Extended Object Tracking with Automotive Radar Using Learned Structural Measurement Model

Yuxuan Xia[1,2], Pu Wang[1], Karl Berntorp[1], Petros Boufounos[1], Philip Orlik[1], Lennart Svensson[2] and Karl Granström[2]
[1]Mitsubishi Electric Research Laboratories (MERL), Cambridge, Massachusetts, USA
[2]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden
Emails: {pwang, berntorp, petrosb, porlik}@merl.com, {firstname.lastname}@chalmers.se

***Abstract*—This paper presents a data-driven measurement model for extended object tracking (EOT) with automotive radar. Specifically, the spatial distribution of automotive radar measurements is modeled as a hierarchical truncated Gaussian with structural geometry parameters (e.g., truncation bounds, their orientation, and a scaling factor) learned from the training data. The contribution is twofold. First, the learned measurement model can provide an adequate resemblance to the spatial distribution of real-world automotive radar measurements. Second, large-scale offline training datasets can be leveraged to learn the geometry-related parameters and offload the computationally demanding model parameter estimation from the state update step. The learned structural measurement model is further incorporated into the random matrix-based EOT approach with a new state update step. The effectiveness of the proposed approach is verified on the nuScenes dataset.**

***Index Terms*—Automotive radar, extended object tracking, random matrix, autonomous driving, nuScenes.**

I. Introduction

Automotive radar plays an important role in autonomous driving, as it provides reliable environmental perception in all-weather conditions with affordable cost. With the advances in radar technology, modern radar sensors can resolve multiple detection points per object. The tracking of an object that generated multiple detections is called extended object tracking (EOT). Compared to conventional point object tracking, EOT can lead to improved tracking capability as multiple detections allow us to not only estimate the object kinematic state more precisely but also to infer the object extent state. An elaborate overview of EOT literature can be found in [1].

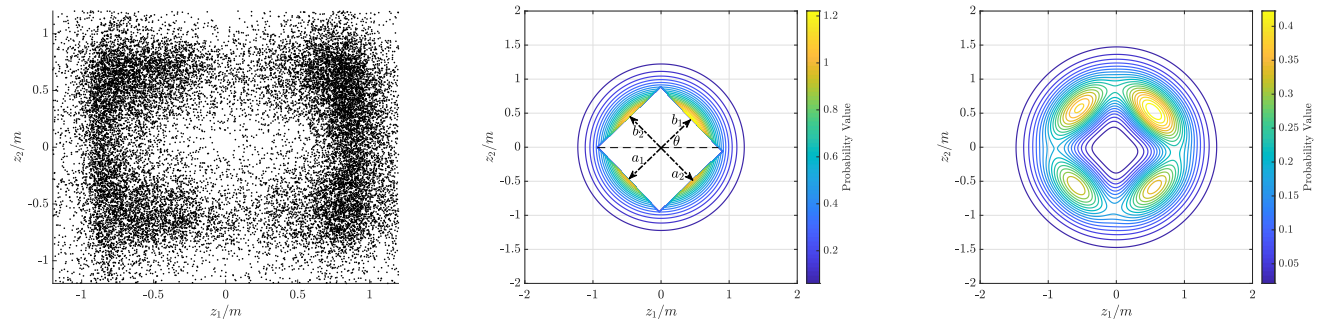
A key component for EOT is the measurement model, which needs to capture the spatial characteristics of radar measurements. One modeling approach is to assume that the extended object has a fixed number of reflection points located on a rigid body shape [2]–[4]. A challenge with this approach is that it requires data association between the fixed set of points and radar measurements. It is also possible to derive the measurement model using a physics-based approach [2], [5], which usually requires some expert knowledge and manual adaption. The most common approach assumes that these radar measurements are spatially distributed as a function of individual measurement likelihoods, also referred to as the spatial distribution.

This work was done during Yuxuan Xia's internship at MERL.

The spatial distribution of automotive radar measurements can be generally divided into three categories: 1) contour models, which reflect the measurement distribution along the object contour; 2) surface models, which assume that the measurements are generated from the inner surface of objects; and 3) surface-volume models, which balance between the above two models with more realistic features. For the contour model, typical examples include rectangular shape models [6], [7], the hypersurface model [8]–[10] and the C B-Spline model [11]. A widely used surface model is the random matrix (RM) approach [12]–[16], which assumes an elliptic object shape. The surface model can lead to computationally simpler algorithms than the contour model, which uses more flexibility to describe more complex shapes.

The spatial characteristics of real-world automotive radar measurements are, however, more complex and cannot be well described by either the contour model or the surface model, see Fig. 1 (a) for an illustration. Two observations can be made from Fig. 1 (a); 1) the measurement density is much lower at the center than in a vicinity around outer edges; and 2) measurements exhibit self-occlusion features: the measurement density is dominant at object parts that are in sight of the automotive radar sensor. These observations have motivated recent developments of surface-volume models, including the variational Gaussian mixture model [18], [19], the volcanormal measurement model [20], and the hierarchical truncated Gaussian (HTG) measurement model [21], [22].

