Uncertainty Estimation of the Bounding Box Centroid for Autonomous Driving

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

ABSTRACT

Object detection provides information needed for target tracking and plays a core role in autonomous driving. In this work, we study the uncertainty in the estimation of the centroid (position) of a bounding box of the measurements from an object detected by the sensor of an autonomous vehicle (AV). The estimated centroid uncertainty will be used in object tracking as measurement noise variance, which is not available from the sensor manufacturer, for measurement association and target state estimation. When the (position) uncertainty that captures the noise inherent in the sensor observations is available for each detected point (this can be done using Bayesian deep learning), the bounding box centroid uncertainty is obtained using a Least-Squares estimator (LS). When the uncertainty for each detected point is not available, one can assume a uniform distribution of the clustered points in a single rectangular bounding box. A Maximum Likelihood estimator is used for the bounding box centroid estimation. Experiments using real data are carried out to show the performance of proposed methods for autonomous driving applications. A comparison with the sample mean approach showed the superiority of new 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)* are crucial in achieving high motion state estimation accuracy and consistency†. Thus, capturing reliable 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. Thus, object detection algorithms using LiDAR observations have been widely studied. In this work, we study the uncertainty in the estimation of the centroid (position) of a bounding box detected by a LiDAR. The centroid uncertainty is necessary for both point object tracking and extended object tracking [4].

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 bounding box centroid uncertainty, which is needed for object tracking as position measurement uncertainty. Ref. [6] introduced the concept of epistemic uncertainty (caused by process noise) and Aleatoric uncertainty (caused by measurement noise) and proposed a Bayesian Deep Learning algorithm for obtaining the aforementioned uncertainties. Ref. [7] proposed an improved LiDAR-based object detector by combing the benefits of the Kullback-Leibler Divergence loss and LiDAR label uncertainty. In [10], the authors proposed 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 [6], Ref. [8] 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. Ref. [9] proposed a novel form of the loss function (via neural networks) to increase the performance of LiDAR-based object detection and discussed the individual measurement uncertainty

Proceedings of SPIE Conf. on Autonomous Systems: Sensors, Processing and Security for Ground, Air, Sea and Space Vehicles and Infrastructure, #12115-4 , April, 2022, Orlando FL.

*The measurement noises are commonly assumed to be white zero-mean Gaussian.

†Consistency means the filter calculated uncertainties (standard deviation) associated with the state estimates match statistically the actually errors.

for prediction. Note that the measurement uncertainty obtained using neural networks in LiDAR detectors is commonly used for bounding box corners. Limited research has focused on the centroid estimation problem of a bounding box and its uncertainty.

In this work, we propose a simple solution for LiDAR bounding box centroid estimation based on detected and clustered LiDAR point clouds. We assume that object detection is a pre-process 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. When the (position) uncertainty that captures the noise inherent in the sensor observations is available for each detected point (this can be done using Bayesian deep learning, i.e, using LaserNett from [8]), the bounding box centroid uncertainty is obtained using a Least-Squares estimator (LS). When the uncertainty for each detected point is not available (a more common situation), one can assume a uniform distribution of the clustered points in a single rectangular bounding box. An unbiased estimator based on the Maximum Likelihood (ML) approach is used for the bounding box (Likelihood Function support boundary) parameters estimation in each coordinate.

Sec. II formulates the problem and introduces a Least Squares Estimator assuming the knowledge of (position) uncertainties from neural network prediction. The detailed discussion for uniform pdf based Maximum Likelihood bounding box centroid estimation (unbiased) is in Sec. III. The scenario setup and numerical results can be found in Sec. IV. Conclusions are given in Sec. V.

II. LEAST SQUARES ESTIMATOR

Assuming each LiDAR point k is a vector that contains 3D position $[x(k)\ y(k)\ z(k)]$, azimuth angle $\theta(k)$ and elevation angle $\eta(k)$. There are N LiDAR points received in the current frame[‡], i.e., $k = 1, \ldots, N$. For each point and class of object, the neural network [8] predicts (assuming it is available) a set of bounding box parameters including $[d_x(k), d_y(k), d_z(k)]$ the relative center, $[\omega_x(k), \omega_y(k), \omega_z(k)]$ the relative orientation, $[l(k), w(k), h(k)]$ are the length, width and height of the 3D rectangular bounding box and the corresponding standard deviations $[\sigma_x(k), \sigma_y(k), \sigma_z(k)]$. Each LiDAR point generates a potential bounding box [8]. The bounding box generated from LiDAR point k is $B_k, k = 1, \ldots, N$.

