calculated centroid variances, from a single run (they vary slightly across runs), are also shown in the same Table. It can be seen that the estimator is consistent, i.e., RMSE matches $\sigma_{x,T}$ statistically. Performance degradation caused by model mismatching is within an acceptable range (maximum RMSE of 0.43m and true centroid is located at 7 m.) with statistically acceptable variance that can be used in object tracking as position measurement noise and its noise variance, respectively. In addition, uniform pdf based estimation is carried out for the same experiments for 100 MC runs and the results are given in Table II. Although the centroid estimation provides practical results, the estimator is not consistent with small calculated variance that should not be utilized in object tracking to avoid inaccurate state estimation results. Additionally, the proposed algorithms are compared with the naïve but commonly used approach in AV applications: Max-Min average (on x-axis for illustration)

$$\hat{c}^i_{x,S} = \frac{1}{2}(o^i_m + o^i_M) \quad (56)$$

Note that the above cannot provide uncertainty for the bounding box centroid.

B. Real-data Testing

In the real-world verification setup that was carried out, a vehicle equipped with a Velodyne VLS-128 mounted on the top of the roof, shown in Figure 2, was parked in a parking lot. Another car was parked in front of the ego-vehicle. A computer receiving the raw Velodyne data passed the data to a machine learning algorithm [14] that was trained to recognize vehicles. The algorithm calculated a 3D bounding box for each detected vehicle, shown in Figure 3. The dimensions and position of the bounding boxes were saved, along with the raw point cloud for offline analysis.

The estimation is carried out at 10 Hz for 1 s (10 scans) and the numerical results are shown in Fig. 4 for x and y. The ground truth is measured manually off-line for performance evaluation. It can be seen that the approach of triangular pdf based unbiased estimation has superior performance with the smallest error among all the algorithms.

The centroid estimation uncertainty is also studied. Tables III and IV show the results from 5 scans: estimation error (estimate (30) minus ground truth) and estimation uncertainty (square root of (31)) from triangular pdf based estimation with different p in (1) and (2) for x and y. It can be seen that, when $p = 2$ for x and $p = 1$ for y, the estimator has better consistency, i.e., this uncertainty (standard deviation) matches the actual error statistically. While the estimation error is small, the consistency is lost with $p = 3$ for x due to model over-fitting — the calculated variance cannot be used in tracking as measurement noise variance. The estimation uncertainty obtained from the uniform pdf based unbiased approach is shown in Table V (estimate (54) minus ground truth). The very small calculated variance is unrealistic and optimistic due to the fact of model mismatch. Note that, prior to centroid estimation, one can apply model-matching for the collected data (find p in (1)) to improve the estimation accuracy.

The proposed centroid estimation algorithm can be run in real-time for both static and moving objects. Since the ground truth was not available for moving objects in the current work, the algorithm performance will be further investigated in a future study.

VI. CONCLUSIONS

In this work, we proposed a simple solution based on order statistic estimation with useful application for lidar bounding box centroid estimation in autonomous driving. Different from data-driven GNN algorithms used in computer vision or complex extended target tracking, the proposed centroid estimation is model-based and independent from state estimation. While providing reliable solution for measurement uncertainty, the complexity of the problem is moderate in this work. Centroid estimation is a pre-process of measurement-to-target association (and target tracking later) using clustered observations, i.e., point cloud from lidar. When the uncertainty for each detected point is not available, one can assume either a triangular density or a uniform density of the detected and clustered points on a single rectangular bounding box. An unbiased estimator is used for the distribution support boundary

Fig. 2: Lexus RX450hL Equipped with a Velodyne VLS-128

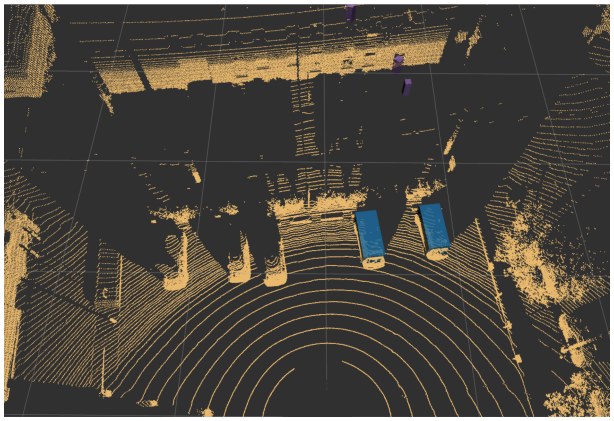


Fig. 3: Vehicle Detection Algorithm Running on Velodyne Point Cloud

TABLE I: Triangle PDF based Centroid Estimation from 100 MC runs

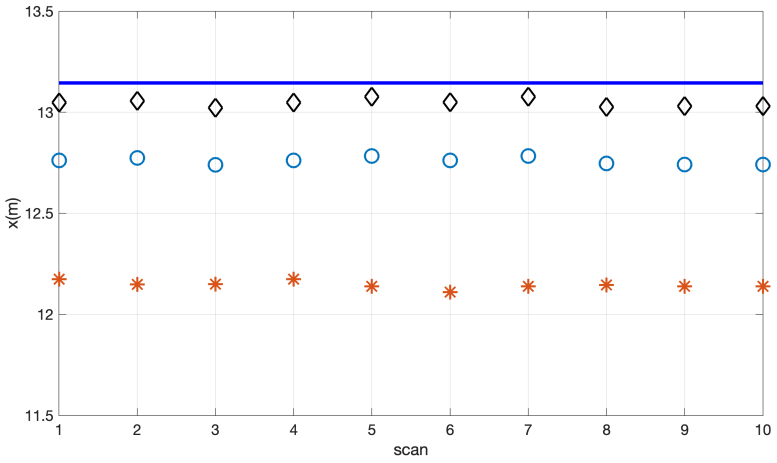
Support Boundary [5 9] m	True Sample Data Distribution					
	p=1 Right Triangle		p=2 Parabolic Triangle		p=3 Cubic Triangle	
	RMSE	$\sigma_{x,T}$	RMSE	$\sigma_{x,T}$	RMSE	$\sigma_{x,T}$
Estimator Model p=1	5.86 cm	5.58 cm	19.3 cm	5.01 cm	37.3 cm	3.92 cm
Estimator Model p=2	20.0 cm	11.1 cm	11.5 cm	10.1 cm	22.3 cm	9.75 cm
Estimator Model p=3	43.3 cm	16.9 cm	25.3 cm	15.4 cm	16.1 cm	13.6 cm

TABLE II: Centroid Estimation from Uniform PDF based Apporcah and Max-Min in Apporch from 100 MC runs

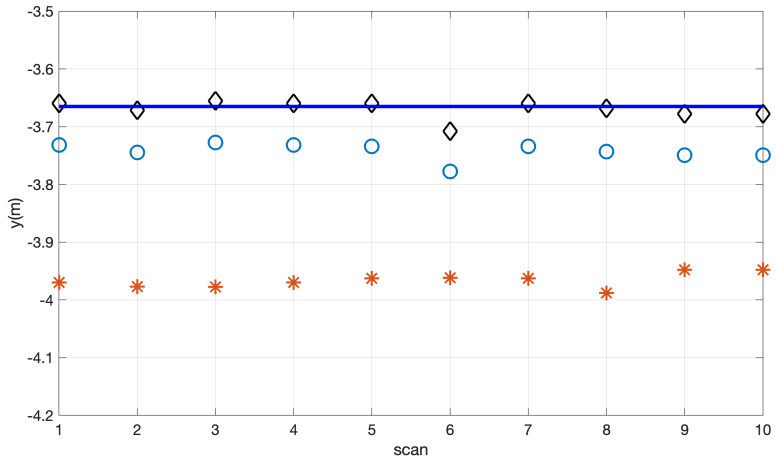
True Sample Data Distribution	p=1 Right Triangle	p=2 Parabolic Triangle	p=3 Cubic Triangle
Uniform pdf based RMSE	10.9 cm	29.1 cm	45.8 cm
Uniform pdf based $\sigma_{x,R}$	0.93 cm	0.89 cm	0.79 cm
Max-Min RMSE	11.4 cm	27.5 cm	46.3 cm

TABLE III: Bounding Box Centroid Estimation Error and Standard Deviation for the: generalized triangular pdf based Estimation with $p = 2$ for x and $p = 1$ for y.