The HTG measurement model can adequately approximate the spatial distribution of real-world automotive radar measurements on vehicles. It has been integrated into the RM-based approach with a modified state update step and online truncation bound estimation in [21] for full-view measurements and in [22] for partial-view measurements due to object self-occlusion. In this paper, we take one step further to offload the bound estimation in the state update step by leveraging large-scale offline training datasets. Particularly, we propose to learn the geometry-related structural model parameters, e.g., the (relative) truncation bounds, their orientation and a scaling factor, of the HTG measurement model and leave the state-related parameters such as the kinematic mean and extent covariance matrix in the measurement likelihood function. Then, the learned structural measurement model is utilized in the EOT algorithm proposed in [21] and [22]. This is in contrast to the full data-driven approach [18], [19] where



(a) (b) (c)

Fig. 1: (a) scatter plot of accumulated automotive radar measurements of vehicles in a unit coordinate system from the nuScenes dataset [17]; (b) the truncated Gaussian $p(y|\xi)$ for the measurement source y with parameters $h(x) = [0, 0]^T$, $X = \text{diag}([1, 1])$, $\rho = 0.183$, $\theta = 0.763$, $B = [0.675, 0.670, 0.611, 0.648]$, obtained via offline training of the measurements in (a); (c) the resulting spatial distribution $p(z|\xi)$ with noise covariance $R_{\text{UC}} = [0.036, 0.001; 0.001, 0.036]$.

the full object extent is learned during the offline training phase. Our structural learning approach can be more robust to measurement model mismatch between the training data and new measurements. The resulting EOT algorithm is validated using the real-world nuScenes dataset [17].

II. Problem Formulation

In this paper, we focus on single EOT. We define the object state as a tuple $\xi_k = (x_k, X_k)$ with x_k denoting the kinematic state and X_k, a symmetric and positive definite matrix, denoting the extent state. The kinematic state $x_k = [p_{k,x}, p_{k,y}, \phi_k, v_k, \omega_k]^{\text{T}}$ comprises the position of the object center given by $p_{k,x}$ and $p_{k,y}$, the heading ϕ_k, the velocity v_k, and the yaw rate ω_k. For object with rectangular shape (e.g., vehicles), its length l and width w can be obtained from the eigen-decomposition of X_k.

We receive n_k measurements $Z_k = \{z_k^j\}_{j=1}^{n_k}$ at each time step k. The objective of object tracking is to recursively compute the posterior density of the object state $p(\xi_k|Z_{1:k})$ given all measurements $Z_{1:k} = \{Z_1, \ldots, Z_k\}$ up to and including the current time step k using Bayesian estimation. The object state ξ_k with corresponding uncertainty measures can then be extracted from the posterior density $p(\xi_k|Z_{1:k})$.

Given the posterior density $p(\xi_{k-1}|Z_{1:k-1})$ at time step $k-1$ and the transition density $p(\xi_k|\xi_{k-1})$, the predicted density is given by the prediction step

$$p(\xi_k|Z_{1:k-1}) = \int p(\xi_{k-1}|Z_{1:k-1})p(\xi_k|\xi_{k-1})d\xi_{k-1}. \quad (1)$$

This density is then updated using measurements Z_k via the Bayes' rule,

$$p(\xi_k|Z_{1:k}) \propto p(\xi_k|Z_{1:k-1})p(Z_k|\xi_k), \quad (2)$$

where $p(Z_k|\xi_k) = \prod_{j=1}^{n_k} p(z_k^j|\xi_k)$ is the joint measurement likelihood with $p(z_k^j|\xi_k)$ denoting the *spatial distribution*. We approximate the predicted and posterior state densities such that they are all of the same functional form, which allows a recursive use of the prediction and update functions.

III. Hierarchical Truncation Model

In this section, we present the HTG spatial distribution for modeling noisy automotive radar measurements and show how the distribution can be learned from the offline training data.

A. Hierarchical Truncated Measurement Model

Each detection point z is modeled as a noisy measurement of a noise-free measurement source y. The distribution $p(z|y)$ that models the sensor noise is Gaussian $\mathcal{N}(z; y, R)$ where R is the measurement noise covariance matrix. The distribution $p(y|\xi)$ that models the spatial characteristics of the measurement source is a truncated Gaussian with the form

$$\mathcal{TN}(y; h(x), \rho X, D) = \frac{\mathbf{1}_D(y)}{c_D}\mathcal{N}(y; h(x), \rho X), \quad (3)$$

where ρ is a scaling factor, D is the density support, $\mathbf{1}_D(\cdot)$ is the indicator function on D, and c_D is the normalization factor such that (3) integrates to one. Note that we use generic D and c_D to symbolize the truncated area and the normalization constant in the equations for the sake of simplicity, although D and c_D might be changing from equation to equation.

