MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

Extended Object Tracking with Spatial Model Adaptation Using Automotive Radar

Yao, Gang; WANG, PU; Berntorp, Karl; Mansour, Hassan; Boufounos, Petros T.; Orlik, Philip V.

TR2021-138 November 13, 2021

Abstract

This paper considers extended object tracking (EOT) using high-resolution automotive radar measurements with spatial model adaptation. This is motivated by the fact that offline learned spatial models maybe over-smoothed due to insufficient offline labels and can be mismatched to onboard radar sensors due to different specifications. To refine the offline learned spatial representation in an online setting, we first derived a modified unscented Rauch-Tung-Striebel (RTS) smoother that explicitly accounts for the offline learned model (i.e., the B-spline chained ellipses model). The smoothed state estimates are then used to create an online batch of *state-independent* training data which is finally utilized by an expectation-maximization algorithm to update the spatial model parameters. Numerical validation is provided to verify the effectiveness of the proposed online model adaptation scheme.

International Conference on Information Fusion (FUSION) 2021

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

Extended Object Tracking with Spatial Model Adaptation Using Automotive Radar

Gang Yao*, Pu Wang, Karl Berntorp, Hassan Mansour, Petros Boufounos, and Philip V. Orlik
Mitsubishi Electric Research Laboratories (MERL)
Cambridge, MA 02139, USA.

***Abstract*—This paper considers extended object tracking (EOT) using high-resolution automotive radar measurements with online spatial model adaptation. This is motivated by the fact that offline learned spatial models may be over-smoothed due to coarsely labeled training data and can be mismatched to onboard radar sensors due to different specifications. To refine the offline learned spatial representation in an online setting, we first apply the unscented Rauch-Tung-Striebel (RTS) smoother that explicitly accounts for the predicted and filtered states based on the offline learned model (i.e., the B-spline chained ellipses model). The smoothed state estimates are then used to create an online batch of state-decoupled training data that are subsequently utilized by an expectation-maximization algorithm to update the spatial model parameters. Numerical validation with synthetic automotive radar measurements is provided to verify the effectiveness of the proposed online model adaptation scheme.**

***Index Terms*—Automotive radar, extended object tracking, smoothing, model adaptation.**

I. Introduction

With increasingly higher angular resolution and rapid advances in automotive radar, more and more detection points per time scan are obtained for a single object and, as a result, extended object tracking (EOT) is well suited to summarize the statistics from the multiple detection points and track the object. Compared with traditional point object tracking, EOT can estimate not only the kinematic state but also the extent state including the length and width of objects [1].

One key issue in EOT is to capture the spatial representation of multiple detection points given the object state including the position, orientation, length, and width. Besides the two main categories, i.e., contour models [2]–[10] and surface models [11]–[16], a third category, i.e., the surface-volume models, was recently considered, to balance between the contour models and the surface models with more realistic features customized to the automotive radar measurements [17]–[23]; see Fig. 2 (a) for an illustration of the real-world accumulated automotive radar measurements [18] in a unit coordinate, where the origin is located at the middle of the rear axle.

More recently, we combined the contour and surface models to introduce a new surface-volume model, i.e., the B-spline chained ellipses model, for the automotive radar measurements in [24]. Particularly, the B-spline chained ellipses model places regularized multiple ellipses around the vehicle contour to describe the spatial representation of automotive radar measurements given the object state. Such a regularization is enforced by the requirement that the center of each ellipsis component (e.g., a typical surface model) has to be located on an enclosed B-spline curve that represents the vehicle contour (e.g., a typical contour model).

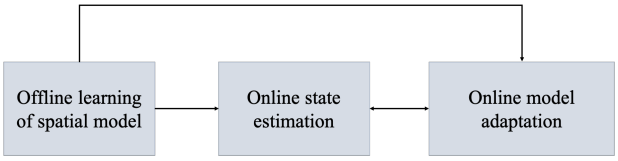


Fig. 1. The overall workflow of an extended object tracking algorithm using the B-spline chained ellipses model: 1) offline learning of spatial model, 2) online state estimation, and 3) online model adaptation. While 1) and 2) were considered in [24], this paper focuses on 3) online model adaptation that refines the offline learned spatial model and further improves the online state estimation performance with a more customized spatial model that may fit better for onboard automotive radar measurements.

As shown in Fig. 1, we build on our previous effort of [24] for 1) offline model parameter learning and 2) online state estimation using the probabilistic multi-hypothesis tracking (PMHT) along with the unscented transform (UT), and focus on 3) online spatial model adaptation that refines the offline learned spatial model and further improves the online state estimation performance with a more customized spatial model that fits for onboard automotive radar measurements. The motivations for online model adaptation are two-fold: First, there might be mismatches on radar sensor specifications between onboard sensors and those used for offline data collection. Second, offline training data with coarse vehicle labels may lead to an over-smoothed offline learned spatial model that averages over different vehicle models. For instance, a coarsely labeled dataset may include sedan and SUV in the same class.

To achieve the online model adaptation, we consider an online smoothing algorithm that uses online state estimates based on the offline learned B-spline chained ellipses model, and performs a backward recursion to smooth the online state filtering with all observed measurements. Particularly, the unscented Rauch-Tung-Striebel (RTS) smoother is applied to compute smoother gain, smoothed mean, and the smoothed covariance matrix at each time step by recursively computing

*The work of G. Yao was done during his internship at MERL. He is a PhD student at University of Connecticut, Storrs, CT

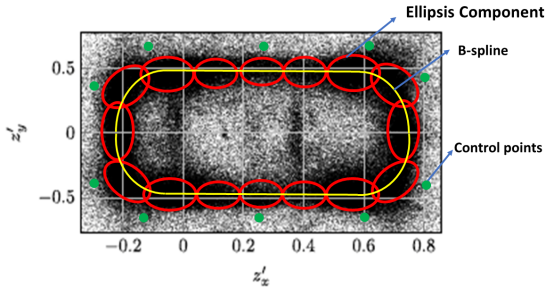


Fig. 2. The B-Spline chained ellipses model [24] that fits for accumulated real-world automotive radar measurements [18].

the posterior of the state conditioned on all observed measurements backward from the filtered state estimate at the last time step. Then, the smoothed state estimates are used to convert all observed measurements in the global coordinate system to a batch of *state-decoupled* training data in a unit coordinate system. As the online batch of training data are state-decoupled and only depends on the spatial model, we use the expectation-maximization (EM) algorithm to update the model parameters within a regularization on the distance to the offline learned model parameters.