	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD
Num. of Detection	167		170		167		168		172	
x (cm)	-16.5	9.5	-17.4	9.5	-13.8	9.7	-13.9	9.7	-18.4	9.6
y (cm)	-2.4	2.9	-1.2	2.9	-1.7	3.0	-2.4	3.1	-3.4	2.9



(a) Centroid Estimation for x



(a) Centroid Estimation for y

Fig. 4: Centroid Estimation from Difference Algorithms
Ground Truth $*$ Max-Min Average ◯ Unbiased ML Estimate
◊ triangular pdf based Estimate ($p = 2$ for x and $p = 1$ for y)

TABLE IV: Bounding Box Centroid Estimation Error and Standard Deviation for the: generalized triangular pdf based Estimation with $p = 3$ for x and $p = 1$ for y.

	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD
Num of Detection	167		170		167		168		172	
x (cm)	-0.1	14.4	-0.1	14.4	-2.9	14.6	-2.8	14.6	-1.8	14.4
y (cm)	-2.4	2.9	-1.2	2.9	-1.7	3.0	-2.4	3.1	-3.4	2.9

TABLE V: Bounding Box Centroid Estimation Error and Standard Deviation for Uniform PDF based Estimation.

	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD
Num of Detection	167		170		167		168		172	
x (cm)	-38.3	1.93	-39.1	1.92	-36.1	1.88	-36.4	1.97	-40.3	1.88
y (cm)	-6.66	1.21	-7.90	1.24	-6.27	1.30	-6.87	1.28	-7.12	1.27

parameters' estimation for each axis and the centroid coordinates are given by the centers of the boundaries. The proposed centroid estimator is shown to be consistent in simulations. Numerical real-data results show that the proposed estimator can successfully estimate the bounding box centroid while providing a consistent uncertainty that is needed for target motion estimation. In the experiment presented, the estimation with an assumption of a triangular pdf has best performance, i.e., smaller estimation error and better consistency, compared with the estimation with assumption of a uniform distribution or the Max-Min average method.

REFERENCES

[1] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya and V. Makarenkov. "A review of uncertainty quantification in deep learning: Techniques, applications and challenges".*Information Fusion*, 76, pp.243-297.

[2] Y. Bar-Shalom, X. R. Li and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*, Wiley, 2001.

[3] Y. Bar-Shalom, P. K. Willett and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, 2011.

[4] X. Bai, Z. Hu, X. Zhu, Q. Huang, Y. Chen, H. Fu and C. Tai, "TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection With Transformers", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 1090-1099

[5] H .K .Chiu, A. Prioletti, J. Li, and J. Bohg, "Probabilistic 3D multi-object tracking for autonomous driving," *arXiv preprint arXiv:2001.05673*, 2020.

[6] D. Feng, L. Rosenbaum, F. Timm, and K. Dietmayer. "Labels Are Not Perfect: Improving Probabilistic Object Detection via Label Uncertainty." *arXiv preprint arXiv:2008.04168*. Aug. 2020.

[7] K. Granström, C. Lundquist, and U. Orguner, "Tracking rectangular and elliptical extended targets using laser measurements," *Proc. 8th International Information Fusion Conference*. July 2011.

[8] K. Granstrom, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview and applications". *arXiv preprint arXiv:1604.00970*.

[9] M. Himmelsbach, F. V. Hundelshausen, and H. J. Wuensche. "Fast segmentation of 3D point clouds for ground vehicles. *2010 IEEE Intelligent Vehicles Symposium*, pp. 560-565. June 2010.

[10] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" *arXiv preprint arXiv:1703.04977*. Mar. 2017.

[11] W. Koch, and S. Roman. "A Bayesian approach to extended object tracking and tracking of loosely structured target groups." *Proc. 8th International Information Fusion Conference*, vol. 1, 2005.

[12] C. Lundquist, O. Umut and G. Fredrik "Extended target tracking using polynomials with applications to road-map estimation." I*EEE Transactions on Signal Processing* 59, no. 1 (2011): 15-26

[13] G. P. Meyer and N. Thakurdesai. "Learning an uncertainty-aware object detector for autonomous driving." *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 10521–10527. Feb. 2020.

[14] G. P. Meyer , A. Laddha, E. Kee, C. Vallespi-Gonzalez, and C. K. Wellington. "Lasernet: An efficient probabilistic 3d object detector for autonomous driving." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12677-12686. 2019.

[15] M. R. Morelande and N. J. Gordon "Target tracking through a coordinated turn." *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP '05), vol. 4, iv/21— iv/24, 2005.

[16] H. Pan, Z. Wang, W. Zhan, and M. Tomizuka. "Towards better performance and more explainable uncertainty for 3d object detection of autonomous vehicles." *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* , pp. 1-7. Sept. 2020.

[17] J. Postels, F. Ferroni, H. Coskun, N. Navab and F. Tombari, "Sampling-Free Epistemic Uncertainty Estimation Using Approximated Variance Propagation," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 2931-2940.

[18] Z. Wang, D. Feng, Y. Zhou, L. Rosenbaum, F. Timm, K. Dietmayer, M. Tomizuka and W. Zhan, "Inferring spatial uncertainty in object detection". *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 5792–5799. Oct. 2020.

[19] S. Ye, Y. Bar-Shalom and P. K. Willett, "Estimation of the Support Parameters of a Uniform PDF and the Cramér-Rao-Leibniz Lower Bound", *IEEE Signal Processing Letters*, pp. 1765 – 1768, August 2020.

[20] Y. Ye, L. Fu and B. Li. "Object detection and tracking using multi-layer laser for autonomous urban driving." *IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 259-264, Nov 2016.

[21] T. Yuan, et al."Extended object tracking using IMM approach for a real-world vehicle sensor fusion system." *Proc. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. pp. 638–643, IEEE, 2017.

[22] T. Yin, X. Zhou, and P. Krahenbuhl, "Center-based 3D object detectionand tracking," *in Proc. CVPR-21*, Jun. 2021, pp. 11 784–11 793.

Order Statistic Estimation with Application to Tracking in Autonomous Driving

Kaipei Yang[1], Yaakov Bar-Shalom[1], Peter Willett[1], Shawn Hunt[2]

[1]Dept. of ECE, University of Connecticut, Storrs, CT 06269-4157

[2]DENSO International America Inc.

Email: [1]{kaipei.yang, peter.willett, yaakov.bar-shalom}@uconn.edu,

[2]shawn.hunt@na.denso.com

Abstract

In this work, we study the order statistic estimation and provide a simple solution to lidar bounding box (BB) centroid estimation for an autonomous vehicle (AV). The estimated BB centroid and its uncertainty will be used in object tracking as measurement and measurement noise variance; the latter is commonly not available from the sensor manufacturer, and is needed for data association and target state estimation. The detected and clustered sensor observations for a single target, on one axis, are assumed to have i) a triangular density, or ii) a uniform (rectangular) density. These densities constitute the likelihood function (LF) of the corresponding support (the interval where they are nonzero) boundaries. Estimators are proposed for the LF support boundary parameter estimation and the centroid coordinates are given by the centers of the boundaries. Experiments using real data are carried out to show the performance of the proposed methods for autonomous driving applications. A comparison with the Max-Min average approach shows the superiority of the proposed algorithm.

I. INTRODUCTION

Object detection provides information needed for target tracking and plays a core role in autonomous driving. The observation uncertainties (standard deviation of measurement noise)[1]

Draft to TAES

[1]The measurement noises are commonly assumed to be white zero-mean Gaussian.

are crucial in achieving high motion state estimation accuracy and consistency[2]. Therefore, capturing reliably the uncertainty in object tracking is indispensable for autonomous driving safety — the major concern. Lidar, as one the most commonly used onboard sensors, can provide color and depth information of the foreground and background. Object detection algorithms using lidar observations have been extensively studied. In this work, we study the order statistic estimation with an application of bounding box centroid (position and its uncertainty) estimation detected by a lidar.