The truncated area can be fully specified by the object center $h(x)$, the truncation bounds $B \triangleq [a_1, a_2, b_1, b_2]$, and an orientation θ with respect to the horizontal axis. One or more truncation bounds can be set to infinity to model partial-view measurements caused by the self-occlusion [22]. Also, note that in contrast to the HTG model introduced in [7], the orientation of the truncated area considered in this work does not need to be aligned with the object orientation. This leads to a better modeling of the feature that the vehicle wheels and wheel houses are typical measurement sources. An illustrative example of $p(y|\xi)$ is given in Fig. 1 (b). The resulting measurement spatial distribution can be computed by marginalizing out the measurement source,

$$\begin{aligned} p(z|\xi) &= \int p(z|y)p(y|\xi)dy \\ &= \frac{1}{c_D}\int_D \mathcal{N}(z; y, R)\mathcal{N}(y; h(x), \rho X)dy, \end{aligned} \quad (4)$$

For the given truncated Gaussian $p(y|\xi)$ in Fig. 1 (b), the resulting spatial distribution $p(z|\xi)$ is shown in Fig. 1 (c) with a given noise covariance matrix R.

B. Learning Hierarchical Truncated Measurement Model

We assume that the object states are available as ground truth, and that the training data to learn the model consists of a set of N two-dimensional accumulated data points in a global

coordinate system, $Z_{\mathrm{GC}} = \{z^j_{\mathrm{GC}}\}_{j=1}^N$, generated by a HTG with given parameter values. The parameters of the model are unknown and the objective is therefore to estimate the parameters given the available data. The maximum likelihood (ML) estimate can be obtained by maximizing the joint measurement likelihood with respect to the parameters that the training data Z_{GC} condition on.

1) Dimension Reduction: To obtain a hierarchical truncation model that well describes the spatial characteristics of automotive radar detections from vehicles, we need to collect a large amount of data samples to account for vehicles of different size and with different poses. To avoid collecting data in such a complex space, we apply the dimension reduction technique proposed in [18]. Specifically, the radar detections in the global coordinate system are first transformed to the object coordinate system to obtain $Z_{\mathrm{OC}} = \{z^j_{\mathrm{OC}}\}_{j=1}^N$, and then the measurements in the normalized object coordinate system $Z_{\mathrm{UC}} = \{z^j_{\mathrm{UC}}\}_{j=1}^N$ are given by

$$z_{\mathrm{UC}} = \mathrm{diag}\left([2/l, 2/w]\right) z_{\mathrm{OC}}, \tag{5}$$

see Fig. 1 (a) for an illustration. After the dimension reduction, the positions of all vehicle detections are transformed to a normalized object coordinate system that are independent of the object states, e.g., the position, length and width.

2) Offline ML Estimation of Model Parameters: The ML estimate for the coordinate-transformed training data is given by

$$\underset{\rho,B,\theta,R_{\mathrm{UC}}}{\arg\max} \frac{1}{c_D^N} \prod_{z_{\mathrm{UC}}\in Z_{\mathrm{UC}}} \int_D \mathcal{N}(z_{\mathrm{UC}}; y, R_{\mathrm{UC}})\mathcal{N}(y; 0_{2\times 1}, \rho \mathrm{I}_2) dy, \tag{6}$$

where I_2 is an identity matrix. Computing (6) involves evaluating the convolution of a bivariate truncated Gaussian and a bivariate Gaussian , which is intractable in general. To obtain a tractable solution of (6), we assume that the noise covariance R_{UC} in the normalized object coordinate has the form

$$R_{\mathrm{UC}} = M(\theta)\mathrm{diag}([r_1, r_2])M(\theta)^{\mathrm{T}}, \tag{7}$$

where $M(\theta)$ is the counterclockwise rotation matrix. Obtaining the ML estimate of (6) is then equivalent to obtaining

$$\begin{aligned}\underset{\rho,B,\theta,r_1,r_2}{\arg\min} \; &- \sum_{z_{\mathrm{UC}}\in Z_{\mathrm{UC}}} \log \int_D \mathcal{N}(y; 0_{2\times 1}, \rho \mathrm{I}_2) \times \\ &\mathcal{N}\left(M(-\theta) z_{\mathrm{UC}}; y, \mathrm{diag}([r_1, r_2])\right) dy + N \log c_D, \end{aligned} \tag{8}$$

where we further transform the measurements Z_{UC} to a new coordinate system such that the orientation of the truncated area is now aligned with the axis and that the transformed noise covariance is a diagonal matrix. With the new parameterization, the above cost function of the ML estimation can be computed as a product of two univariate Gaussians and, hence, decoupled. Specifically, we can rewrite (8) by applying the convolution formula as

$$\begin{aligned}\underset{\rho,B,\theta,r_1,r_2}{\arg\min} \; &- \sum_{[z_x,z_y]^{\mathrm{T}}\in Z_{\mathrm{UC}}} \log \Big(\mathcal{N}(\tilde{z}_x; 0, r_1 + \rho) \\ &\times \mathcal{N}(\tilde{z}_y; 0, r_2 + \rho) - \lambda(\tilde{z}_x, a_1, b_1, r_1) \\ &\times \lambda(\tilde{z}_y, a_2, b_2, r_2)(1 - c_D) \Big) + N \log c_D, \end{aligned} \tag{9}$$

$$\text{s.t. } \rho \in (0, 1], \theta \in [-\pi/2, \pi/2], \{B, r_1, r_2\} \in [0, \infty),$$