The remainder of this paper is organized as follows. Section II introduces the B-spline chained ellipses model designed to resemble the empirical distribution of real-world automotive radar measurements. The online state estimation and online model adaptation are introduced in Section III. Section IV provides numerical evaluation and performance comparison in terms of state estimation errors with and without online modal adaptation. It is then followed by conclusions in Section V.

II. B-Spline Chained Ellipses Model

As shown in Fig. 2, the proposed spatial model consists of L Gaussian components (i.e., ellipses) with their component means located on a B-spline curve. Given N_{offline} offline training data $\tilde{\mathbf{Z}} = \{\tilde{\mathbf{z}}_i\}_{i=1}^{N_{\text{offline}}}$, we can associate each measurement, e.g., the i-th measurement, to the l-th ellipsis component (centered at μ_l with an extent covariance matrix Σ_l) with an association probability ρ_i^l. Given the measurement-to-ellipse assignment, the likelihood function becomes

$$\phi\left(\tilde{\mathbf{Z}}|\bar{N}_l,\mu_l,\Sigma_l,\rho^l\right) \propto \mathcal{N}\left(\bar{\mathbf{z}}_l;\mu_l,\frac{\Sigma_l}{\bar{N}_l}\right) \times W\left(\overline{Z}_l;\bar{N}_l-1,\Sigma_l\right)$$

where $\bar{N}_l = \sum_{i=1}^{N_{\text{offline}}} \rho_i^l$,

$$\bar{\mathbf{z}}_l = \frac{\sum_{i=1}^{N_{\text{offline}}} \rho_i^l \tilde{\mathbf{z}}_i}{\sum_{i=1}^{N_{\text{offline}}} \rho_i^l}, \tag{1}$$

$$\overline{\mathbf{Z}}_l = \sum_{i=1}^{N_{\text{offline}}} \rho_i^l \left(\tilde{\mathbf{z}}_i - \bar{\mathbf{z}}_l\right)\left(\tilde{\mathbf{z}}_i - \bar{\mathbf{z}}_l\right)^{\mathrm{T}}, \tag{2}$$

are, respectively, the sample mean and spread of the l-th ellipse, $\mathcal{N}$ denotes the Gaussian distribution and $\mathcal{W}$ is the Wishart distribution. With all L ellipses and given the measurement-to-ellipse association, the L random matrices model is defined as

$$p(\tilde{\mathbf{Z}}|\theta,\rho) = \sum_{l=1}^{L} \pi_l \phi\left(\tilde{\mathbf{Z}}|\bar{N}_l,\mu_l,\Sigma_l,\rho^l\right), \tag{3}$$

where θ includes μ_l and Σ_l and the mixture weights π_l are assumed to be equal $\pi_l = 1/L$.

Moreover, we assume that the ellipse centers are located on a B-spline curve defined by $\mathbf{c}(r) \in \mathbb{R}^{2\times 1}$ of degree d [25]

$$\mathbf{c}(r) = \sum_{j=0}^{m} \mathbf{p}_j B_{j,d}(r), \quad 0 \le r \le m-d+1, \tag{4}$$

where $\mathbf{p}_j \in \mathbb{R}^{2\times 1}$ is the j-th control point, $m+1$ is the number of control points, and $B_{j,d}(r)$ is the basis function with a parameter r [25]. By enforcing $\mu_l = \mathbf{c}(r_l)$ with r_l denoting the corresponding parameter of the l-th ellipse center μ_l, the B-spline chained ellipses model is defined as

$$p(\tilde{\mathbf{Z}}|\theta_{\text{offline}},\rho) = \sum_{l=1}^{L} \pi_l \phi\left(\tilde{\mathbf{Z}}|\bar{N}_l,\mathbf{c}(r_l),\Sigma_l,\rho^l\right), \tag{5}$$

where θ_{offline} includes the parameters for the proposed spatial modal that consist of the number of measurements for each component $\bar{N}_l$, control points of the B-spline curve $\{\mathbf{p}_j\}_{j=0}^{m}$, and the covariance matrices of ellipsis components $\{\Sigma_l\}_{l=1}^{L}$.

Given N_{offline} offline training measurements $\tilde{\mathbf{Z}}$, one can estimate the modal parameters θ_{offline} associated with the B-spline chained ellipses model with the offline model learning algorithm considered in [24, Section 2.2]. The details are skipped here due to the space limit.

III. Online State Estimation and Spatial Model Adaptation

In this section, we introduce two main building blocks for extended object tracking using the B-spline chained ellipses spatial model for 1) online state estimation and 2) online model adaptation, corresponding to the upper and lower branches of Fig. 3, respectively.

The upper branch mainly deals with the prediction and update of the random state variables in a *fast-time* scale, i.e., at each time step, by using either an offline learned spatial model or a recently updated one. On the contrary, the lower branch represents a *slow-time* update on the deterministic model parameters associated with the spatial model over all of the T time steps.

A. Online State Estimation

With the offline learned spatial model and an initial state, our earlier effort in [24] considered the online state prediction and update that accounts for the nature of multiple ellipses in the learned model with fixed relative geometry, as shown in the upper branch of Fig. 3. Particularly, the unscented Kalman filter-PMHT (UKF-PMHT) algorithm was considered