While the uncertainty of bounding box labeling has been studied in the literature using machine learning (ML) and neural networks (NN) with deep learning, little work has addressed the connection between the label uncertainty and measurement uncertainty, which is needed for object tracking. The concept of epistemic uncertainty (caused by process noise) and aleatoric uncertainty (caused by measurement noise) was introduced in [10], which proposed a Bayesian Deep Learning algorithm for obtaining the aforementioned uncertainties. Meyer et al. [13] proposed an improved lidar-based object detector by combing the benefits of the Kullback-Leibler Divergence loss and lidar label uncertainty. The authors proposed in [18] a generative model to estimate bounding box label uncertainties from lidar point clouds, and defined a new representation of the probabilistic bounding box through spatial distribution. Based on [10], Meyer et al. [14] introduced an efficient probabilistic 3D object detector using lidar. The mean and variance of the bounding box are discussed with the assumption that each point from the bounding box shares the same variance. A novel form of the loss function (via neural networks) was proposed in [16] to improve the performance of lidar-based object detection and discussed the uncertainty of an individual measurement for prediction. A sampling-free approach for computing the epistemic uncertainty of a neural network was presented in [17]. A review of uncertainty quantification in deep learning is given in [1]. Center-based object tracking via machine learning using 3D bounding box is introduced in [5]. Bai et al. [4] proposed TransFusion for lidar-camera fusion where a feed-forward network (FFN) is used to predicts the bounding box center offset and shape.

Model-based extended object tracking can provide target shape (including the centroid) information, however, it is done in the process of filtering along with motion state estimation

[2]Consistency means the filter calculated uncertainties (standard deviations) associated with the state estimates match statistically the actual errors.

[22]. Granström el al. [7] proposed a method to solve the problem of tracking a rectangular target using laser measurements by using sample covariance eigenvalues and the reference point, i.e., the nearest angle of the rectangular to the laser range sensor. A Gaussian mixture probability hypothesis density (GM-PHD) filter is used for the estimation of an extended state vector including target motion and shape parameters. More details are given in [8] for extended object tracking with various approaches for shape estimation, which is coupled with motion state estimation. Turning to real-world sensor fusion systems, the application for extended object tracking was introduced in [21], in which the authors presented a tracker using an interacting multiple model (IMM) estimator for kinematic information and a truncated Gaussian scheme for shape (width/length/orientation) estimation, and a hierarchical association method according to both kinematic and shape information. In [11], the authors considered ellipsoidal object extensions, which are defined by symmetric, positive definite random matrices. Note that the above literature assumed known measurement noise statistics.

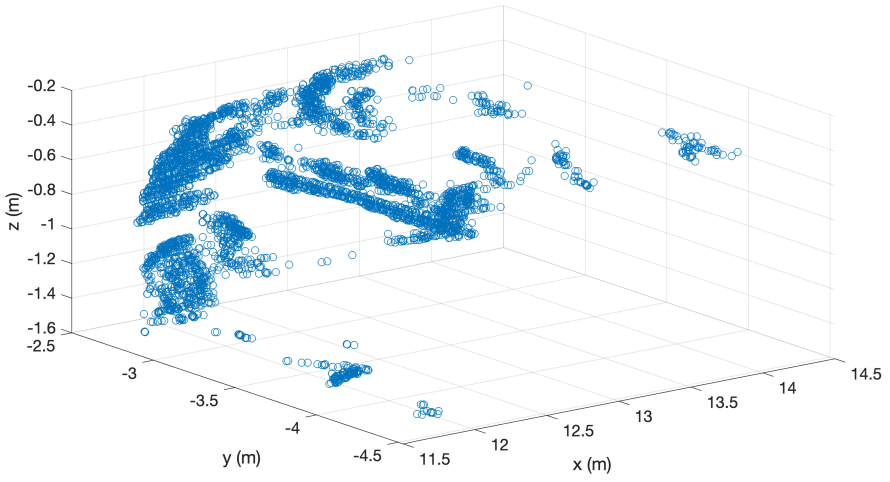


Fig. 1: Detected and clustered position observations for a single target from lidar

In some AV applications, the representative kinematic behaviors of the object are described by the state of a reference point, which is commonly chosen to be the nearest observable corner to the ego vehicle. However, the shape estimation suffers from problems such as an erratic reference point as the object-sensor geometry changes, and insufficient object kinematic information, i.e., the reference point cannot fully represent the object behaviors (e.g., a smoothly turning object might be observed as static if the reference point is the rotation pivot point). In consequence, it

is necessary to estimate the bounding box centroid and its uncertainty.

In this work, we investigate the order statistic estimation for a general triangular density and, focusing on a rectangular AV bounding box, proposed a simple solution for centroid estimation, which is decoupled from state estimation (this is not extended target tracking). Centroid estimation is a pre-process of measurement-to-object association and target motion tracking. Fig. 1 shows the detected and clustered observations for a target of interest. The observations, along each axis, are assumed to have: 1) a triangular density or 2) a uniform density. Empirically, the uniform is a questionable assumption compared to the triangular one, but also provides significant improvement in centroid estimation compared to the Max-Min average method. The high resolution lidar measurements from an oncoming vehicle are more likely to have a triangular pdf, i.e., there will be more measurements around its corner nearest to the lidar than the farthest — this motivated the use of a triangular pdf. We also considered a "generalized triangular" pdf[3] where the hypotenuse is not a straight line but a (convex "U") parabola or a cubic. The distribution of sensor observations depends on sensor-object geometry. An unbiased estimator is then used to obtain the support boundary parameters and its uncertainty via order statistics estimation for the triangular density and Maximum Likelihood estimation for the rectangular density. For AV applications, the approach is used for lidar bounding box centroid estimation based on detected and clustered lidar point clouds. We assume that object detection is a preprocessing of the bounding box centroid estimation and it is available in the detector. The bounding box estimation results will then be used in target motion tracking.

Sec. II formulates the problem and introduces order statistics for a generalized triangular density. The order statistic estimation is introduced in Sec. II-A and the support boundary estimation is shown in Sec. II-B. Generalized triangular pdf based centroid estimation is introduced in Sec. III. The detailed discussion for uniform pdf based Maximum Likelihood bounding box centroid estimation is in Sec. IV. The scenario setup and numerical results can be found in Sec. V. Conclusions are given in Sec. VI.

II. ORDER STATISTICS FOR GENERALIZED TRIANGULAR DENSITY

In this section, we discuss the order statistic estimation for a generalized triangular density. Assuming the lidar measurement (clustered) in a generic coordinate x has a density with Prob-

[3]We admit that the term "generalized triangular" can be misleading but there does not seem to be one that is better.

ability Density Function (pdf) and Cumulative Distribution Function (cdf)

$$f(x) = \begin{cases} \frac{(p+1)(b-x)^p}{(b-a)^{p+1}} & ; \quad a \le x \le b \\ 0 & \text{else} \end{cases} \tag{1}$$

$$F(x) = \begin{cases} 0 & x < a \\ 1 - \left(\frac{b-x}{b-a}\right)^{p+1} & a \le x \le b \\ 1 & x > b \end{cases} \tag{2}$$

within the support boundaries a, b; $p = 1$ is a right triangle, $p = 2$ is a "parabolic triangle" (convex "U"), $p = 3$ is a "cubic triangle". The clustered measurements are assumed to be i.i.d. and the total number is n. Specially, for target i we have $n = n^i$ (27). The superscript is dropped here for simplicity. Define $x_{(k)}$ as the k^{th} order statistic, i.e, the k^{th} element when they are sorted in an ascending order in the coordinate considered.