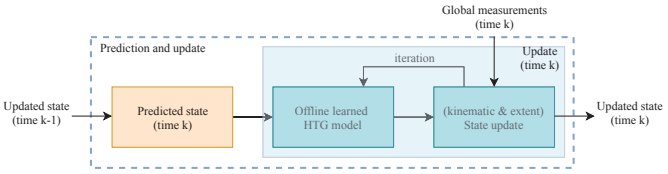


Fig. 2: Diagram of proposed random matrix approach for learned HTG model.

where

$$\begin{aligned}[\tilde{z}_x, \tilde{z}_y] &= [\cos(\theta)z'_x + \sin(\theta)z'_y, -\sin(\theta)z'_x + \cos(\theta)z'_y], \\ c_D &= 1 - \big((\Phi(b_1\rho^{-1/2}) - \Phi(-a_1\rho^{-1/2})) \\ &\quad \times (\Phi(b_2\rho^{-1/2}) - \Phi(-a_2\rho^{-1/2}))\big), \\ \lambda(\tilde{z}, a, b, r) &= \gamma(a, b, r)\Bigg(\Phi\left(\frac{\tilde{z} + a - \varepsilon(\tilde{z}, r)}{\beta(r)}\right) \\ &\quad - \Phi\left(\frac{\tilde{z} - b - \varepsilon(\tilde{z}, r)}{\beta(r)}\right)\Bigg) \exp\left(-\frac{\tilde{z}\tilde{z}}{2(r + \rho)}\right), \\ \varepsilon(\tilde{z}, r) &= r\tilde{z}(r + \rho)^{-1}, \;\; \beta(r) = r^{1/2}\rho^{1/2}(r + \rho)^{-1}, \\ \gamma(a, b, r) &= \frac{\sqrt{2\pi}\beta(r)}{2\pi r^{1/2}\rho^{1/2}\left(\Phi(-b\rho^{-1/2}) - \Phi(a\rho^{-1/2})\right)}, \end{aligned}$$

with $\Phi(\cdot)$ denoting the cumulative density function (CDF) of a standard normal distribution. The ML estimate of all unknown parameters can be found by using the coordinate descent optimization which iteratively determines a coordinate (one of unknown parameters here) and minimizes (9) over the selected coordinate with proper constraints (e.g., the bounds B are non-negative) while fixing all other coordinates.

3) Measurement Model Dependence on Aspect Angles: Automotive radar measurements are likely reflected from parts of the vehicle that are in sight of the radar sensor. To account for this self-occlusion feature, we can split the training data set into groups according to the annotated aspect angle in the training dataset [18]

$$\phi' = \phi_{\mathrm{SC}} - \mathrm{atan2}(p_{y,\mathrm{SC}}, p_{x,\mathrm{SC}}), \tag{11}$$

where ϕ_{SC} and $[p_{x,\mathrm{SC}}, p_{y,\mathrm{SC}}]^{\mathrm{T}}$, respectively, denote the object orientation and center position in the sensor coordinate system. Then the conditional measurement models on the aspect angle can be learning using the above ML estimation; see Fig. 3 for an illustration.

IV. Random Matrix Approach for Learned Hierarchical Truncation Model

In this section, we introduce the modified RM approach with the learned hierarchical truncation model in Fig. 2. In particular, we present a new RM state update step. It is assumed that both the predicted and posterior densities have the factorized form [13]

$$\begin{aligned} p(\xi_k|Z_{1:k'}) &\approx p(x_k|Z_{1:k'})p(X_k|Z_{1:k'}) \\ &= \mathcal{N}(x_k; m_{k|k'}, P_{k|k'})\mathcal{IW}(X_k; \nu_{k|k'}, V_{k|k'}), \end{aligned} \tag{12}$$

where $k' \in \{k-1, k\}$. The kinematic state x_k is Gaussian distributed with mean $m_{k|k'}$ and covariance matrix $P_{k|k'}$, whereas the extent matrix X_k is inverse-Wishart distributed with $\nu_{k|k'}$ degrees of freedom and scale matrix $V_{k|k'}$.

