A multiple model PHD approach to tracking of cars under an assumed rectangular shape

Karl Granström
Department of Electrical Engineering
Linköping University,
Linköping, Sweden
Email: karl@isy.liu.se

Stephan Reuter, Daniel Meissner, and Alexander Scheel
Institute of Measurement, Control and Microtechnology,
Ulm University, Ulm, Germany
Email: stephan.reuter@uni-ulm.de,
daniel.meissner@uni-ulm.de, alexander.scheel@uni-ulm.de

***Abstract*—This paper presents an extended target tracking method for tracking cars in urban traffic using data from laser range sensors. Results are presented for three real world datasets that contain multiple cars, occlusions, and maneuver changes. The car's shape is approximated by a rectangle, and single track steering models are used for the target kinematics. A multiple model approach is taken for both the dynamics modeling and the measurement modeling. A comparison to ground truth shows that the estimation errors are generally very small: on average the absolute error is less than half a degree for the heading. Multiple cars are handled using a multiple model PHD filter, where a variable probability of detection is integrated to enable tracking of occluded cars.**

I. Introduction

Multiple target tracking can be defined as the processing of multiple measurements obtained from multiple targets in order to maintain estimates of the targets' current states, see e.g. [1]. In this context, a point target is defined as a target which is assumed to give rise to at most one measurement per time step, and an extended target is defined as a target that potentially gives rise to more than one measurement per time step.

For single extended target tracking spatial distribution models appeared in [2], [3]. Under this model type the extended target measurements are random samples of a distribution that is dependent on the extended target state. A number of different extended target measurement models that fall into this category have been presented. A popular choice is to assume a specific geometric shape for the target, e.g. a stick [3]–[6], a circle [7], an ellipse [8]–[14], or a rectangle [10]. General shapes, i.e. ones that cannot be described by a specific geometric shape, can also be estimated, see e.g. [15]–[17].

For multiple target tracking finite set statistics (FISST) represent a rigorous approach, see [18]. It has led to the point target probability hypothesis density (PHD) filter [19], with its Gaussian mixture (GM) implementation [20]. An extension of the PHD filter to handle extended targets with spatial distributions of the type presented in [2] is given in [21], with implementations in [22]–[24].

Target tracking approaches are typically cast as prediction and measurement update recursions, and as such rely heavily on motion modeling and measurement modeling. In many practical scenarios it is difficult to capture all target dynamics and measurement characteristics in a single model pair. For such cases multiple models (MM), also called jump Markov system models, can be used. The interactive multiple model (IMM) algorithm [25] represents a good trade-off between computational complexity and tracking performance, and has been shown to perform well for maneuvering single point target tracking.

The use of multiple models can be integrated into a multiple target tracking frame work. An overview of MM-PHD filters for point targets is given in [26], and [27] is pointed out as the preferred MM-PHD approach. An MM-PHD filter for extended targets is presented in [6].

In this paper we consider tracking multiple cars in urban and rural scenarios using laser range sensors, both for stationary and for moving sensors. Laser range sensors have been used extensively for mapping and localization in field robotics for the past decade, and are attracting increasing attention in the automotive industry. The laser sensors emit light along pre-defined bearings and measure the range to the nearest reflecting point along the bearing. The geometric appearance of the measurements can change rapidly, mainly because of the cars' movements but also due to the complicated reflectance properties of the sensors. It is therefore necessary to use multiple measurement models. Further, multiple motion models are necessary to model, e.g., cars that may either drive straight ahead or turn at an intersection. To handle tracking of multiple cars an MM-PHD filter [6] is used.

II. Multiple model framework for car tracking using laser data

The notation used in this paper is given in Table I. The kinematic state vector $\mathbf{x}_k$ is defined as

$$\mathbf{x}_k = \left[\mathsf{x}_k,\ \mathsf{y}_k,\ \mathsf{v}_k,\ \varphi_k,\ \theta_k,\ \ell_k,\ w_k\right]^{\mathrm{T}}, \tag{1}$$

where $(\mathsf{x}_k, \mathsf{y}_k)$ is the position in Cartesian coordinates of the center of the rear-axle, v_k is the speed, φ_k is the heading, θ_k is the steering angle, ℓ_k is the length and w_k is the width of the car. The length of the wheel base, i.e. from the rear axle to the front axle, is assumed to be $\ell_k^w = 0.7\ell_k$. An example illustration of the kinematic state is given in Figure 1.