In general, for a population of i.i.d. random variables having pdf $f(\cdot)$ and cdf $F(\cdot)$, the pdf of the k^{th} order statistic is

$$f_k(x_{(k)}) = \begin{pmatrix} n \\ k \end{pmatrix} k f(x_{(k)}) F(x_{(k)})^{k-1} (1 - F(x_{(k)}))^{n-k} \tag{3}$$

A. *Order Statistic Estimation*

For the variable defined in (1) and (2), the k^{th} order statistic has the expected value

$$\mathcal{E}\{x_{(k)}\} = \int_a^b x \begin{pmatrix} n \\ k \end{pmatrix} k f(x_{(k)}) F(x_{(k)})^{k-1} (1 - F(x_{(k)}))^{n-k} dx \tag{4}$$

$$= \int_a^b x \begin{pmatrix} n \\ k \end{pmatrix} k \frac{(p+1)(b-x)^p}{(b-a)^{p+1}} \left(1 - \left(\frac{b-x}{b-a}\right)^{p+1}\right)^{k-1} \times \left(\left(\frac{b-x}{b-a}\right)^{p+1}\right)^{n-k} dx \tag{5}$$

$$= b - \int_a^b \begin{pmatrix} n \\ k \end{pmatrix} k \frac{(p+1)(b-x)^{p+1}}{(b-a)^{p+1}} \left(1 - \left(\frac{b-x}{b-a}\right)^{p+1}\right)^{k-1} \times \left(\left(\frac{b-x}{b-a}\right)^{p+1}\right)^{n-k} dx \tag{6}$$

Letting

$$m = \left(\frac{b-x}{b-a}\right)^{p+1} \tag{7}$$

we have

$$dx = -\frac{b-a}{p+1} m^{\frac{-p}{p+1}} dm \tag{8}$$

and

$$\begin{aligned}
\mathcal{E}\{x_{(k)}\} &= b - \int_0^1 \left(\begin{array}{c} n \\ k \end{array} \right) k(p+1) m \left(1-m\right)^{k-1} m^{n-k} \frac{b-a}{p+1} m^{\frac{-p}{p+1}} dm & (9)\\
&= b - \left(\begin{array}{c} n \\ k \end{array} \right) k(b-a) \int_0^1 m^{\alpha_1 - 1} \left(1-m\right)^{\beta_1 - 1} & (10)\\
&= b - \left(\begin{array}{c} n \\ k \end{array} \right) k(b-a) B(\alpha_1, \beta_1) \int_0^1 \frac{1}{B(\alpha_1, \beta_1)} m^{\alpha_1 - 1} \left(1-m\right)^{\beta_1 - 1} & (11)\\
&= b - \left(\begin{array}{c} n \\ k \end{array} \right) k B(\alpha_1, \beta_1)(b-a) & (12)
\end{aligned}$$

where

$$\alpha_1 = n - k + \frac{1}{p+1} + 1 \quad \beta_1 = k \tag{13}$$

and $B(\alpha_1, \beta_1) = \frac{\Gamma(\alpha_1)\Gamma(\beta_1)}{\Gamma(\alpha_1+\beta_1)}$ is the beta function. Note in (11), we have an integral of a beta distribution pdf.

The second moment of the k^{th}order statistic is

$$\begin{aligned}
\mathcal{E}\{x_{(k)}^2\} &= \int_a^b x^2 \left(\begin{array}{c} n \\ k \end{array} \right) k f(x_{(k)}) F(x_{(k)})^{k-1} (1 - F(x_{(k)}))^{n-k} dx & (14)\\
&= \int_a^b \left[(b-x)^2 + 2xb - b^2\right] \left(\begin{array}{c} n \\ k \end{array} \right) k \frac{(p+1)(b-x)^p}{(b-a)^{p+1}} \left(1 - \left(\frac{b-x}{b-a}\right)^{p+1}\right)^{k-1} \\
&\qquad \times \left(\left(\frac{b-x}{b-a}\right)^{p+1}\right)^{n-k} dx & (15)\\
&= \int_a^b \left(\begin{array}{c} n \\ k \end{array} \right) k(p+1)(b-a) \frac{(b-x)^{p+2}}{(b-a)^{p+2}} \left(1 - \left(\frac{b-x}{b-a}\right)^{p+1}\right)^{k-1} \\
&\qquad \times \left(\left(\frac{b-x}{b-a}\right)^{p+1}\right)^{n-k} dx + 2b\mathcal{E}\{x_{(k)}\} - b^2 & (16)
\end{aligned}$$

Again, we use (7) and the above becomes

$$\mathcal{E}\{x_{(k)}^2\} = \int_0^1 \binom{n}{k} k(p+1)(b-a)m^{\frac{p+2}{p+1}}(1-m)^{k-1}m^{n-k}\frac{b-a}{p+1}m^{\frac{-p}{p+1}}dm + 2b\mathcal{E}\{x_{(k)}\} - b^2 \quad (17)$$

$$= \int_0^1 \binom{n}{k} k(b-a)^2 m^{\alpha_2-1}(1-m)^{\beta_2-1}dm + 2b\mathcal{E}\{x_{(k)}\} - b^2 \quad (18)$$

$$= \binom{n}{k} k(b-a)^2 B(\alpha_2,\beta_2) + 2b\mathcal{E}\{x_{(k)}\} - b^2 \quad (19)$$

where

$$\alpha_2 = n - k + \frac{2}{p+1} + 1 \quad \beta_2 = k \quad (20)$$

B. PDF Support Boundary Estimation

Define

$$\gamma_{n,k} = \binom{n}{k} kB(\alpha_1,\beta_1) \quad (21)$$

then the expected value of the k^{th}order statistic (12) becomes

$$\mathcal{E}\{x_{(k)}\} = \gamma_{n,k}a + (1-\gamma_{n,k})b \quad (22)$$

We propose the following estimators, based on [19]

$$\begin{aligned}\hat{a} &\equiv \mu_k^a x_{(k)} + \mu_l^a x_{(l)} \\ \hat{b} &\equiv \mu_k^b x_{(k)} + \mu_l^b x_{(l)}\end{aligned} \quad (23)$$

for arbitrary order statistics k and l. In order that these be unbiased estimators of a and b we require

$$\begin{pmatrix} (1-\gamma_{n,k}) & (1-\gamma_{n,l}) \\ \gamma_{n,k} & \gamma_{n,l} \end{pmatrix} \begin{pmatrix} \mu_k^a \\ \mu_l^a \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (24)$$

$$\begin{pmatrix} (1-\gamma_{n,k}) & (1-\gamma_{n,l}) \\ \gamma_{n,k} & \gamma_{n,l} \end{pmatrix} \begin{pmatrix} \mu_k^b \\ \mu_l^b \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (25)$$

III. TRIANGULAR PDF BASED BOUNDING BOX CENTROID ESTIMATION

In object clustering (by discretizing the top-down view into bins [14]), a set of lidar points S_i whose centroids – absolute box centers – fall into the same bin (cluster) i form a single bounding box for object i. Define the bin indicator

$$\alpha^i(k) = \begin{cases} 1 & k \in S_i \ (\mathbf{c}(k) \text{ falls in bin } i) \quad k = 1, \ldots, N \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

where N is the total number of detections in the scan of interest.

The total number of points in the set S_i, with membership indicator $\alpha^i(k)$, is

$$n^i = \sum_{k=1}^{N} \alpha^i(k) \tag{27}$$

After clustering, the position observation set for target i is defined as

$$\mathbf{O}^i = \text{col}\left[[x(k)\ y(k)\ z(k)]'\right]_{k \in S_i} \tag{28}$$

which is a stacked vector of all the clustered 3D position vectors.

The estimation for the centroid will be done independently across coordinates. Thus, as an illustration, the following analysis is only for the x-axis, where the results can be used for any axis. A subset of (28), only considering the x-axis, is defined as

$$\mathbf{O}_x^i \triangleq \text{col}\left[x(k)\right]_{k \in S_i} \tag{29}$$

which contains n^i, given in (27), position measurements. Assuming the observations are i.i.d. random variables having pdf (1) and cdf (2), the support boundary estimate is $(\hat{a}^i, \hat{b}^i)$. The centroid of the bounding box can be obtained by

$$\hat{c}_{x,T}^i = \frac{1}{2}(\hat{a}^i + \hat{b}^i) \tag{30}$$

The subscript T indicates triangular pdf based estimation.