A. Prediction Step

We assume that the state transition density is approximated as a product of Gaussian and Wishart distributions [12]

$$p(\xi_k|\xi_{k-1}) \approx p(x_k|x_{k-1})p(X_k|X_{k-1}, x_{k-1}) = \mathcal{N}(x_k; g(x_{k-1}), Q)\mathcal{W}(X_k; \kappa_{k-1}, E_{x_{k-1}} X_{k-1} E_{x_{k-1}}^{\mathrm{T}}/\kappa_{k-1}), \tag{13}$$

where $g(\cdot)$ denotes the (nonlinear) motion model, Q denotes the process noise covariance and E_x denotes the transformation matrix, typically a rotation matrix depending on kinematic state x. Given the state transition density (13) and the posterior density $p(\xi_{k-1}|Z_{1:k-1})$ in (12), the predicted parameters $\{m, P, v, V\}_{k|k-1}$ of $p(\xi_k|Z_{1:k-1})$ are [13]:

$$m_{k|k-1} = g(m_{k-1|k-1}), \tag{14a}$$
$$P_{k|k-1} = G_{k-1} P_{k-1|k-1} G_{k-1}^{\mathrm{T}} + Q_{k-1}, \tag{14b}$$
$$\nu_{k|k-1} = 6 + e^{-T_s/\tau}(\nu_{k-1|k-1} - 6), \tag{14c}$$
$$V_{k|k-1} = e^{-T_s/\tau} E_{m_{k-1}} V_{k-1|k-1} E_{m_{k-1}}^{\mathrm{T}}, \tag{14d}$$

where $G_{k-1} = \nabla_x g(x)|_{x=m_{k-1|k-1}}$, T_s is the sampling time and τ is a maneuvering correlation constant. The kinematic state prediction in (14a) and (14b) follows the standard prediction step of a (nonlinear) Kalman filter, whereas the extent state prediction is given by (14c) and (14d).

B. Update Step with Learned HTG Measurement Model

As the learned HTG measurement model depends on the object state, the object state is updated in a recursive fashion. The complete state update step for the learned hierarchical truncation model runs iteratively over two building blocks: 1) selecting a learned hierarchical truncation model conditioned on the aspect angle, and 2) object state update using converted measurement statistics [21], until a convergence criteria is met. From one iteration to the next, a refined HTG model can be obtained by having a more accurate object state estimate.

Specifically, at the t-th iteration, we first compute the aspect angle (11) using the updated object state $\xi_{k|k}^{(t-1)}$ estimated at the $(t-1)$-th iteration, and then we select a learned HTG measurement model. The RM approach uses the statistics (mean and spread) of Gaussian distributed measurements to update the predicted state density [12], [13]. To integrate the HTG measurement model into the state update step of RM, an effective way is to construct Gaussian measurement statistics using the HTG distributed measurements [21].

The procedure is given as follows; refer [21] for more details. We first compute the analytical mean and spread of $n_k^c = n_k(1-c_{D_k})/c_{D_k}$ pseudo measurements that follow a truncated Gaussian distribution with density support $D_k^c = \mathbb{R}^2 \setminus D_k$. Next, we convert the learned sensor noise covariance R_{UC} and the computed pseudo measurement statistics from normalized object coordinate system to the global coordinate system using $\xi_{k|k}^{(t-1)}$. Then we can take the weighted sum of the transformed pseudo measurement mean/spread and the received measurement mean/spread to obtain the converted Gaussian measurement mean $\bar{\breve{z}}_k$ and spread $\Sigma_{\breve{Z}_k}$. Given the learned HTG model (4) and the predicted density $p(\xi_k|Z_{1:k-1})$, the updated parameters $\{m, P, v, V\}_{k|k}$ determining the posterior density are given as [13]:

$$m_{k|k} = m_{k|k-1} + K\varepsilon, \tag{15a}$$
$$P_{k|k} = P_{k|k-1} - KHP_{k|k-1}, \tag{15b}$$
$$\nu_{k|k} = \nu_{k|k-1} + (n_k + n_k^c), \tag{15c}$$
$$V_{k|k} = V_{k|k-1} + \hat{N} + \hat{Z}, \tag{15d}$$

where

$$\hat{N} = \hat{X}^{1/2} S^{-1/2} \varepsilon \varepsilon^{\mathrm{T}} S^{-\mathrm{T}/2} \hat{X}^{\mathrm{T}/2}, \ \varepsilon = \bar{\breve{z}}_k - H m_{k|k-1}$$
$$\hat{Z} = \hat{X}^{1/2} \hat{R}^{-1/2} \Sigma_{\breve{Z}_k} \hat{R}^{-\mathrm{T}/2} \hat{X}^{\mathrm{T}/2},$$
$$\hat{R} = \rho \hat{X} + R, \ \hat{X} = V_{k|k-1}/(\nu_{k|k-1} - 6),$$
$$S = HP_{k|k-1}H^{\mathrm{T}} + \hat{R}/(n_k + n_k^c), \ K = P_{k|k-1} H S^{-1}.$$

At the first iteration, we may use the predicted state estimate at time step k to initialize the algorithm by setting $\xi_{k|k}^{(0)} = \xi_{k|k-1}$.