We consider two different motion modes and two different measurement modes. The motion modes are for the two cases

TABLE I
NOTATION

- $\mathbb{R}^n$ is the set of real n-vectors, $\mathbb{S}^n_+$ is the set of symmetric positive semi-definite $n \times n$-matrices, and $\mathbb{N}_+$ is the set of natural numbers.
- At discrete time t_k, $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the extended target kinematic state, $o_k \in M \subset \mathbb{N}_+$ is the extended target mode, and $\xi_k = (\mathbf{x}_k\ ,\ o_k) \in \mathcal{X}_0 = \mathbb{R}^{n_x} \times M$ is the augmented extended target state.
- $\mathcal{N}(x\,;\, m, P)$ denotes the probability density function (pdf) of a Gaussian distribution defined over the vector x, with mean vector $m \in \mathbb{R}^n$ and covariance matrix $P \in \mathbb{S}^n_+$. The short hand notation $\mathcal{N}(x, \Theta)$ is also used, where $\Theta = (m, P)$.
- $\mathbf{Z}_k = \left\{\mathbf{z}_k^{(j)}\right\}_{j=1}^{N_{z,k}}$ is a measurement set at time t_k, $\mathbf{z}_k^{(j)} \in \mathbb{R}^{n_z}$, $\forall j$.
- $\mathrm{p} \angle \mathbf{Z}$ denotes all the partitions p of the set $\mathbf{Z}$. A partition p is a set of non-empty subsets W called cells. The union of all cells W is equal to the set $\mathbf{Z}$. The cardinality of a cell W is denoted $|W|$.
- $\delta_{i,j}$ is the Kronecker delta, and $\otimes$ is the Kronecker product.
- $f[g(x)]$ denotes the integral $\int f(x)g(x)dx$.
- $\mathbf{I}_n$ is a $n \times n$ identity matrix and $\mathbf{0}_n$ is a $n \times n$ all zero matrix.

where the steering angle θ_k is constant or zero. These two motion models are in a way analogous to constant turn-rate and constant velocity models, and analogously we refer to them as constant steering (CS) and constant velocity (CV). The measurement modes are for the two alternatives that the sensor detections are from either one side of the car, or from two sides. This gives four modes in total ($M = \{1, 2, 3, 4\}$),

1) CS motion, one sided measurements (CS1),
2) CV motion, one sided measurements (CV1),
3) CS motion, two sided measurements (CS2),
4) CV motion, two sided measurements (CV2).

A. Mode transitions

The modes follow a discrete Markov chain, where the transition probability matrix is decomposed as follows

$$T_{k+1|k} = T^{\text{Meas}}_{k+1|k} \otimes T^{\text{Motion}}_{k+1|k}, \tag{2a}$$

$$T^{\text{Meas}}_{k+1|k} = \begin{bmatrix} P^{\text{meas}}_{\text{same}} & 1 - P^{\text{meas}}_{\text{same}} \\ 1 - P^{\text{meas}}_{\text{same}} & P^{\text{meas}}_{\text{same}} \end{bmatrix}, \tag{2b}$$

$$T^{\text{Motion}}_{k+1|k} = \begin{bmatrix} P^{\text{mot}}_{\text{same}} & 1 - P^{\text{mot}}_{\text{same}} \\ 1 - P^{\text{mot}}_{\text{same}} & P^{\text{mot}}_{\text{same}} \end{bmatrix}. \tag{2c}$$

The probabilities of staying in the same measurement mode and motion mode are given by $P^{\text{meas}}_{\text{same}}$ and $P^{\text{mot}}_{\text{same}}$, respectively.

In previous work the measurement mode transition probability has been modeled as a function of the kinematic state [6]. Making this choice for car tracking using laser data can seem intutitive: the number of sides that can be seen from the sensor depend on the position and heading of the car. However, in this work we found that this type of modeling does not improve results. The main reason is that the reflectance properties of the cars make it much more difficult to predict how many sides of the car the sensor should be seeing, i.e. the mode transitions depend on factors other than the kinematic state. Because these factors are not modeled, we found it is a better choice to model the measurement mode transitions as independent of the kinematic state.