The variance of the centroid is

$$\sigma_{x,T}^2 \triangleq P(\hat{c}_{x,T}^i) = \frac{1}{4}[P(a^i) + 2P(a^i, b^i) + P(b^i)] \tag{31}$$

From Sec. II, using the 1st and nth order statistics in (23), i.e., $l = 1$ and $k = n$, we have

$$\begin{aligned} \hat{a} &= \mu_{(n^i)}^a x_{(n^i)} + \mu_1^a x_{(1)} \\ \hat{b} &= \mu_{(n^i)}^b x_{(n^i)} + \mu_1^b x_{(1)} \end{aligned} \tag{32}$$

where

$$\mu^a_{(n^i)} = \frac{B(n^i + \frac{1}{p+1}, 1) - 1}{n^i \left[B(n^i + \frac{1}{p+1}, 1) - B(1 + \frac{1}{p+1}, n^i)\right]} \tag{33}$$

$$\mu^a_1 = \frac{1 - B(1 + \frac{1}{p+1}, n^i)}{n^i \left[B(n^i + \frac{1}{p+1}, 1) - B(1 + \frac{1}{p+1}, n^i)\right]} \tag{34}$$

$$\mu^b_{(n^i)} = \frac{B(n^i + \frac{1}{p+1}, 1)}{n^i \left[B(n^i + \frac{1}{p+1}, 1) - B(1 + \frac{1}{p+1}, n^i)\right]} \tag{35}$$

$$\mu^b_1 = \frac{-B(1 + \frac{1}{p+1}, n^i)}{n^i \left[B(n^i + \frac{1}{p+1}, 1) - B(1 + \frac{1}{p+1}, n^i)\right]} \tag{36}$$

Similarly, the variance of the centroid can be obtained using (31) with

$$\begin{aligned} P(a^i) &= \mathrm{Var}(\mu^{\mathrm{a^i}}_{\mathrm{n}} \mathrm{x}_{(\mathrm{n^i})} + \mu^{\mathrm{a^i}}_1 \mathrm{x}_{(1)}) \\ &\approx (\mu^{a^i}_{n^i})^2 \mathrm{Var}(\mathrm{x}_{(\mathrm{n^i})}) + (\mu^{\mathrm{a^i}}_1)^2 \mathrm{Var}(\mathrm{x}_{(1)}) \end{aligned} \tag{37}$$

$$\begin{aligned} P(b^i) &= \mathrm{Var}(\mu^{\mathrm{b}}_{\mathrm{n}} \mathrm{x}_{(\mathrm{n^i})} + \mu^{\mathrm{b^i}}_1 \mathrm{x}_{(1)}) \\ &\approx (\mu^{b^i}_n)^2 \mathrm{Var}(\mathrm{x}_{(\mathrm{n^i})}) + (\mu^{\mathrm{b^i}}_1)^2 \mathrm{Var}(\mathrm{x}_{(1)}) \end{aligned} \tag{38}$$

$$\begin{aligned} P(a^i, b^i) &= \mathrm{Cov}[(\mu^{\mathrm{a^i}}_{\mathrm{n}} \mathrm{x}_{(\mathrm{n^i})} + \mu^{\mathrm{a^i}}_1 \mathrm{x}_{(1)}), (\mu^{\mathrm{b^i}}_{\mathrm{n}} \mathrm{x}_{(\mathrm{n^i})} + \mu^{\mathrm{b^i}}_1 \mathrm{x}_{(1)})] \\ &\approx \mu^{a^i}_{n^i} \mu^{b^i}_{n^i} \mathrm{Var}(\mathrm{x}_{(\mathrm{n^i})}) + \mu^{\mathrm{a^i}}_1 \mu^{\mathrm{b^i}}_1 \mathrm{Var}(\mathrm{x}_{(1)}) \end{aligned} \tag{39}$$

assuming independence between $x_{(1)}$ and $x_{(n^i)}$. We can use the first and second moments to obtain the variance

$$\mathrm{Var}(x_{(1)}) = \mathcal{E}\{x^2_{(1)}\} - \mathcal{E}\{x_{(1)}\}^2 \tag{40}$$

$$\mathrm{Var}(x_{(n^i)}) = \mathcal{E}\{x^2_{(n^i)}\} - \mathcal{E}\{x_{(n^i)}\}^2 \tag{41}$$

Note that since the true boundary $[a\ b]$ are not available, boundary estimates $[\hat{a}\ \hat{b}]$ should be used to obtain the above centroid variance with

$$\mathcal{E}\{x_{(1)}\} = \hat{b} - n^i B(n^i + \frac{1}{p+1}, 1)(\hat{b} - \hat{a}) \tag{42}$$

$$\mathcal{E}x^2_{(1)} = n^i(\hat{b} - \hat{a})^2 B(n^i + \frac{2}{p+1}, 1) + 2\hat{b}\mathcal{E}\{x_{(1)}\} - \hat{b}^2 \tag{43}$$

$$\mathcal{E}\{x_{(n^i)}\} = \hat{b} - n^i B(1 + \frac{1}{p+1}, n^i)(\hat{b} - \hat{a}) \tag{44}$$

$$\mathcal{E}\{x^2_{(n^i)}\} = n^i(\hat{b} - \hat{a})^2 B(1 + \frac{2}{p+1}, n^i) + 2\hat{b}\mathcal{E}\{x_{(n^i)}\} - \hat{b}^2 \tag{45}$$

Thus, one can use (30) and (31) as centroid measurement and measurement uncertainty, respectively, in object tracking.

IV. UNIFORM PDF BASED ESTIMATION

Assume the observation set (29) is a set of independent variables drawn from a uniform distribution $\mathcal{U}(x_L^i, x_U^i)$, i.e., a rectangular pdf, and

$$\mathbf{x}^i = [x_L^i \ x_U^i]' \tag{46}$$

is the nonrandom parameter vector that defines the lower (L) and upper (U) boundaries of the support of the likelihood function of $\mathbf{x}$ based on $\mathbf{O}^i$ (the pdf of $\mathbf{O}^i$, with the subscript dropped, conditioned on x)

$$\Lambda(\mathbf{x}^i; \mathbf{O}^i) = \prod_{j=1}^{n^i} p(x(\ell_j^i)|x_L^i, x_U^i) = \frac{1}{(x_U^i - x_L^i)^{n^i}} \tag{47}$$

The unbiased ML (maximum likelihood) estimator for (46) is [19]

$$\hat{x}_L^i = \frac{n^i o_m^i - o_M^i}{n^i - 1}, \quad \hat{x}_U^i = \frac{n^i o_M^i - o_m^i}{n^i - 1} \tag{48}$$

where

$$o_m^i \triangleq \min\{o_x^i(\ell_1^i), \ldots, o_x^i(\ell_{n^i}^i)\} \tag{49}$$

$$o_M^i \triangleq \max\{o_x^i(\ell_1^i), \ldots, o_x^i(\ell_{n^i}^i)\} \tag{50}$$

The corresponding covariance matrix is

$$P^i(\mathbf{x}) = E\left[(\hat{\mathbf{x}}^i - \mathbf{x}^i)(\hat{\mathbf{x}}^i - \mathbf{x}^i)'\right] = \begin{bmatrix} P_{11}^i & P_{12}^i \\ P_{21}^i & P_{22}^i \end{bmatrix} \tag{51}$$

where

$$P_{11}^i = P_{22}^i = \frac{n^i (x_U^i - x_L^i)^2}{(n^i - 1)(n^i + 1)(n^i + 2)} \tag{52}$$

and

$$P_{12}^i = P_{21}^i = \frac{-(x_U^i - x_L^i)^2}{(n^i - 1)(n^i + 1)(n^i + 2)} \tag{53}$$

The centroid of the bounding on x-axis with the boundary of $[\hat{x}_L^i \ \ \hat{x}_U^i]$ can be obtained by

$$\hat{c}_{x,R}^i = \frac{1}{2}(\hat{x}_L^i + \hat{x}_U^i) \tag{54}$$

with corresponding variance

$$\begin{aligned} \sigma_{x,R}^2 =& E\left[(\hat{c}_{x,U}^i - \frac{1}{2}(x_l^i + x_u^i))^2\right] \\ =& \frac{1}{4}P_{11}^i + \frac{1}{2}P_{12}^i + \frac{1}{4}P_{22}^i \\ =& \frac{(x_U^i - x_L^i)^2}{2(n^i + 1)(n^i + 2)} \end{aligned} \tag{55}$$

Note that, to implement the above, one needs to use the estimates $(\hat{x}_L^i, \hat{x}_U^i)$ instead of (x_l^i, x_u^i) for approximation since the true values are not available. Following the same procedure, the centroid of the bounding on y-axis and z-axis can be obtained using $\mathbf{O}_y^i$ and $\mathbf{O}_z^i$, respectively.