Compared with our previous update step of [21] and [22], the update step for the learned HTG model directly uses the offline learned model parameters B, θ, ρ and R_{UC}. This makes the new update step computationally simpler as it skips the bounds update and converges faster. By incorporating the learned measurement model into the update step, the proposed method also shows improved capability of EOT with sparse measurements.

V. Experimental Validation

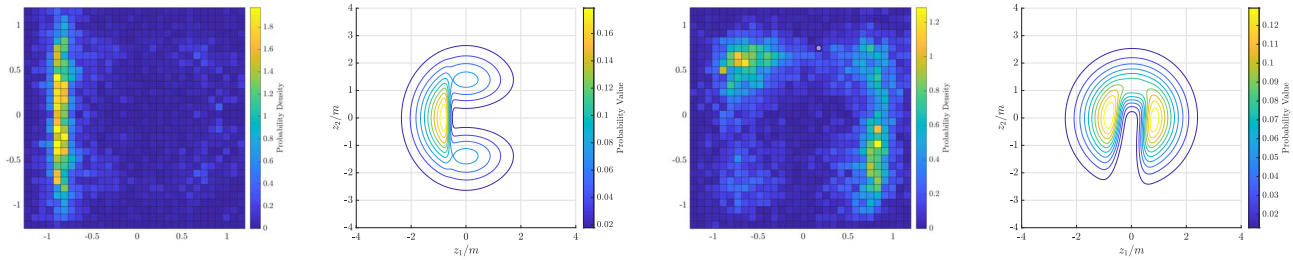
The real-world automotive radar data from the nuScenes dataset [17] are used to learn the geometry-related model parameters. The annotated keyframe for radar measurements is sampled at 2 Hz for 1000 scenes of 20 seconds duration. The ego vehicle is mounted with five 77-GHz FMCW radars at the four corners and the front bumper. Each radar sensor has a 13-Hz capture frequency and up to 250-m range. The proposed EOT algorithm is validated on annotated trajectories of objects that are continuously covered by at least one radar point for a certain time period.

A. Offline Measurement Model Training

To obtain the accumulated radar point cloud for vehicle detections, we first extract the set of objects labeled as car[1], and then for each object we extract all the radar points within its annotated 3D bounding box with a scaling ratio of 1.2. This is to account for the spread of the measurements around object edges. After the initial measurement extraction, (5) is applied to transform the radar point cloud to a unit coordinate system. To prevent the training data mainly consisting of data points collected from dominating aspect angles (e.g., from the rear), we utilized the annotation information to only consider radar detections with "cross-moving" dynamic property[2], "unambiguous" radial velocity and probability being an artifact

[1]There are other vehicle types in the dataset like truck and bus, our learned measurement model can be extended to cover different sizes of vehicles.

[2]This increases the proportion of radar detections that appear on the left and right sides of the car.



(a) $\phi' \in [-\pi/8, \pi/8)$ (b) $\phi' \in [-\pi/8, \pi/8)$ (c) $\phi' \in [5\pi/8, 7\pi/8)$ (d) $\phi' \in [5\pi/8, 7\pi/8)$

Fig. 3: Histograms of radar points cloud and pdfs of trained HTG model, conditioned on different aspect angle ranges.

due to multipath propagation less than 25%. This results in a more balanced training dataset with 20743 radar points in total as shown in Fig. 1 (a).

Similar to [19], a set of eight hierarchical truncation models are trained based on a subset of radar points conditioned on the aspect angle ϕ' in the equally spaced intervals $[-\pi, \pi]$ in the unit coordinate system. To utilize the symmetry information of vehicle detections, pseudo reflection points with respect to the horizontal axis are added to the training data to enrich the information. Subsets of radar points cloud conditioned on different aspect angle ranges and their corresponding learned HTG densities are shown are shown in Fig. 3.

B. Tracking Performance Evaluation on nuScenes Dataset

We compare the tracking performance between RM [13] and the proposed method HTG-RM. The object is assumed to move following a coordinated turn motion model with standard deviations of polar and angular acceleration noise $\sigma_v = 1.5$ and $\sigma_\omega = 0.1$, respectively. For RM, the standard deviation of sensor noise is $\sigma_r = 0.2$. For HTG-RM, the learned sensor noise covariance is applied and the number of iterations used in the update step is set to 5. For both RM and HTG-RM, the initial kinematic state is set to the same as the first annotated object state and the initial extent state is determined by the average size of all the annotated cars in the nuScenes dataset.

To validate the performance of the proposed method, a set of three courses representing common driving routes was extracted from the nuScenes ground truth data. For each course, the annotations at the keyframes and the ego vehicle trajectory are shown in Fig. 4. In scenario 1, the object moves following the ego vehicle on a nearly straight lane. In scenario 2, the object first moves on a straight lane behind and to the left of the ego vehicle, and then it takes a right turn and moves behind and to the right of the ego vehicle. In scenario 3, the object moves in an S curve behind the ego vehicle.