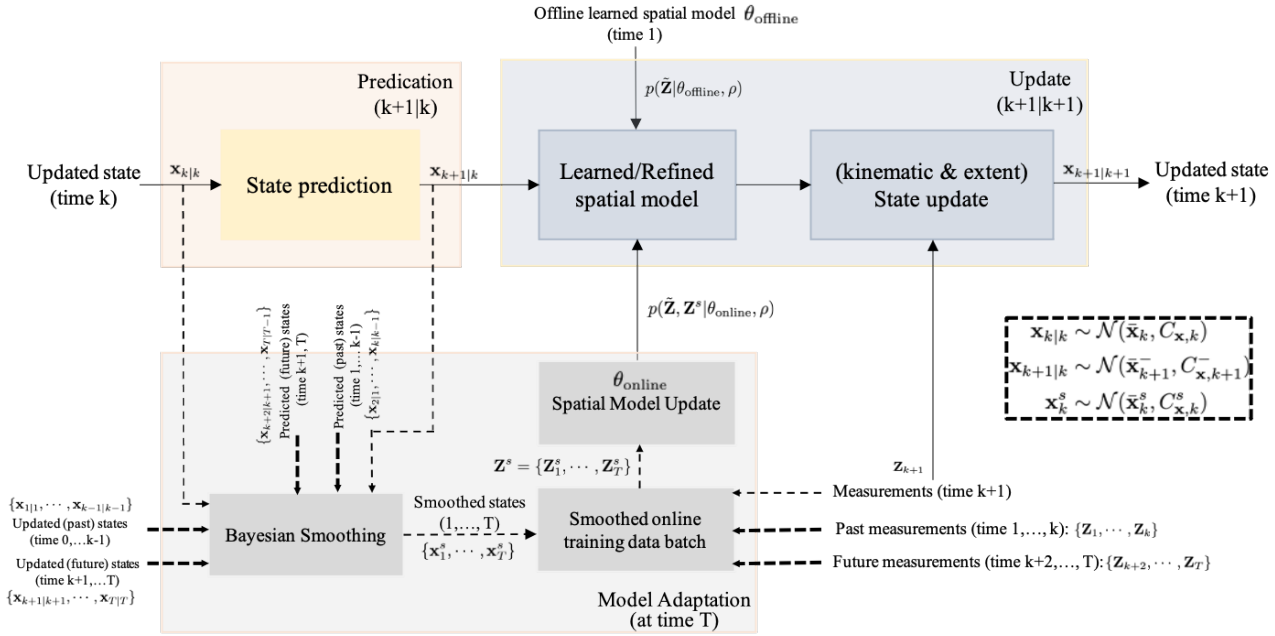


Fig. 3. The diagram for 1) extended object state prediction and update at each time step k (the upper branch) and 2) online modal adaptation at time T (the lower branch). The box in dash line shows the distribution of the updated states $\mathbf{x}_{k|k}$, predicted states $\mathbf{x}_{k+1|k}$, and smoothed states $\mathbf{x}^s_k$ with corresponding mean and covariance matrix given in Section III-A and Sections III-B.

to predict and update the vehicle state, given the motion model and the measurements $\mathbf{Z}_1, \cdots, \mathbf{Z}_k$,

$$\mathbf{x}_k = [x_{m,k}, y_{m,k}, v_k, \psi_k, \omega_k, l_k, w_k]^{\mathrm{T}}, \tag{6}$$

where $[x_{m,k}, y_{m,k}]^{\mathrm{T}}$ is the center of the vehicle, v_k is the polar velocity of the vehicle [26], ψ_k is the heading orientation, ω_k is the turning rate, l_k and w_k are length and width of the vehicle, respectively, and $\mathbf{Z}_k = \{\mathbf{z}_{i,k}\}_{i=1}^{N_k}$ with N_k denoting the number of measurements at time step k.

Since this step was covered in [24, Section 3], we will only highlight the key steps here. The UKF-PMHT follows the two-step prediction and update procedure shown in the upper branch of Fig. 3.

First, for the state prediction (the first block in the upper branch of Fig. 3), the approximate mean $\bar{\mathbf{x}}^-_{k+1}$ and covariance matrix $C^-_{\mathbf{x},k+1}$ of the predicted state $\mathbf{x}_{k+1|k}$ is obtained with the unscented transform by generating sigma points and weights according to the mean $\bar{\mathbf{x}}_k$ and covariance matrix $C_{\mathbf{x},k}$ of the recently updated state $\mathbf{x}_{k|k}$ and propagating through the (nonlinear) motion model (e.g., coordinated turn (CT) with polar velocity [26]).

Second, for the state update (the second block in the upper branch of Fig. 3), it consists of two parts: one is to use the predicted state and the offline learned spatial model to generate *predicted measurements* (i.e., the block of learned/refined spatial model); and the other is to generate *synthetic measurements* and update the vehicle state using both the predicted and synthetic measurements (i.e., the block of (kinematic & extent) state update).

To generate the *predicted measurements*, each ellipsis component (indexed by l) of the offline learned B-spline chained ellipses model $p(\tilde{\mathbf{Z}}|\theta_{\text{offline}}, \rho)$ in the unit coordinate system is mapped to the global coordinate system by going through another unscented transform or the Taylor-series expansion via the following unit-to-global coordinate transform

$$f_{l,k}(\mathbf{x}_{k+1|k}) = \mathbf{m}_{k+1|k} + R(\psi_{k+1|k}) S(l_{k+1|k}, w_{k+1|k}) \mathbf{x}_{l,\mu} \tag{7}$$

where $\mathbf{x}_{l,\mu} \sim \mathcal{N}(\mu_l, \Sigma_l)$ with μ_l and Σ_l determined by the offline learned model parameters θ_{offline}. And the predicted state $\mathbf{x}_{k+1|k}$ specifies $\mathbf{m}_{k+1|k} = [x_m, y_m]$ as the predicted vehicle center position, $R(\psi_{k+1|k}) \in \mathbb{R}^{2\times 2}$ as the rotation matrix with angle given by the orientation $\psi_{k+1|k}$, and a scaling matrix of $S(l_{k+1|k}, w_{k+1|k}) = \mathrm{diag}(l_{k+1|k}, w_{k+1|k})$ given by length and width. The sigma points for each ellipsis component are generated as an augmented vector that combines the predicted state and the ellipsis component $\mathbf{x}_{l,\text{aug}} = [\mathbf{x}^T_{k+1|k}, \mathbf{x}^T_{l,\mu}]^T$. As a result, we may form the *predicted measurement* $\mathbf{z}^-_{l,k+1}$, its covariance matrix $C^-_{l,\mathbf{x}}$, and cross-covariance matrix $C_{l,\mathbf{xz}}$ for each ellipsis component using the propagated sigma points via (7) plus the noise. It is noted that all predicted measurements $\mathbf{z}^-_{l,k+1}$ in the learned offline spatial model are mapped to the global coordinate system via the same set of (predicted) vehicle state $\mathbf{x}_{k+1|k}$.

To generate the *synthetic measurements* using the current measurements $\mathbf{Z}_{k+1}$ at time $k+1$, we use the PMHT to associate the measurements to each ellipse in the global coordinate system. By treating each ellipsis in our learned

spatial model as an individual extended object, we can directly apply the data association step in the PMHT-E in [27] by probabilistically assigning associate weights of measurements to each ellipsis component and generating synthetic measurements, i.e., the synthetic mean and the synthetic spread, for each ellipsis.