The centroid of the bonding box B_k can be obtained by

$$\mathbf{c}(k) = \begin{bmatrix} c_x(k) \\ c_y(k) \\ c_z(k) \end{bmatrix} = \begin{bmatrix} x(k) \\ y(k) \\ z(k) \end{bmatrix} + J_{\theta,\eta}(k) \begin{bmatrix} d_x(k) \\ d_y(k) \\ d_z(k) \end{bmatrix} \tag{1}$$

where $J_{\theta,\eta}(k)$ ([8]) is the rotation matrix parameterized by $\theta(k)$ and $\eta(k)$.

In object clustering (by discretizing the top-down view into bins [8]), a set of LiDAR points S_i whose centroids, i.e., absolute box centers fall into the same bin (cluster) i form a single bounding box for object i. Define the bin indicator

$$\alpha^i(k) = \begin{cases} 1 & k \in S_i \ (\mathbf{c}(k) \text{ falls in bin } i) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The total number of points in the set S_i is

$$n^i = \sum_{k=1}^{N} \alpha^i(k) \tag{3}$$

and the (clustered) bounding box centroid and its uncertainty can be obtained via Least Squares Technique described in the following.

For all the LiDAR points in S_i, rewrite (1) as

$$\begin{bmatrix} d_x(k) \\ d_y(k) \\ d_z(k) \end{bmatrix} = J_{\theta,\eta}(k)^{-1} \left[\begin{bmatrix} c_x^i \\ c_y^i \\ c_z^i \end{bmatrix} - \begin{bmatrix} x(k) \\ y(k) \\ z(k) \end{bmatrix} \right] \qquad k \in S_i \tag{4}$$

[‡]The time argument is omitted for simplicity.

where

$$\mathbf{c}^i = \begin{bmatrix} c_x^i \\ c_y^i \\ c_z^i \end{bmatrix} \tag{5}$$

is the centroid of the clustered bounding box (for the bin i) that needs to be estimated, i.e., it is the parameter vector of interest. Given the estimates and standard deviations of the relative center, we shall construct a pesdo-measurement (to account for the uncertainty caused by neural network prediction) for the parameter vector

$$\begin{bmatrix} \hat{d}_x(k) \\ \hat{d}_y(k) \\ \hat{d}_z(k) \end{bmatrix} = J_{\theta,\eta}(k)^{-1} \left[\begin{bmatrix} c_x^i \\ c_y^i \\ c_z^i \end{bmatrix} - \begin{bmatrix} x(k) \\ y(k) \\ z(k) \end{bmatrix} \right] + \begin{bmatrix} w_x(k) \\ w_y(k) \\ w_z(k) \end{bmatrix} \tag{6}$$

where $w_x(k)$, $w_y(k)$ and $w_z(k)$ are the pesdo-measurement noises assumed to be white zero-mean Gaussian with covariance matrix

$$R(k) = \begin{bmatrix} \sigma_x^2(k) & 0 & 0 \\ 0 & \sigma_y^2(k) & 0 \\ 0 & 0 & \sigma_z^2(k) \end{bmatrix} \tag{7}$$

Note that $[\hat{d}_x(k), \hat{d}_y(k), \hat{d}_z(k)]$ and $R(k)$ are obtained via deep learning [6] .

Define the observation vector, by moving the known values (obtained in neural network prediction) on the r.h.s. of (6) to the left,

$$\mathbf{z}(k) = \begin{bmatrix} \hat{d}_x(k) \\ \hat{d}_y(k) \\ \hat{d}_z(k) \end{bmatrix} + J_{\theta,\eta}(k)^{-1} \begin{bmatrix} x(k) \\ y(k) \\ z(k) \end{bmatrix} = J_{\theta,\eta}(k)^{-1} \begin{bmatrix} c_x^i \\ c_y^i \\ c_z^i \end{bmatrix} + \begin{bmatrix} w_x(k) \\ w_y(k) \\ w_z(k) \end{bmatrix} = J_{\theta,\eta}(k)^{-1}\mathbf{c}^i + \mathbf{w}^i \tag{8}$$

The stacked observation vector (column vector) for set S_i is

$$\mathbf{Z}^i = \text{col}\left[\mathbf{z}(k)\right]_{k=\ell_1^i \cdots \ell_{n^i}^i} \tag{9}$$

where $\ell_j^i, j = 1, \cdots, n^i$ are the indices of the measurements (from N measurements in the frame considered) that belong to S_i according to (2).