To evaluate the object kinematic and extent states estimation performance on annotated keyframes, we compute the Gaussian Wasserstein Distance (GWD) metric [23] between the ground truth ξ and the estimate $\hat{\xi}$:

$$d_{\mathrm{GW}}(\xi - \hat{\xi}) = ||h(x) - h(\hat{x})||^2 + \mathrm{Tr}(X + \hat{X} - 2(X^{\frac{1}{2}}\hat{X}X^{\frac{1}{2}})^{\frac{1}{2}}), \tag{16}$$

where the extent state X of the annotation bounding box with length l and width w is given by $\mathrm{diag}(l^2/4, w^2/4)$. The GWDs of RM and HTG-RM for the three different scenarios are compared in Fig. 5. The results show that the proposed method HTG-RM outperforms RM by a large margin except for a few annotated keyframes. We found that such performance degradation happens when the predicted state deviates from the ground truth to a large extent. This means that the update step of HTG-RM could be sensitive to initialization.

In addition, snap-shots of object state estimates of RM and HTG-RM are illustrated in Fig. 6. It can be seen that the RM estimates tend to fit the part of the object that is in sight of the ego vehicle whereas HTG-RM does not present this behavior in general and shows improved performance in terms of both the kinematic state and the extent state.

VI. Conclusion

In this paper, we have presented a data-driven HTG measurement model learned from real-world automotive radar data. We have also developed a new RM update step tailored to the learned measurement model. The effectiveness of the proposed approach has been verified on data from the nuScenes dataset.

References

[1] K. Granström, M. Baum, and S. Reuter, "Extended object tracking: Introduction, overview, and applications," *Journal of Advances in Information Fusion*, vol. 12, no. 2, 2017.

[2] M. Buhren and B. Yang, "Simulation of automotive radar target lists using a novel approach of object representation," in *Proc. IV Symp.* IEEE, 2006, pp. 314–319.

[3] J. Gunnarsson, L. Svensson, L. Danielsson, and F. Bengtsson, "Tracking vehicles using radar detections," in *Proc. IV Symp.* IEEE, 2007, pp. 296–302.

[4] L. Hammarstrand, M. Lundgren, and L. Svensson, "Adaptive radar sensor model for tracking structured extended objects," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 1975–1995, 2012.

[5] C. Knill, A. Scheel, and K. Dietmayer, "A direct scattering model for tracking vehicles with high-resolution radars," in *Proc. IV Symp.* IEEE, 2016, pp. 298–303.

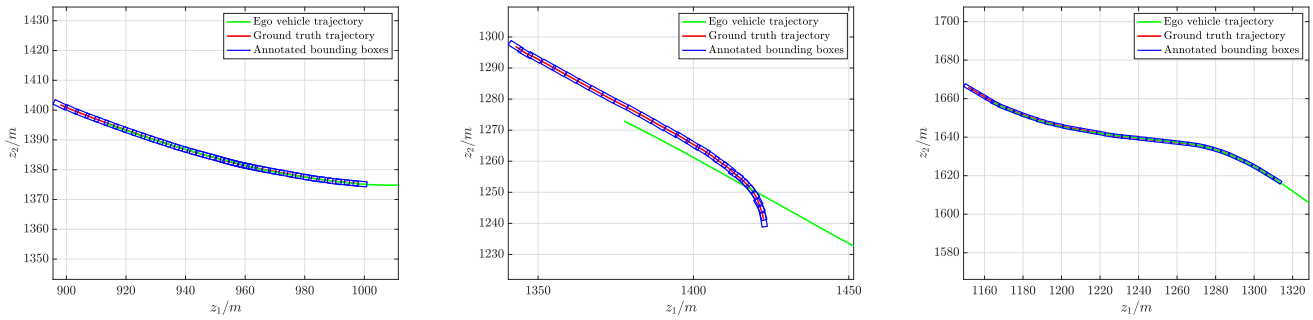
[6] P. Broßeit, M. Rapp, N. Appenrodt, and J. Dickmann, "Probabilistic rectangular-shape estimation for extended object tracking," in *Proc. IV Symp.* IEEE, 2016, pp. 279–285.

[7] X. Cao, J. Lan, X. R. Li, and Y. Liu, "Extended object tracking using automotive radar," in *Proc. 21st Int. Conf. Inf. Fusion*. IEEE, 2018, pp. 1–5.

[8] M. Baum and U. D. Hanebeck, "Extended object tracking with random hypersurface models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 149–159, 2014.