To update the vehicle state $\mathbf{x}_{k+1}$, we proposed an iterative state update step over the ellipsis l by using the predicted and synthetic measurements in [24]. Since the predicted measurements for all ellipsis components are a function of the same vehicle state $\mathbf{x}_{k+1|k}$ (see (7)), this iterative state update enforces that all (predicted and synthetic) measurements are inherently used to update the same predicted state $\mathbf{x}_{k+1|k}$ to the newly updated state $\mathbf{x}_{k+1|k+1}$. This iterative state update step is different from [27] where the predicted measurements were generated according to multiple independent vehicle states.

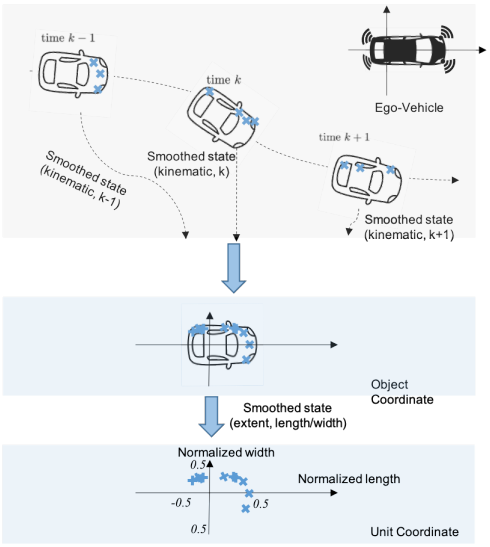


Fig. 4. Creating an online batch of smoothed training data in the unit coordinate system for updating spatial model parameters.

B. Online Model Adaptation

The online model adaptation aims to refine the offline spatial model (in terms of θ) after certain time steps, say T time steps. As shown in the lower branch of Fig. 3, the online model adaptation accumulates the past and future (w.r.t. time step k) updated states $\mathbf{x}_{k|k}$, predicted states $\mathbf{x}_{k+1|k}$, and measurements $\mathbf{Z}_k$ to improve the state estimates (by Bayesian state smoothing), create an online batch of training data, and update the spatial model parameters.

1) Unscented Rauch-Tung-Striebel (RTS) Smoother: To improve the state estimation performance, all of the measurements obtained during the T time steps are used to smooth the estimates. Compared with the random matrix-based Bayesian smoothing algorithm which has to recursively deal with the extent covariance matrix [28], the offline learned B-spline model enables a simple definition of the overall vehicle state as the concatenation of the conventional kinematic state elements (i.e., the first 5 elements in (6)) and the extent state elements in terms of the length and width in (6). Subsequently, this further leads to a direct application of standard point-target Bayesian smoothing algorithms on the augmented kinematic-and-extent state vector. Here, we consider the unscented RTS smoother [29], [30] to smooth the vehicle states after T time steps.

Particularly, from the above state update step, the update state $\mathbf{x}_{k|k}$ at time k follows the Gaussian distribution with mean $\bar{\mathbf{x}}_k$ and covariance $C_{\mathbf{x},k}$. First, the sigma points $\mathcal{X}_i$ and corresponding weights $W_i^{(m)}$ and $W_i^{(C)}$ are generated according to the statistics of the update state:

$$
\begin{aligned}
\mathcal{X}_0 &= \bar{\mathbf{x}}_k \\
\mathcal{X}_i &= \bar{\mathbf{x}}_k + \left(\sqrt{(M+\lambda)C_{\mathbf{x},k}}\right)_i \qquad i = 1,\dots,M \\
\mathcal{X}_i &= \bar{\mathbf{x}}_k - \left(\sqrt{(M+\lambda)C_{\mathbf{x},k}}\right)_i \qquad i = M+1,\dots 2M \\
W_0^{(m)} &= \lambda/(M+\lambda) \\
W_0^{(C)} &= \lambda/(M+\lambda) + (1-\alpha^2+\beta) \\
W_i^{(m)} &= W_i^{(C)} = 1/\left[2(M+\lambda)\right] \qquad i = 1,\dots,2M
\end{aligned}
\tag{8}
$$

where M is the dimension of the overall state vector, λ is a scaling parameter as $\lambda = \alpha^2(M+\kappa) - M$, α is the parameter determines the spread of the sigma points around the mean $\bar{\mathbf{x}}_k$, κ is the secondary scaling parameter usually set to 0, and β is used to incorporate the prior knowledge of the state distribution.

These sigma points propagate through the (nonlinear) motion model $\mathbf{x}_{k+1|k} = g(\mathbf{x}_{k|k})$ (e.g., coordinated turn (CT) with polar velocity [26]) and we have $\mathcal{Y}_i = g(\mathcal{X}_i)$. Then, the predicted mean $\bar{\mathbf{x}}_{k+1}^-$ and its covariance $C_{\mathbf{x},k+1}^-$ of the predicted state $\mathbf{x}_{k+1|k}$ can be computed using the propagated sigma points:

$$
\bar{\mathbf{x}}_{k+1}^- = \sum_{i=0}^{2M} W_i^{(m)} \mathcal{Y}_i \tag{9}
$$

$$
C_{\mathbf{x},k+1}^- = \sum_{i=0}^{2M} W_i^{(C)} \left[\mathcal{Y}_i - \bar{\mathbf{x}}_{k+1}^-\right]\left[\mathcal{Y}_i - \bar{\mathbf{x}}_{k+1}^-\right]^T \tag{10}
$$

In addition, one can compute the cross-covariance matrix between the predicted states $\mathbf{x}_{k+1|k}$ (via unscented

propagation) and the updated state $\mathbf{x}_{k|k}$ (via the UKF-PMHT of the above online state estimation section) as

$$D_{k+1} = \sum_{i=0}^{2M} W_i^{(C)} \left[\mathcal{X}_i - \bar{\mathbf{x}}_k\right] \left[\mathcal{Y}_i - \bar{\mathbf{x}}_{k+1}^-\right]^T. \quad (11)$$