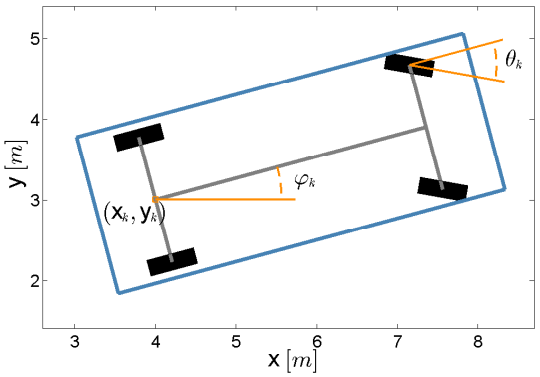

Fig. 1. Example of kinematic state vector $\mathbf{x}_k$, where $(\mathsf{x}_k, \mathsf{y}_k)$ is the position in Cartesian coordinates of the center of the rear-axle, φ_k is the heading, and θ_k is the steering angle. The speed v_k, the length ℓ_k, and the width w_k, are not illustrated. The length of the wheel base, i.e. from the rear axle to the front axle, is assumed to be $\ell^w_k = 0.7\ell_k$.

B. Kinematic state motion models

The time evolution of the kinematic state is modeled as

$$p\left(\mathbf{x}_k \,|o_k, \mathbf{x}_{k-1}\right) = \mathcal{N}\left(\mathbf{x}_k\,;\, f\left(\mathbf{x}_{k-1}, o_k\right), Q_k(o_k)\right). \tag{3}$$

where $f(\cdot) : \mathcal{X}_0 \to \mathbb{R}^{n_x}$ is a state space motion model corresponding to either CS motion or CV motion. For the four modes listed above we have $f\left(\mathbf{x}_k, o_{k+1} = \{1, 3\}\right) = f_{\text{CS}}\left(\mathbf{x}_k\right)$ and $f\left(\mathbf{x}_k, o_{k+1} = \{2, 4\}\right) = f_{\text{CV}}\left(\mathbf{x}_k\right)$.

1) Constant steering: For a non-zero steering angle θ_k, the distance traveled is $d_k = Tv_k$, the turning angle is $\beta_k = \frac{d_k}{\ell^w_k}\tan(\theta_k)$, and the turning radius is $R_k = d_k/\beta_k$. The constant steering angle motion model, see e.g. [28], is

$$f_{\text{CS}}\left(\mathbf{x}_k\right) = \begin{bmatrix} \mathsf{x}_k - R_k\sin(\varphi_k) + R_k\sin(\varphi_k + \beta_k) \\ \mathsf{y}_k + R_k\cos(\varphi_k) + R_k\cos(\varphi_k + \beta_k) \\ \left[\mathsf{v}_k,\ \varphi_k + \beta_k,\ \theta_k,\ \ell_k,\ w_k\right]^{\mathrm{T}} \end{bmatrix}. \tag{4}$$

This motion model approximates the car as having a single front wheel and a single rear wheel (analogously to a bicycle), where the rear wheel follows along a circle with radius R_k.

If the steering angle approaches zero the following result,

$$\lim_{\theta_k \to 0} \beta_k = 0, \qquad \lim_{\theta_k \to 0} R_k = \infty, \tag{5}$$

holds for the turning angle and the turning radius. Using Taylor expansion it is simple to show that for a steering angle θ_k close to zero the motion models reduce to a standard constant velocity motion model,

$$f_{\text{CS}}\left(\mathbf{x}_k\right) = \begin{bmatrix} \mathsf{x}_k + d_k\cos(\varphi_k) \\ \mathsf{y}_k + d_k\sin(\varphi_k) \\ \left[\mathsf{v}_k,\ \varphi_k,\ \theta_k,\ \ell_k,\ w_k\right]^{\mathrm{T}} \end{bmatrix}. \tag{6}$$

In the implemented tracking filters we use this model if $|\hat{\theta}_{k|k}| < 10^{-9}$ degrees, where $\hat{\theta}_{k|k}$ is the estimated steering angle.

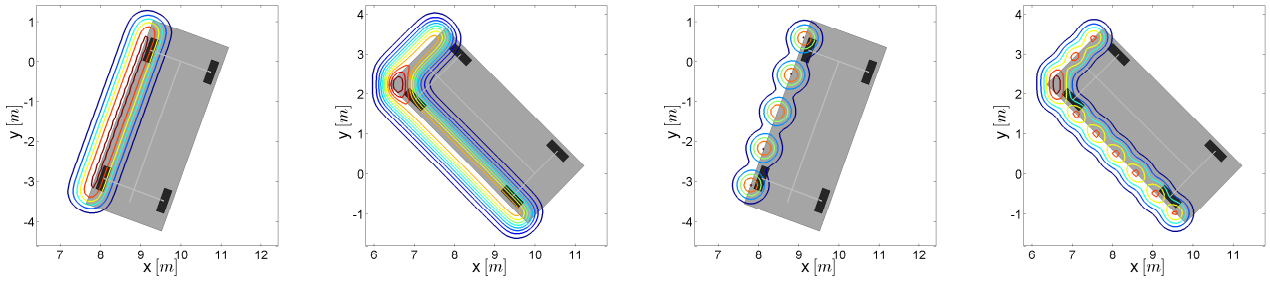

Fig. 2. Rectangular target laser sensor measurement likelihood. Sensor located in origin, directed along positive x-axis. Left: measurement likelihood when one and two sides are seen by sensor, respectively. Right: corresponding Gaussian mixture approximations with five and ten components, respectively.