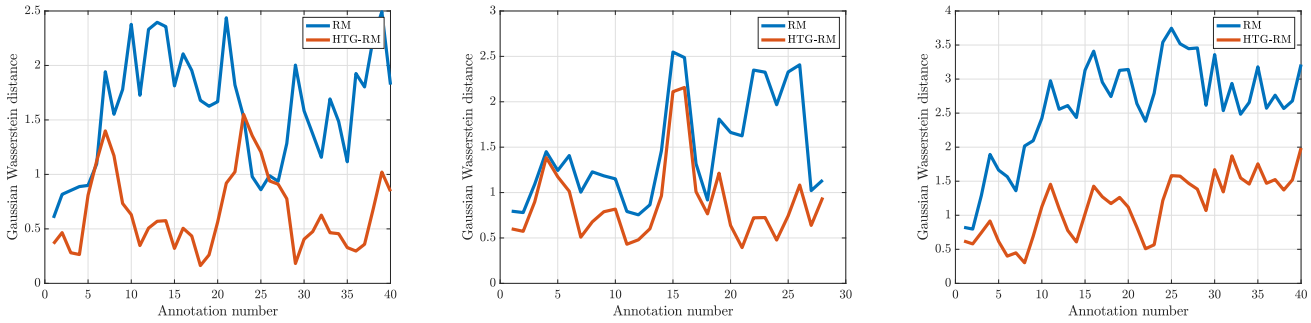
[9] N. Wahlström and E. Özkan, "Extended target tracking using Gaussian processes," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4165–4178, 2015.

[10] K. Thormann, M. Baum, and J. Honer, "Extended target tracking using Gaussian processes with high-resolution automotive radar," in *Proc. 21st Int. Conf. Inf. Fusion*. IEEE, 2018, pp. 1764–1770.



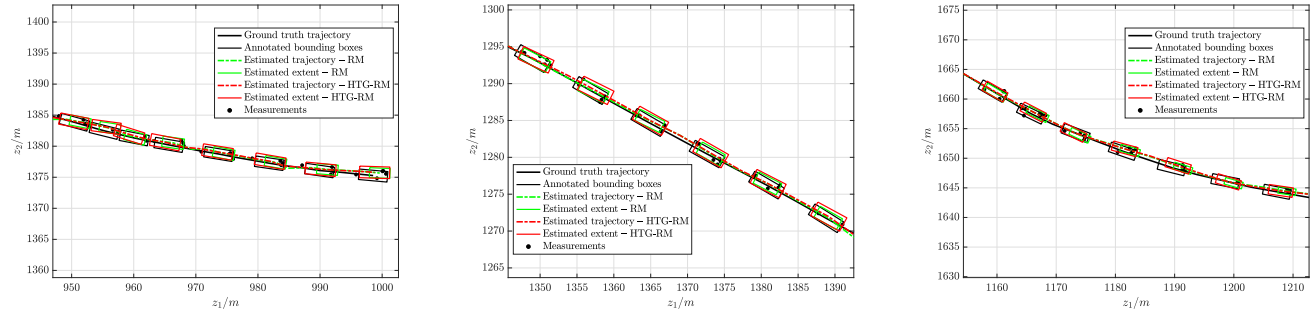
(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

Fig. 4: Ground truth data in global coordinate system for three different scenarios.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

Fig. 5: Tracking performance comparison in terms of Gaussian Wasserstein distance for the three different scenarios.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

Fig. 6: Snap-shots of object state estimates in global coordinate system for the three different scenarios.

[11] H. Kaulbersch, J. Honer, and M. Baum, "A Cartesian B-spline vehicle model for extended object tracking," in *Proc. 21st Int. Conf. Inf. Fusion*. IEEE, 2018, pp. 1–5.

[12] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 3, pp. 1042–1059, 2008.

[13] M. Feldmann, D. Franken, and W. Koch, "Tracking of extended objects and group targets using random matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1409–1420, 2010.

[14] U. Orguner, "A variational measurement update for extended target tracking with random matrices," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3827–3834, 2012.

[15] M. Schuster, J. Reuter, and G. Wanielik, "Probabilistic data association for tracking extended targets under clutter using random matrices," in *Proc. 18th Int. Conf. Inf. Fusion*. IEEE, 2015, pp. 961–968.

[16] G. Vivone, P. Braca, K. Granstrom, and P. Willett, "Multistatic bayesian extended target tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 6, pp. 2626–2643, 2016.

[17] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," *arXiv preprint:1903.11027*, 2019.

[18] A. Scheel and K. Dietmayer, "Tracking multiple vehicles using a variational radar model," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3721–3736, 2018.

[19] H. Kaulbersch, J. Honer, and M. Baum, "EM-based extended target tracking with automotive radar using learned spatial distribution models," in *Proc. 22nd Int. Conf. Inf. Fusion*, 2019, pp. 1–8.

[20] P. Broßeit, B. Duraisamy, and J. Dickmann, "The volcanormal density for radar-based extended target tracking," in *IEEE ITSC*, 2017, pp. 1–6.

[21] Y. Xia, P. Wang, and K. Berntorp, et. al., "Extended object tracking using hierarchical truncation measurement model with automotive radar," in *IEEE ICASSP*, 2020, pp. 4900–4904.

[22] ——, "Extended object tracking using hierarchical truncation model with partial-view measurements," in *IEEE SAM*, 2020.

[23] S. Yang, M. Baum, and K. Granström, "Metrics for performance evaluation of elliptic extended object tracking methods," in *Proc. MFI Conf.* IEEE, 2016, pp. 523–528.