Finally, the smoothed states $\mathbf{x}_k^s$ is Gaussian distributed with the smoothed mean $\bar{\mathbf{x}}_k^s$ and smoothed covariance matrix $C_{\mathbf{x},k}^s$ as [29], [30]:

$$\bar{\mathbf{x}}_k^s = \bar{\mathbf{x}}_k + G_k \left(\bar{\mathbf{x}}_{k+1}^s - \bar{\mathbf{x}}_{k+1}^-\right) \quad (12)$$

$$C_{\mathbf{x},k}^s = C_{\mathbf{x},k} + G_k \left(C_{\mathbf{x},k+1}^s - C_{\mathbf{x},k+1}^-\right) G_k^{\mathrm{T}}, \quad (13)$$

where $G_k = D_{k+1}[C_{\mathbf{x},k+1}^-]^{-1}$ is the smoother gain. To start with, one can set $\bar{\mathbf{x}}_{T,k}^s = \bar{\mathbf{x}}_T$ and $C_{\mathbf{x},T}^s = C_{\mathbf{x},T}$ at time step T and iteratively compute (12) and (13) backwards in time step $k = T-1, \cdots, 1$.

2) Smoothed Measurements: A Batch of Online Training Data: To create an online batch of training data that may fit better for onboard radar sensors, we make use of all measurements $\mathbf{Z} = \{\mathbf{Z}_1, \ldots \mathbf{Z}_T\}$ up to time step T and corresponding smoothed states $\{\mathbf{x}_1^s, \cdots, \mathbf{x}_T^s\}$ as $\mathbf{x}_k^s \sim \mathcal{N}(\bar{\mathbf{x}}_k^s, C_{\mathbf{x},k}^s)$ with $\bar{\mathbf{x}}_k^s$ and $C_{\mathbf{x},k}^s$ given by (12) and, respectively, (13).

Since our goal is to create *state-decoupled* online training data for the spatial model update, we remove the underlying unknown states $\{\mathbf{x}_1, \cdots, \mathbf{x}_T\}$ from the measurements $\mathbf{Z}$ in the global coordinate system using the smoothed mean $\{\bar{\mathbf{x}}_1^s, \cdots, \bar{\mathbf{x}}_T^s\}$

$$\mathbf{z}_{i,k}^s = [S(\bar{l}_k^s, \bar{w}_k^s)]^{-1}[R(\bar{\psi}_k^s)]^{-1}(\mathbf{z}_{i,k} - \bar{\mathbf{m}}_k^s), \quad (14)$$

where $\{\bar{\mathbf{m}}_k^s, \bar{\psi}_k^s, \bar{l}_k^s, \bar{w}_k^s)\}$ denote the same state variables as defined in (7) except that they now correspond to the smoothed state mean $\bar{\mathbf{x}}_k^s$ of (12). For each time step k, the online training dataset is grouped as $\mathbf{Z}_k^s = \{\mathbf{z}_{i,k}^s\}_{i=1}^{N_k}$, $k = 1, \cdots, K$.

The above state-decoupling process can be illustratively represented by Fig. 4. First, the measurements in the global coordinate system are transformed into a coordinate system that is positioned in the center of the vehicle and oriented such that the x-axis of the new coordinate system points to the vehicle front using the orientation angle $\bar{\psi}_k^s$ and object center $\bar{\mathbf{m}}_k^s$. Then, these measurements are further normalized by the extent states, i.e., length $\bar{l}_k^s$ and width $\bar{w}_k^s$ via the scaling matrix S. The length and width used in (14) can also be chosen as the average values of smoothed length and width over all time steps, i.e., $S = 1/T\mathrm{diag}(\sum_k \bar{l}_k^s, \sum_k \bar{w}_k^s)$.

3) Online Update of Spatial Model Parameters: For the annotation simplicity, we group all smoothed measurements $\mathbf{Z}^s = \{\mathbf{Z}_1^s, \cdots, \mathbf{Z}_T^s\}$ as the online batch of training data and re-assign the measurement index in $\mathbf{Z}^s$ as $\mathbf{Z}^s = \{\mathbf{z}_1^s, \cdots, \mathbf{z}_{N_{\text{online}}}^s\}$ with $N_{\text{online}} = \sum_k N_k$ denoting the total number of measurements over T steps. Given these N_{online} online training data, the model adaptation is conducted by the EM algorithm as follows:

Expectation step is to probabilistically assign the n-th online measurement $\mathbf{z}_n^s$ from the online training batch $\mathbf{Z}^s$ to the l-th ellipsis component via the posterior association probability w_n^l [27]

$$w_n^l = \frac{\frac{1}{L} \times \mathcal{N}(\mathbf{z}_n^s; \mu_l, 4\Sigma_l)}{\frac{1}{L} \times \sum_{l=1}^{L} \mathcal{N}(\mathbf{z}_n^s; \mu_l, 4\Sigma_l) + \epsilon} \quad (15)$$

where μ_l and $4\Sigma_l$ is the mean and the covariance matrix of each component. The scaling factor 4 is introduced by assuming the measurement in each component is uniformly distributed [12] and ϵ is the probability of the uniformly distributed outliers. Then, the synthetic measurements $\bar{\mathbf{z}}_l$, synthetic measurement spread $\overline{\mathbf{Z}}_l$, and the sum of weights for the l-th component are calculated, respectively, as

$$\bar{\mathbf{z}}_l = \frac{\sum_{n=1}^{N_{\text{online}}} w_n^l \mathbf{z}_n^s}{\sum_{n=1}^{N_{\text{online}}} w_n^l} \quad (16)$$

$$\overline{\mathbf{Z}}_l = \sum_{n=1}^{N_{\text{online}}} w_n^l \left(\mathbf{z}_n^s - \bar{\mathbf{z}}_l\right) \left(\mathbf{z}_n^s - \bar{\mathbf{z}}_l\right)^{\mathrm{T}} \quad (17)$$

$$\bar{M}_l = \sum_{n=1}^{N_{\text{online}}} w_n^l. \quad (18)$$

Maximization step is to update the model parameters $\theta = \{\mathbf{p}, \Sigma_l\}$ around the offline learned modal parameters θ_{offline} based on a regularized log-likelihood function that enforces the maximal allowable change on the control points $\mathbf{p}$ of the B-spline curve in (4):

$$\mathcal{L}(\theta) \propto \sum_{l=1}^{L} \left\{ -\frac{\bar{M}_l}{2} \left(\mu_l - \bar{\mathbf{z}}_l\right)^{\mathrm{T}} \Sigma_l^{-1} \left(\mu_l - \bar{\mathbf{z}}_l\right) - \frac{\bar{M}_l + 1}{2} \log|\Sigma_l| - \frac{1}{2}\mathrm{tr}\left(-\frac{1}{2}\overline{\mathbf{Z}}_l \Sigma_l^{-1}\right) \right\} + \lambda \left\|\mathbf{p} - \mathbf{p}_{\text{offline}}\right\|_2^2, \quad (19)$$

where $\mathbf{p}_{\text{offline}} \in \mathbb{R}^{2(m+1)\times 1}$ are the control points corresponding to the offline learned model, λ is regularization parameter that controls the spatial model adaptation rate, and $\|\cdot\|_2$ denotes the l_2 norm.