Under the LS criterion [1] one has the LS estimate of the center of the bounding box (for bin i)

$$\hat{\mathbf{c}}^i = \left[(H^i)'(R^i)^{-1}H^i\right]^{-1}(H^i)'(R^i)^{-1}\mathbf{Z}^i \tag{10}$$

where

$$H^i = \begin{bmatrix} J_{\theta,\eta}(\ell_1^i)^{-1} \\ \vdots \\ J_{\theta,\eta}(\ell_{n^i}^i)^{-1} \end{bmatrix} \tag{11}$$

and

$$R^i = \begin{bmatrix} R(\ell_1^i) & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & R(\ell_{n^i}^i) \end{bmatrix} \tag{12}$$

The corresponding covariance matrix is

$$P^i = \left[(H^i)'(R^i)^{-1}H^i\right]^{-1} \tag{13}$$

which should be used for target motion tracking and sensor data fusion (along with (10) in filtering) as the measurement (position) noise uncertainty.

III. UNIFORM PDF BASED CENTROID ESTIMATION

In many cases where the uncertainty of the relative center is not available after pre-processing for the point cloud, we can estimate the centroid of the (clustered measurements' bounding box) by assuming that the position observations are obtained from a uniform distribution with certain boundaries (to be estimated).

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

$$\mathbf{O}^i = \mathrm{col}\big[[x(k)\ y(k)\ z(k)]'\big]_{k\in S_i} \tag{14}$$

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

Note that the estimation for the centroid is done independently across coordinates. Thus, as an illustration, the following analysis is only for x-axis, where the terminology can be used for any axis. A subset of (14), only considering the x-axis, is defined as

$$\mathbf{O}_x^i \triangleq \mathrm{col}\,[x(k)]_{k\in S_i} \tag{15}$$

which contains n^i, given in (3), position measurements.

Assume the observation set (15) is a set of independent identical distributed (i.i.d.) random variables drawn from a uniform distribution $\mathcal{U}(x_L^i, x_U^i)$ and

$$\mathbf{x}^i = [x_L^i\ x_U^i]' \tag{16}$$

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)

$$\Lambda(\mathbf{x}^i;\mathbf{O}^i) = \prod_{j=1}^{n^i} p(x(\ell_j^i)|x_L^i, x_U^i) = \frac{1}{(x_U^i - x_L^i)^{n^i}} \tag{17}$$

The unbiased ML estimator for (16) is [11]

$$\hat{x}_L^i = \frac{n^i o_m^i - o_M^i}{n^i - 1}, \qquad \hat{x}_U^i = \frac{n^i o_M^i - o_m^i}{n^i - 1} \tag{18}$$

where

$$o_m^i \triangleq \min\{o_x^i(\ell_1^i), \ldots, o_x^i(\ell_{n^i}^i)\} \tag{19}$$

$$o_M^i \triangleq \max\{o_x^i(\ell_1^i), \ldots, o_x^i(\ell_{n^i}^i)\} \tag{20}$$

The corresponding covariance matrix is

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

where

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

and

$$P_{12}^i = P_{21}^i = \frac{-(x_U^i - x_L^i)^2}{(T^i - 1)(T^i + 1)(T^i + 2)} \tag{23}$$

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

$$\hat{c}_x^i = \frac{1}{2}(\hat{x}_L^i + \hat{x}_U^i) \tag{24}$$

with corresponding variance

$$\begin{aligned}\sigma^2_{\hat{c}^i_x} =& E\left[(\hat{c}^i_{x,U} - \frac{1}{2}(x^i_l + x^i_u))^2\right] \\ =& \frac{1}{4}P^i_{11} + \frac{1}{2}P^i_{12} + \frac{1}{4}P^i_{22} \\ =& \frac{(x^i_U - x^i_L)^2}{2(n^i+1)(n^i+2)}\end{aligned} \tag{25}$$

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.

IV. NUMERICAL RESULTS

In the real-world verification setup that was carried out, a vehicle equipped with a Velodyne VLS-128 mounted on the top of the vehicle roof, shown in Figure 1, 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 that was trained to recognize vehicles and the algorithm calculated a 3D bounding box for each detected vehicle, shown in Figure 2. The dimensions and position of the bounding boxes were saved off along with the raw point cloud for offline analysis.