Fig. 2: Lexus RX450hL Equipped with a Velodyne VLS-128

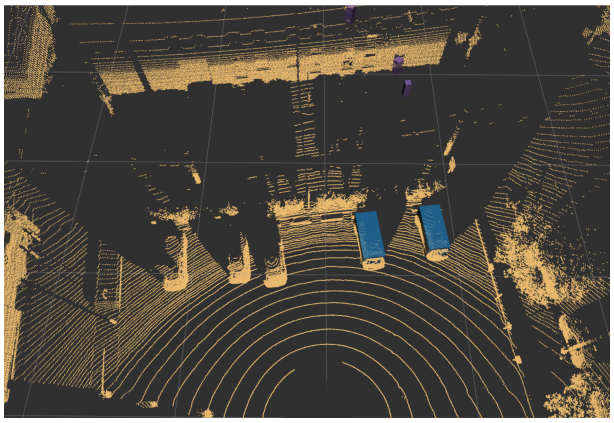


Fig. 3: Vehicle Detection Algorithm Running on Velodyne Point Cloud

V. NUMERICAL RESULTS

A. *Simulation Results*

In this subsection, performance (sensitivity) of the proposed centroid estimator is tested by using various generalized triangular densities, i.e., different p in (1). Random samples (300

samples in simulation) are drawn from the true support interval $[5\ 9]$ with a given triangular density for each experiment. The centroid (center of the support boundaries) estimation Root Mean Square Error (RMSE) from 100 Monte-Carlo runs are given in Table I. Calculated centroid variances, from a single run (they vary slightly across runs), are also shown in the same Table. It can be seen that the estimator is consistent, i.e., RMSE matches $\sigma_{x,T}$ statistically. Performance degradation caused by model mismatch is within an acceptable range (maximum RMSE of 0.43 m and true centroid is located at 7 m) with statistically acceptable variance that can be used in object tracking as position measurement noise and its noise variance, respectively. In addition, uniform pdf based estimation is carried out for the same experiments for 100 MC runs and the results are given in Table II. Although the centroid estimation provides practical results, the estimator is not consistent with small calculated variance that should not be utilized in object tracking to avoid inaccurate state estimation results. Additionally, the proposed algorithms are compared with the naïve but commonly used approach in AV applications: Max-Min average (on x-axis for illustration)

$$\hat{c}^i_{x,S} = \frac{1}{2}(o^i_m + o^i_M) \tag{56}$$

Note that the above cannot provide uncertainty for the bounding box centroid.

TABLE I: Triangle PDF based Centroid Estimation from 100 MC runs

Support Boundary [5 9] m	True Sample Data Distribution					
	p=1 Right Triangle		p=2 Parabolic Triangle		p=3 Cubic Triangle	
	RMSE	$\sigma_{x,T}$	RMSE	$\sigma_{x,T}$	RMSE	$\sigma_{x,T}$
Estimator Model p=1	5.86 cm	5.58 cm	19.3 cm	5.01 cm	37.3 cm	3.92 cm
Estimator Model p=2	20.0 cm	11.1 cm	11.5 cm	10.1 cm	22.3 cm	9.75 cm
Estimator Model p=3	43.3 cm	16.9 cm	25.3 cm	15.4 cm	16.1 cm	13.6 cm

TABLE II: Centroid Estimation from Uniform PDF based Approach and Max-Min Approach from 100 MC runs

True Sample Data Distribution	p=1 Right Triangle	p=2 Parabolic Triangle	p=3 Cubic Triangle
Uniform pdf based RMSE	10.9 cm	29.1 cm	45.8 cm
Uniform pdf based $\sigma_{x,R}$	0.93 cm	0.89 cm	0.79 cm
Max-Min RMSE	11.4 cm	27.5 cm	46.3 cm

B. Real-data Testing

In the real-world verification setup that was carried out, a vehicle equipped with a Velodyne VLS-128 mounted on the top of the roof, shown in Figure 2, was parked in a parking lot. Another car was parked in front of the ego-vehicle. A computer receiving the raw Velodyne data passed the data to a machine learning algorithm [14] that was trained to recognize vehicles. The algorithm calculated a 3D bounding box for each detected vehicle, shown in Figure 3. The dimensions and position of the bounding boxes were saved, along with the raw point cloud for offline analysis.

The estimation is carried out at 10 Hz for 1 s (10 scans) and the numerical results are shown in Fig. 4 for x and y. The ground truth is measured manually off-line for performance evaluation. It can be seen that the approach of triangular pdf based unbiased estimation has superior performance with the smallest error among all the algorithms.

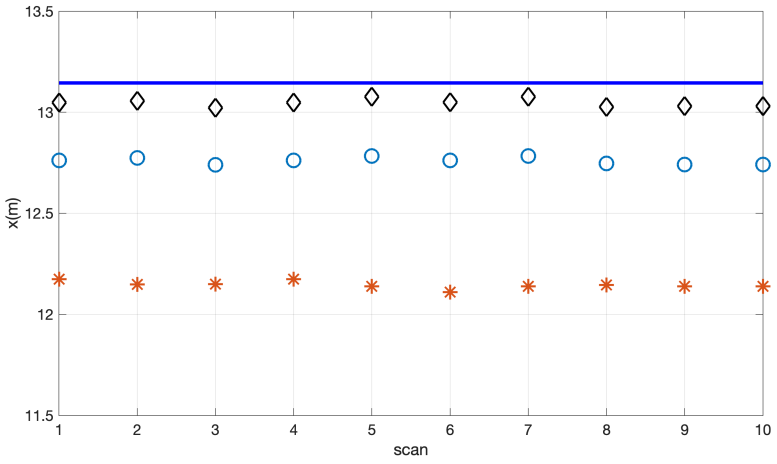
TABLE III: Bounding Box Centroid Estimation Error and Standard Deviation for the: generalized triangular pdf based Estimation with $p = 2$ for x and $p = 1$ for y.

	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD
Num. of Detection	167		170		167		168		172	
x (cm)	-16.5	9.5	-17.4	9.5	-13.8	9.7	-13.9	9.7	-18.4	9.6
y (cm)	-2.4	2.9	-1.2	2.9	-1.7	3.0	-2.4	3.1	-3.4	2.9

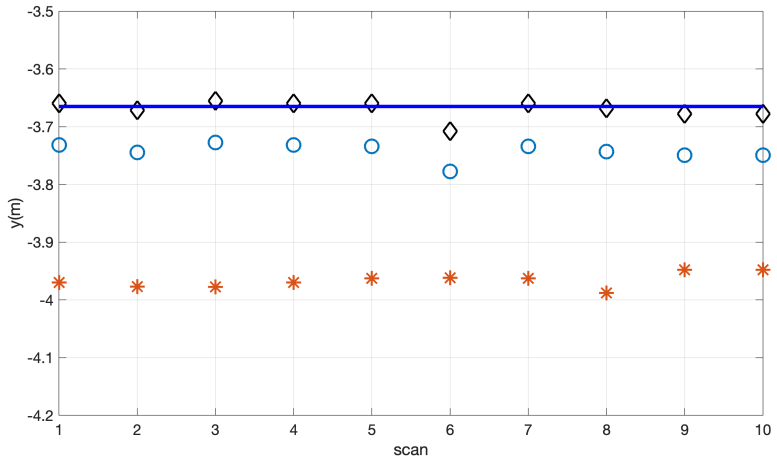
TABLE IV: Bounding Box Centroid Estimation Error and Standard Deviation for the: generalized triangular pdf based Estimation with $p = 3$ for x and $p = 1$ for y.