In a more compact form, the B-spline curve can be represented in a matrix-vector form as $\mu_l = \mathbf{B}_l\mathbf{p}$, , $\mathbf{B}_l = \mathrm{blkdiag}\left(\mathbf{n}_l^{\mathrm{T}}, \mathbf{n}_l^{\mathrm{T}}\right)$, $\mathbf{n}_l = \left[B_{0,d}(r_l), \cdots B_{m,d}(r_l)\right]^{\mathrm{T}}$, and $\mathbf{p} = \left[\mathbf{p}_x^{\mathrm{T}}, \mathbf{p}_y^{\mathrm{T}}\right]^{\mathrm{T}}$ with $\mathbf{p}_x^{\mathrm{T}}$ and $\mathbf{p}_y^{\mathrm{T}}$ denoting the control points in the x- and y- coordinates, respectively. Taking the gradients of $\mathcal{L}(\theta)$ with respect to the parameters θ and set them to 0 results in

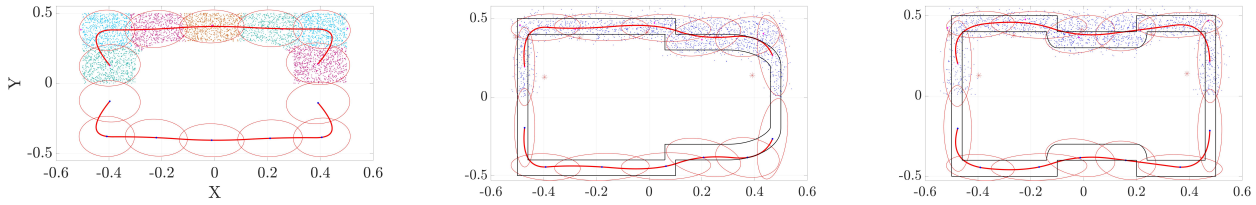
$$\mathbf{p} = \mathbf{H}^{+}\mathbf{M}, \quad (20)$$

where $\mathbf{M} = \sum_{l=1}^{L}\left(\bar{M}_l \mathbf{B}_l^{\mathrm{T}} \Sigma_l^{-1} \bar{\mathbf{z}}_l\right) + 2\lambda\mathbf{p}_{\text{offline}}$ and $\mathbf{H}^{+}$ is the Moore-Penrose inverse of $\mathbf{H} = \sum_{l=1}^{L}\left(\bar{M}_l \mathbf{B}_l^{\mathrm{T}} \Sigma_l^{-1} \mathbf{B}_l\right) + 2\lambda\mathbf{I}$ with $\mathbf{I}$ denoting the identity matrix, and

$$\Sigma_l = \frac{1}{\bar{M}_l + 1}\left[\bar{M}_l \left(\bar{\mathbf{z}}_l - \mu_l\right)\left(\bar{\mathbf{z}}_l - \mu_l\right)^{\mathrm{T}} + \overline{\mathbf{Z}}_l^{\mathrm{T}}\right]. \quad (21)$$

The iteration between the expectation and maximization steps is carried out until the relative changes of the modal parameters is smaller than predefined values, or if the optimization reaches the empirical maximum iteration number.

Finally, given the updated modal parameters $\theta_{\text{online}} = \{\mathbf{p}, \Sigma_l\}$ from the above iterations, the updated spatial modal



(a) an over-smoothed offline learned model (b) online updated model for the case of truck (c) online updated model for the case of tractor

Fig. 5. Vehicle model adaptation with one exemplar run: (a) an over-smoothed offline learned model possibly due to a coarsely labeled offline training dataset, and the updated B-spline chained ellipses models using online radar measurements (blue dots) reflected from (b) the truck (b) and (c) the tractor. In (b) and (c), true vehicle contours are shown in solid black curves, while the online updated contours (B-spline) are shown in bold red curves.

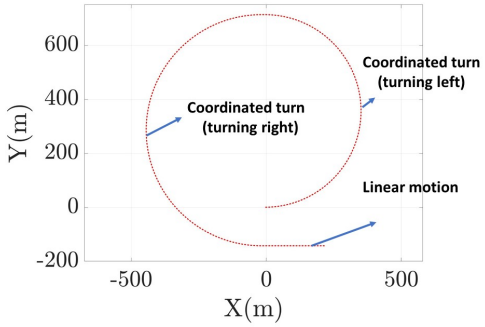


Fig. 6. The simulated trajectory of the vehicle (truck or tractor). It starts with a coordinated turn motion (left turn) for a duration of 101 s, then turns right with a constant velocity for another duration of 121 s, and finally reaches a straight motion for the last 20 s.

$p(\tilde{\mathbf{Z}}, \mathbf{Z}^s|\theta_{\text{online}}, \rho)$ replaces the offline learned spatial model $p(\tilde{\mathbf{Z}}|\theta_{\text{offline}}, \rho)$ in the update block of Fig. 3.

IV. Performance Evaluation

To evaluate the state estimation performance with and without online modal adaptation, we ran 100 Monte-Carlo simulations for each scenario to obtain the root mean square errors (RMSE) for both kinematic and extent state vectors.

A. Simulation Configuration

Truck and tractor vehicles are simulated with different contours (black solid lines) as shown in Fig. 5 (b) and (c). The length and width of the vehicles are 5 and 2 meters, respectively. For either case, the measurements are uniformly distributed 0.2 m inside its contour. The number of measurements in each scan is generated based on $\max(0, \mathbf{Pois}(20))$, where $\mathbf{Pois}(20)$ is the Poisson distribution with mean 20. The variance of the sensor noise is $\text{diag}\left(0.05^2, 0.05^2\right)$ in the unit of m^2.