Figure 1: Lexus RX450hL Equipped with a Velodyne VLS-128

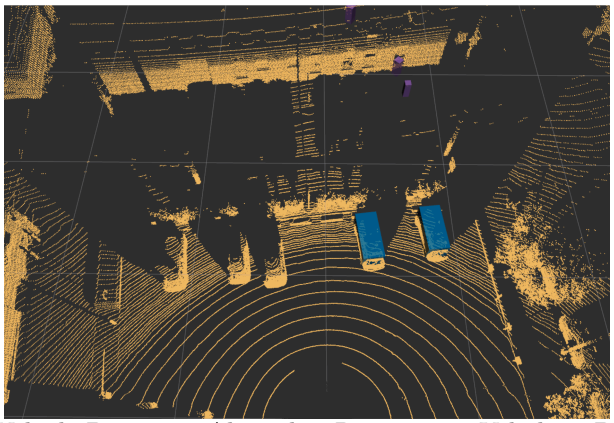
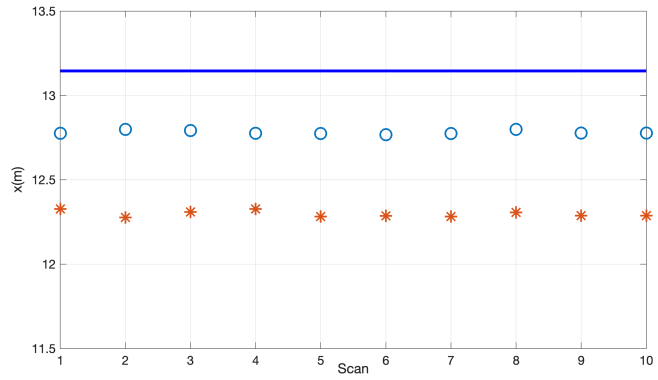
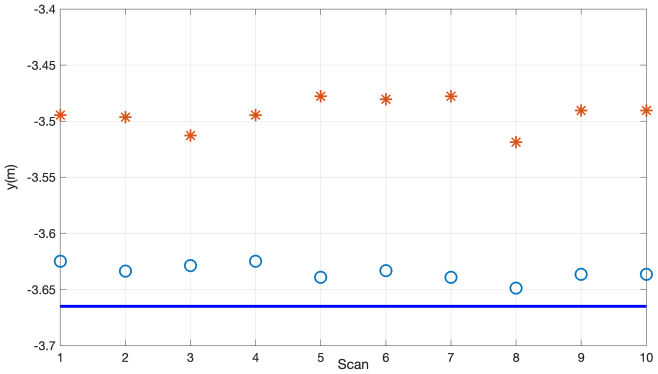


Figure 2: Vehicle Detection Algorithm Running on Velodyne Point Cloud



(a) Centroid Estimation for x



(a) Centroid Estimation for y

Figure 3: Centroid Estimation Error
Ground Truth $*$ Sample Mean ◯ Unbiased ML Estimate

The estimation is carried out at 10 Hz and the numerical results are shown in Fig. 3 for x and y. The proposed unbiased ML estimation is compared with the naive but commonly used approach in AV applications: the sample mean (on the x-axis for illustration)

$$\hat{c}^i_{x,S} = \frac{1}{2}(o^i_m + o^i_M) \tag{26}$$

Note that the above cannot provide uncertainty for the bounding box centroid using the sample data. The ground truth is measured manually off-line. It can be seen that the approach of Uniform PDF based ML unbiased estimation (18) has superior performance with smaller error.

The estimate uncertainty (25) is shown in Table 1 for the unbiased ML estimation assuming i.i.d. observations (16). Note that the LS estimate and its uncertainty are not available in the current examination due to insufficient data. The estimation errors in Fig. 3 are potentially caused by: 1) biased sensor measurements and; 2) model mismatch. The performance testing for a moving object will be investigated for LS estimation (with uncertainty obtained from machine learning) and Uniform PDF based ML estimation with noisy measurements in a future study.

V. CONCLUSIONS

In this work, we proposed a simple solution for LiDAR-based target bounding box centroid estimation in autonomous driving. When the (position) uncertainty that captures the noise inherent in the sensor observations

Table 1: Bounding Box Centroid Estimate Standard Deviation

Scan	1	2	3	4	5	6	7	8	9	10
x (cm)	0.30	0.30	0.31	0.39	0.31	0.30	0.37	0.30	0.30	0.38
y (cm)	0.20	0.25	0.20	0.26	0.20	0.20	0.25	0.21	0.20	0.25

is available for each detected point (this can be done using Bayesian deep learning), the bounding box centroid uncertainty can be obtained using a Least-Squares estimator. When the uncertainty for each detected point is not available, one can assume a uniform distribution of the clustered points on a single rectangle bounding box. An unbiased estimator is used for the likelihood function support boundary parameters' estimation for each axis and the centroid coordinates are given by the centers of the boundaries. Numerical results show that the proposed ML estimator can successfully estimate the bounding box centroid, and the uncertainty is also available. Moving object centroid estimation will be studied in future research.

REFERENCES

1. Y. Bar-Shalom, X. R. Li and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*, Wiley, 2001.
2. Y. Bar-Shalom, P. K. Willett and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, 2011.
3. 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.
4. K. Granström, C. Lundquist, and U. Orguner, "Tracking rectangular and elliptical extended targets using laser measurements," *Proc. IEEE Int. Conf. on Information Fusion*. pp. 1–8. July 2011.
5. 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.
6. A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" *arXiv preprint arXiv:1703.04977*. Mar. 2017.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.