	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD
Num of Detection	167		170		167		168		172	
x (cm)	-0.1	14.4	-0.1	14.4	-2.9	14.6	-2.8	14.6	-1.8	14.4
y (cm)	-2.4	2.9	-1.2	2.9	-1.7	3.0	-2.4	3.1	-3.4	2.9

The centroid estimation uncertainty is also studied. Tables III and IV show the results from 5 scans: estimation error (estimate (30) minus ground truth) and estimation uncertainty (square root of (31)) from triangular pdf based estimation with different p in (1) and (2) for x and y. It can be seen that, when $p = 2$ for x and $p = 1$ for y, the estimator has better consistency, i.e., this uncertainty (standard deviation) matches the actual error statistically. While the estimation



(a) Centroid Estimation for x



(a) Centroid Estimation for y

Fig. 4: Centroid Estimation from Difference Algorithms

— Ground Truth $*$ Max-Min Average ◯ Unbiased ML Estimate

♢ triangular pdf based Estimate ($p = 2$ for x and $p = 1$ for y)

TABLE V: Bounding Box Centroid Estimation Error and Standard Deviation for Uniform PDF based Estimation.

	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD	Est Error	SD
Num of Detection	167		170		167		168		172	
x (cm)	-38.3	1.93	-39.1	1.92	-36.1	1.88	-36.4	1.97	-40.3	1.88
y (cm)	-6.66	1.21	-7.90	1.24	-6.27	1.30	-6.87	1.28	-7.12	1.27

error is small, the consistency is lost with $p = 3$ for x due to model over-fitting — the calculated variance cannot be used in tracking as measurement noise variance. The estimation uncertainty obtained from the uniform pdf based unbiased approach is shown in Table V (estimate (54) minus ground truth). The very small calculated variance is unrealistic and optimistic due to the fact of model mismatch. Note that, prior to centroid estimation, one can apply model-matching for the collected data (find p in (1)) to improve the estimation accuracy.

The proposed centroid estimation algorithm can be run in real-time for both static and moving objects. Since the ground truth was not available for moving objects in the current work, the algorithm performance will be further investigated in a future study.

VI. CONCLUSIONS

In this work, we proposed a simple solution based on order statistic estimation with useful application for lidar bounding box centroid estimation in autonomous driving. Different from data-driven GNN algorithms used in computer vision or complex extended target tracking, the proposed centroid estimation is model-based and independent from state estimation. While providing reliable solution for measurement uncertainty, the complexity of the problem is moderate in this work. Centroid estimation is a pre-process of measurement-to-target association (and target tracking later) using clustered observations, i.e., point cloud from lidar. When the uncertainty for each detected point is not available, one can assume either a triangular density or a uniform density of the detected and clustered points on a single rectangular bounding box. An unbiased estimator is used for the distribution support boundary parameters' estimation for each axis and the centroid coordinates are given by the centers of the boundaries. The proposed centroid estimator is shown to be consistent in simulations. Numerical real-data results show that the proposed estimator can successfully estimate the bounding box centroid while providing a consistent uncertainty that is needed for target motion estimation. In the experiment presented, the estimation with an assumption of a triangular pdf has best performance, i.e., smaller estimation error and better consistency, compared with the estimation with assumption of a uniform distribution or the Max-Min average method.

REFERENCES

[1] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya and V. Makarenkov. "A review of uncertainty quantification in deep learning: Techniques, applications and challenges".*Information Fusion*, 76, pp.243-297.

[2] Y. Bar-Shalom, X. R. Li and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*, Wiley, 2001.

[3] Y. Bar-Shalom, P. K. Willett and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, 2011.

[4] X. Bai, Z. Hu, X. Zhu, Q. Huang, Y. Chen, H. Fu and C. Tai, "TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection With Transformers", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 1090-1099

[5] H .K .Chiu, A. Prioletti, J. Li, and J. Bohg, "Probabilistic 3D multi-object tracking for autonomous driving," *arXiv preprint arXiv:2001.05673*, 2020.

[6] D. Feng, L. Rosenbaum, F. Timm, and K. Dietmayer. "Labels Are Not Perfect: Improving Probabilistic Object Detection via Label Uncertainty." *arXiv preprint arXiv:2008.04168*. Aug. 2020.

[7] K. Granström, C. Lundquist, and U. Orguner, "Tracking rectangular and elliptical extended targets using laser measurements," *Proc. 8th International Information Fusion Conference*. July 2011.

[8] K. Granstrom, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview and applications". *arXiv preprint arXiv:1604.00970*.

[9] M. Himmelsbach, F. V. Hundelshausen, and H. J. Wuensche. "Fast segmentation of 3D point clouds for ground vehicles. *2010 IEEE Intelligent Vehicles Symposium*, pp. 560-565. June 2010.

[10] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" *arXiv preprint arXiv:1703.04977*. Mar. 2017.

[11] W. Koch, and S. Roman. "A Bayesian approach to extended object tracking and tracking of loosely structured target groups." *Proc. 8th International Information Fusion Conference*, vol. 1, 2005.

[12] C. Lundquist, O. Umut and G. Fredrik "Extended target tracking using polynomials with applications to road-map estimation." I*EEE Transactions on Signal Processing* 59, no. 1 (2011): 15-26

[13] G. P. Meyer and N. Thakurdesai. "Learning an uncertainty-aware object detector for autonomous driving." *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 10521–10527. Feb. 2020.

[14] G. P. Meyer , A. Laddha, E. Kee, C. Vallespi-Gonzalez, and C. K. Wellington. "Lasernet: An efficient probabilistic 3d object detector for autonomous driving." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12677-12686. 2019.

[15] M. R. Morelande and N. J. Gordon "Target tracking through a coordinated turn." *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP '05), vol. 4, iv/21— iv/24, 2005.

[16] H. Pan, Z. Wang, W. Zhan, and M. Tomizuka. "Towards better performance and more explainable uncertainty for 3d object detection of autonomous vehicles." *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)* , pp. 1-7. Sept. 2020.

[17] J. Postels, F. Ferroni, H. Coskun, N. Navab and F. Tombari, "Sampling-Free Epistemic Uncertainty Estimation Using Approximated Variance Propagation," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 2931-2940.

[18] Z. Wang, D. Feng, Y. Zhou, L. Rosenbaum, F. Timm, K. Dietmayer, M. Tomizuka and W. Zhan, "Inferring spatial

uncertainty in object detection". *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 5792–5799. Oct. 2020.

[19] S. Ye, Y. Bar-Shalom and P. K. Willett, "Estimation of the Support Parameters of a Uniform PDF and the Cramér-Rao-Leibniz Lower Bound", *IEEE Signal Processing Letters*, pp. 1765 – 1768, August 2020.

[20] Y. Ye, L. Fu and B. Li. "Object detection and tracking using multi-layer laser for autonomous urban driving." *IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 259-264, Nov 2016.

[21] T. Yuan, et al."Extended object tracking using IMM approach for a real-world vehicle sensor fusion system." *Proc. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. pp. 638–643, IEEE, 2017.

[22] T. Yin, X. Zhou, and P. Krahenbuhl, "Center-based 3D object detectionand tracking," *in Proc. CVPR-21*, Jun. 2021, pp. 11 784–11 793.