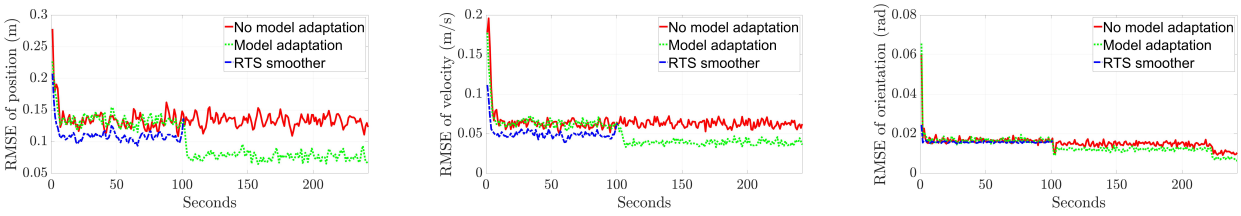
Fig. 6 shows the simulated trajectory that consists of several types of motions. Specifically, the vehicle starts with a CT motion model (left turn) for a duration of 101 s, then turns right with a constant velocity for another duration of 121 s, and finally reaches a straight motion for 20 s. The velocity of the vehicle is always kept at 11.2 m/s or, equivalently, 25 mph. The sampling interval is 1 s which gives in total 242 time steps.

The states are initialized as random variables with mean $\mathbf{x} = [x_m, y_m, v, \psi, \omega, l, w]^{\text{T}} = [0, 0, 11.2, 0, 0.01, 5, 2]^{\text{T}}$ and covariance $\text{diag}\left(0.5^2, 0.5^2, 0.05, 0.1^2, 0.035^2, 0.005^2, 0.005^2\right)$ (with position in meters and orientation in radians). The CT motion model with polar velocity [26] is used as the dynamic motion model for the kinematic states with $\sigma_{\dot{V}}^2 = 0.1^2$ and $\sigma_{\dot{\omega}}^2 = 0.035^2$. The process noise variances for the extent state, i.e., the length and width, are set to be $\sigma_l^2 = \sigma_w^2 = 1e^{-10}$, considering the changes in the vehicle size are small over time.

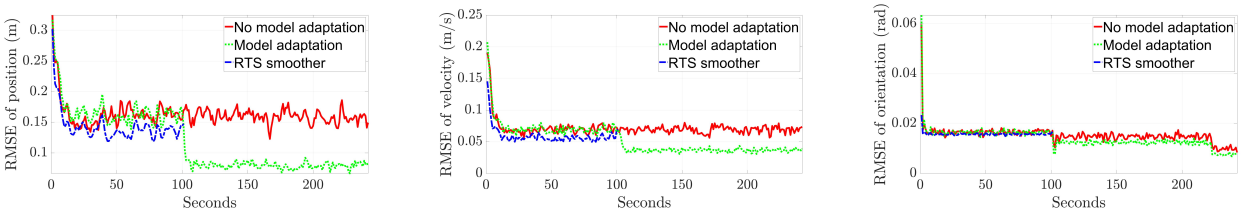
B. Qualitative Evaluation of Online Model Adaptation

We first illustrate the performance of online model adaptation from one exemplar run for both cases of truck and tractor in Fig. 5. To mimic an over-smoothed learned B-spline chained ellipses model due to a coarsely labeled offline training dataset, we use the B-spline chained ellipses model in Fig. 5 (a) as the offline learned spatial model. It is clear that this over-smoothed offline learned model does not reflect the contour feature of either the truck or the tractor.

For the first 101 s, i.e., the left-turn motion, the UKF-PMHT-based state estimation uses the offline learned modal to update the state. At time step $T = 101$, the online modal adaptation is performed using the measurements collected between time step 1 and 101, following Section III-B. Fig. 5 (b) and (c) show, respectively, the online updated B-spline chained spatial models for the truck and tractor at time step $T = 101$ using corresponding online automotive radar measurements (blue dots). It is seen that updated B-spline curves (in bold red curves) follow the trend of the underlying true vehicle contours (in solid black curves), particularly around the front portion of the truck and the middle portion of the tractor. Besides the refined contours, the ellipsis components (centers and covariance matrices) are also updated to better capture the underlying distribution of the online radar measurements. This can be seen from the horizontally stretched ellipses around the back portion of the truck.

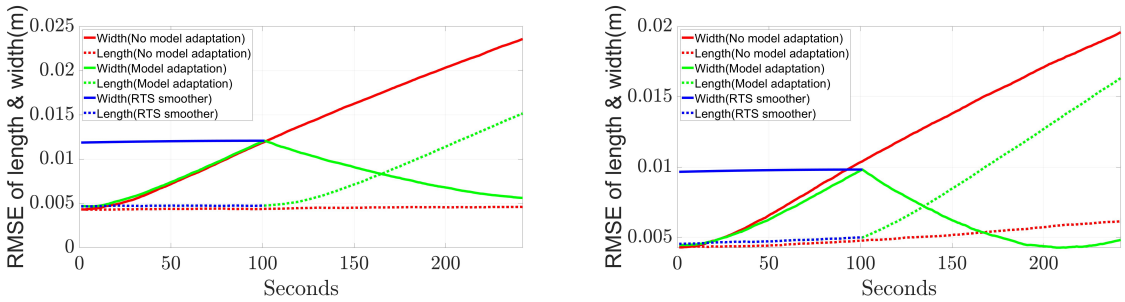


(a) RMSE of position (b) RMSE of velocity (c) RMSE of orientation



(d) RMSE of position (e) RMSE of velocity (f) RMSE of orientation

Fig. 7. Performance comparison in terms of RMSE for the UKF-PMHT of [24] 1) without modal adaptation (red solid curves) and 2) with modal adaptation (green dotted curves) for the kinematic (i.e., position, velocity, orientation) states over 100 Monte-Carlo runs. In each plot, we also include the state smoothing performance (blue dash curves) from the unscented RTS smoother before the time step $T = 101$ when the modal adaptation takes place. Top row shows the results for the case of truck, while the bottom is for tractor.



(a) Truck (b) Tractor

Fig. 8. Performance comparison in terms of RMSE for the UKF-PMHT of [24] 1) without modal adaptation (red solid curves) and 2) with modal adaptation (dotted curves) for the extent (i.e., length and width) states.