Replies to Reviewers' Comments

Please note that the line numbers indicated in the following replies are w.r.t. the one column version manuscript.

Reviewer 1.

1.1 In the first instance, I think that some additional effort in reviewing background research should be undertaken. The paper makes a number of statements such as 'little work has addressed the connection between the label uncertainty and measurement uncertainty' and 'There are no model-based results in the literature on the centroid estimation problem of a bounding box'. The topic of extended object tracking yields 2.2M results. The review paper by Ganstrom et al. 2016 outlines a number of different models attempting to estimate the bounding box of a lidar cluster. Furthermore, there have been some works which have considered the challenge of measuring both aleatoric and epistemic uncertainties in computer vision based detection/classification techniques (for example Postels et al. 2019 and Abdar et al. 2021.

Reply-1.1: Thank you for the comment, we have improved the introduction.

1.1a Additional Suggested Reference.

Rely-1.1a: Thank you for the suggestion. Recommended references have been added. See lines 44–57

1.2 My second major comment is regarding the results. This paper claims 'superiority of the proposed algorithm' based on only a single result. For an IEEE transactions paper, I feel that more thorough testing is required to justify this claim. Specifically, the results should introduce simulated results (e.g. using the MATLAB AD toolbox) with statistically well-defined measurements, should consider different ranges (i.e. lower resolutions and lower number of observation points) and different motion considerations.

Reply 1.2: Thank you for your suggestions. Simulation results have been added in Sec.V-A (lines 174–190) to test the sensitivity of the proposed estimator for different generalized triangular density models. Matlab AD toolbox is not available in the current project but will be incorporated in future research.

1.3 One of the major points that the paper didn't consider is the challenge of associating measurement points to the target objects. Specifically, when a sensor such as a lidar observes an extended object the relationship between the measurements and the extended object is unknown. My understanding was that this paper considered that the object would be fully sampled across two dimensions, and that the bounding box would fully encompass the object. This may be true for close in objects, but for those at range, or only presenting a single dimension (e.g. observing cars in front on a highway), it is unknown how well the proposed technique will perform.

Reply 1.3: Model-based support boundary estimation is done independently across axes. In a 3-D case, assuming independence, the centroid estimate can be obtained by finding the estimates on x-axis, y-axis and z-axis individually. Clustering is beyond the scope of the current paper, i.e., centroid estimation is based on clustered measurements. Measurement-to-target association is carried out after centroid estimation and this is why reliable estimate and uncertainty are crucial since they are used to represent sensor measurement in association.

1.4 The paper uses 'Ref. [x]' to refer to works. The common nomenclature is to state the name and cite the work (e.g. smith et al. [4]). Saying 'Ref.' makes the reader search for the reference to understand it.

Reply 1.4: Thank you, they have been modified.

1.5 Page 3, Line 8. Turning to real-world driving sensor fusion system ... Consider saying 'Turning to a real-world ...' or 'sensor fusion systems'

Reply 1.5: Thank you, fixed as "sensor fusion systems".

1.6 Page 3 Line 55. Surely this is only for point object tracking? The challenge of extended object tracking is that you are trying to track a target which is not fully observed and therefore there is no known measurement-to-object association ?

Reply 1.6: The motion state of a target is represented by a point, i.e., the centroid of the bounding box. Clustering is assumed to be done prior to centroid estimation. Please also see Reply 1.7. Thus one can talk about measurement-to-target association.

1.7 Page 4, Line 13. There is an additional consideration here in that you have two challenges.

The first is the measurement-to-Estimated Object association (i.e. where on the vehicle is each lidar point measuring) and the estimatedObject-to-ObjectTrack association. While I am not sure it is necessary to discuss this in this paper, it may be a useful consideration for future work!

Reply 1.7: Thank you. The first challenge described by the reviewer is known as clustering (gather data points that belong to one single object). Note that the proposed centroid estimation is based on clustered data. Clustering techniques can be found in the literature and are not included in this paper for simplicity. The second challenge is measurement-to-track association which uses centroid (or corner point) as reference point. The focus of this paper is to find a simple but accurate solution of estimating the centroid and its uncertainty in order to achieve reliable measurement-to-track association results.

1.8 Section 2 Equation 26. It would be useful to have a short summary to this section outlining what the derivation achieved. Specifically, explaining what the parameters $\hat{a}$ and $\hat{b}$ represent would be useful here and will greatly aid in readability of the paper when you are considering the bounding box based estimation in later sections. When reviewing the paper, I needed to go back and forth in order to follow the derivation and its relation to the triangular and uniform pdf functions.

Reply 1.8: Section III has been modified with detailed equations (30)–(45) added to improve the readability for non-experts in the field as suggested by the reviewer.

1.9 Page 9, Line 18. The paper states that 'assuming independence between x(1) and x(ni)' - Is this strictly true? This might be a more philosophical discussion, but surely the clustered lidar points are not truly independent nor identical? The measurements themselves are recorded intervals against the same physical object, and the uncertainty covariance for each lidar beam will be a function of angle/range and not x/y. I think that it would be useful to comment on this!

Reply 1.9: Conditional independence between x(1) and x(n) is a reasonable assumption here (it is conditioned on the distribution parameter) for centroid estimation.

1.10 How is ground truth computed in the results? The paper states that it is estimated 'manually offline' (page 11 line 53) - what does this mean exactly? Surely the most accurate sensor is itself the lidar scanner? This is important as the results are showing errors which are below the precision of the sensor based on the theoretic range-resolution to bandwidth consideration (I believe that this

is 4cm for a velodyne but I may be wrong).

Reply 1.10: Ground truth, i.e., centroid of the vehicle in local coordinates are measurements obtained using an extra laser range measuring device. The standard deviation (uncertainty level), matches sensor resolution. As shown in the newly added simulation section, the estimator is consistent, i.e, RMSE matches calculated SD statistically.

1.11 Consider placing a short summary of the derivations at the end of sections 3 and 4. It would be useful for readability by non-experts in the field.

Reply 1.11: Please see Reply. 1.8.

1.12 My final recommendation would be to expand the conclusions section. At present, the conclusions provide a very good summary of the work. However, a little expansion on the importance and impact of the work (so what) will help the reader to understand the importance of the work in this domain.

Reply 1.12: Thank you, we have modified the conclusion. See lines 218–228.

Reviewer 2.

2.1 No references are provided for the fact that max-min averaging is the standard procedure. In fact, a quick literature study (see below) showed that most methods directly take the center of the bounding box as measurement. This should be discussed by the authors and using the center of the bounding box as a measurement should also be considered in the numerical results.

Reply 2.1: Max-min averaging is an empirical simplest method for getting the centroid. To clarify, the center of the bounding box is the centroid that is analyzed in our numerical results. The proposed technique uses all the clustered measurements to estimate the support boundaries of the pdf and is significantly better than Max-min as shown in Fig.4.

2.2 The numerical results need to be discussed in more detail and expanded. No reference is provided for the used dataset and the machine learning method used for bounding box extraction. This is somewhat surprising since there is a very large number of datasets and extracted bounding boxes available online, e.g., provided by Waymo and nuScenes. I strongly recommend using publicly

available data and providing appropriate references for reproducibility.

Reply 2.2: Thank you for the suggestion. Clustering/detection technique used in the experiment is from LaserNET, a reference as been added. Unfortunately, more bounding box data analysis is not available in the current project. In addition, the current work is based on clustered data and clustering technique is beyond the scope of this paper. We will incorporate more data testing in future study.

2.3

T. Yin, X. Zhou, and P. Krahenbuhl, "Center-based 3D object detectionand tracking," in Proc. CVPR-21, Jun. 2021, pp. 11 784–11 793.

H.-k. Chiu, A. Prioletti, J. Li, and J. Bohg, "Probabilistic 3D multi-object tracking for autonomous driving,"arXiv preprint arXiv:2001.05673,2020.

Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, Chiew-Lan Tai; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 1090-1099

Reply 2.3: Thank you for the suggestions. Suggested references added in Introduction, see lines 45–52.