C. Quantitative Evaluation of Online Model Adaptation

To further evaluate the quantitative performance, we compute the RMSE from the 100 Monte-Carlo runs on the state estimates over all 242 time steps. Similar to the above, the online modal adaptation takes place at $T = 101$. Once the spatial model is updated, it is then used by the UKF-PMHT-based state estimation to update the vehicle state from time step 102 until the end of the trajectory. Before $T = 101$, we also include the RMSE of the smoothed state from the unscented RTS smoother.

The RMSE results for the kinematic (i.e., position, velocity, orientation) and extent (length and width) states are shown in Fig. 7 and, respectively, Fig. 8. The top row shows the results for the case of the truck, while the bottom is for the tractor. Several observations can be made here: First, the unscented RTS smoother (blue dash curves) often leads to improved state estimates than the UKF-PMHT-based state estimation before the online modal adaptation $T = 101$. This confirms the use of smoothed states, rather than updated states, to decouple the state from the online measurements and create an online batch of state-decoupled training data. Second, the UKF-PMHT-based state estimation gives much improved RMSE results after the online modal adaptation. Finally, the performance improvement between the truck and tractor cases is comparable as both updated spatial models appear to capture the true vehicle contours.

V. Conclusions

In this paper, we continued our previous effort on using a new surface-volume model to represent automotive radar measurements. Instead of the *fast-time* state update, we focus on the *slow-time* spatial model adaptation to refine the offline learned spatial model. This was achieved by performing online smoothing over the UKF-PHMT-based predicted and updated states, creating an online batch of smoothed training data, and updating the spatial modal parameters within a regularized distance to the offline learned ones. The slow-time spatial model adaptation has been numerically verified to improve the

online state estimation performance when the offline learned B-spline chained ellipses model is over-smoothed.

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] P. Broßeit, M. Rapp, N. Appenrodt, and J. Dickmann, "Probabilistic rectangular-shape estimation for extended object tracking," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, 2016, pp. 279–285.

[3] X. Cao, J. Lan, X. R. Li, and Y. Liu, "Extended object tracking using automotive radar," in *2018 FUSION*, 2018, pp. 1–5.

[4] M. Baum and U. D. Hanebeck, "Extended object tracking with random hypersurface models," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 50, no. 1, pp. 149–159, 2014.

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

[6] K. Thormann, M. Baum, and J. Honer, "Extended target tracking using Gaussian processes with high-resolution automotive radar," in *2018 FUSION*, 2018, pp. 1764–1770.

[7] H. Kaulbersch, J. Honer, and M. Baum, "A cartesian B-spline vehicle model for extended object tracking," in *2018 FUSION*, 2018, pp. 1–5.

[8] J.-L. Yang, P. Li, and H.-W. Ge, "Extended target shape estimation by fitting B-spline curve," *Journal of Applied Mathematics*, vol. 2014, 2014.

[9] Y. Guo, Y. Li, R. Tharmarasa, T. Kirubarajan, M. Efe, and B. Sarikaya, "GP-PDA filter for extended target tracking with measurement origin uncertainty," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 4, pp. 1725–1742, 2019.

[10] W. Aftab, R. Hostettler, A. De Freitas, M. Arvaneh, and L. Mihaylova, "Spatio-temporal Gaussian process models for extended and group object tracking with irregular shapes," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2137–2151, 2019.

[11] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, 2008.

[12] M. Feldmann, D. Fränken, and W. Koch, "Tracking of extended objects and group targets using random matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1409–1420, 2011.

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

[14] S. Yang and M. Baum, "Tracking the orientation and axes lengths of an elliptical extended object," *IEEE Trans. on Signal Processing*, vol. 67, no. 18, pp. 4720–4729, 2019.

[15] G. Yao, R. Saltus, and A. Dani, "Image moment-based extended object tracking for complex motions," *IEEE Sensors Journal*, vol. 20, no. 12, pp. 6560–6572, 2020.

[16] K. Wyffels and M. Campbell, "Precision tracking via joint detailed shape estimation of arbitrary extended objects," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 313–332, 2016.

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

[18] A. Scheel and K. Dietmayer, "Tracking multiple vehicles using a variational radar model," *IEEE Transactions on Intelligent Transportation Systems*, 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 *2019 FUSION*, 2019, pp. 1–8.

[20] Y. Xia, P. Wang, K. Berntorp, T. Koike-Akino, H. Mansour, M. Pajovic, P. Boufounos, and P. V. Orlik, "Extended object tracking using hierarchical truncation measurement model with automotive radar," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 4900–4904.

[21] Y. Xia, P. Wang, K. Berntorp, H. Mansour, P. Boufounos, and P. V. Orlik, "Extended object tracking using hierarchical truncation model with partial-view measurements," in *2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2020, pp. 1–5.

[22] Y. Xia, P. Wang, K. Berntorp, P. Boufounos, P. Orlik, L. Svensson, and K. Granström, "Extended object tracking with automotive radar using learned structural measurement model," in *2020 IEEE Radar Conference (RadarConf)*, 2020, pp. 1–6.

[23] Y. Xia, P. Wang, K. Berntorp, L. Svensson, K. Granström, H. Mansour, P. Boufounos, and P. V. Orlik, "Learning-based extended object tracking using hierarchical truncation measurement model with automotive radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 1013–1029, 2021.

[24] G. Yao, P. Wang, K. Berntorp, H. Mansour, P. Boufounos, and P. V. Orlik, "Extended object tracking with automotive radar using B-Spline chained ellipses model," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8408–8412.

[25] C. de Boor, *A practical guide to splines*. Springer-Verlag New York, 1978, vol. 27.

[26] X. Rong Li and V. P. Jilkov, "Survey of maneuvering target tracking. part I. dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, 2003.

[27] M. Wieneke and W. Koch, "A PMHT approach for extended objects and object groups," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 3, pp. 2349–2370, 2012.

[28] K. Granstrom and J. Bramstang, "Bayesian smoothing for the extended object random matrix model," *IEEE Transactions on Signal Processing*, vol. 67, no. 14, pp. 3732–3742, Jul 2019.

[29] S. Särkkä, "Unscented Rauch–Tung–Striebel smoother," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 845–849, 2008.

[30] ——, *Bayesian Filtering and Smoothing*. USA: Cambridge University Press, 2013.