2) Constant velocity: In the CV mode the steering angle is exactly zero, and the corresponding state could be removed from the kinematic state vector. However, to avoid having kinematic vectors of different lengths in the different modes, we instead estimate the steering angle and use an exponential decay model for the time evolution,

$$p(\theta_{k+1}|\theta_k) = \mathcal{N}\left(\theta_{k+1}\,;\, e^{-0.5}\theta_k, q_{k+1}\right). \tag{7}$$

In other words: in the CV mode we use the same motion models as in (4) and (6), with the exception that θ_k is replaced by $e^{-0.5}\theta_k$. This is not an exactly zero steering angle motion model, however it is close and empirically we have found that this model is sufficient.

C. Measurement model

In this section we drop time indexing (subindex $_k$) to avoid the notation becoming too cluttered. Let $\mathbf{W} = \left\{\mathbf{z}^{(j)}\right\}_{i=1}^{|\mathbf{W}|}$ be a set of measurements, where all measurements originate from the same target. As is common in much extended target tracking, we assume that the measurements are independent of each other, which gives the measurement likelihood

$$p\left(\mathbf{W}\,|\xi\right) = \prod_{j=1}^{|\mathbf{W}|} p\left(\mathbf{z}^{(j)}\middle|\,\xi\right). \tag{8}$$

A planar laser range sensor sweeps counter clockwise through the surveillance area, measuring the range $r_k^{(j)}$ to the closest object for a set of bearings $\alpha_k^{(j)}$. A target generated measurement $\mathbf{z}$ can be seen as a random measurement generating point $\mathbf{y}$ that is measured with some noise $\mathbf{e}$. The measurement likelihood $p(\mathbf{z}|\xi)$ is then given by the convolution

$$p(\mathbf{z}|\xi) = \int p(\mathbf{z}|\mathbf{y})p(\mathbf{y}|\xi)\mathrm{d}\mathbf{y} \tag{9}$$

As an example, let the target be a $2D$ stick located in the origin and aligned with the x-axis. In this case the measurement generating points $\mathbf{y}$ can be modeled as uniformly distributed along the stick, and the noise $\mathbf{e}$ can be modeled as additive zero mean Gaussian distributed. The pdf of a target generated measurement $\mathbf{z} = [z_\mathrm{x},\ z_\mathrm{y}]^\mathrm{T}$ is (see e.g. [3, Eq. 19])

$$\begin{aligned} p(\mathbf{z}|\xi) =& \frac{1}{2\ell\sqrt{2\pi|\Sigma|}}\exp\left(-\frac{1}{2}\left(\frac{z_\mathrm{y}}{\sigma_\mathrm{across}}\right)^2\right) \\ &\times\left(\mathrm{erf}\left(\frac{z_\mathrm{x}+\frac{1}{2}\ell}{\sqrt{2}\sigma_\mathrm{along}}\right) - \mathrm{erf}\left(\frac{z_\mathrm{x}-\frac{1}{2}\ell}{\sqrt{2}\sigma_\mathrm{along}}\right)\right). \end{aligned} \tag{10}$$

Here $\Sigma = \mathrm{diag}\left([\sigma^2_\mathrm{along},\ \sigma^2_\mathrm{across}]\right)$ denotes the noise variance along the length of the stick, and across the length. If the stick target is not located in the origin, and/or is not aligned with the x-axis, the measurement likelihood is obtained by rotation and translation of (10).

In the case of a rectangular target, e.g. a car as shown in Figure 1, the sensor can see either one or two of the target's four sides (left, rear, right, front). The measurement generating points $\mathbf{y}$ are then uniformly distributed along either one or two of the sides. Two example measurement likelihoods are illustrated in the left part of Figure 2.

In this work the measurement likelihood (10) is approximated by a Gaussian mixture,

$$p(\mathbf{z}|\xi) \approx \sum_{i=1}^{N} w^{(i)}\mathcal{N}\left(\mathbf{z}\,;\,\mathbf{y}^{(i)}(\xi), R^{(i)}\right),\ \sum_{i=1}^{N} w^{(i)} = 1. \tag{11}$$

Here the measurement generating points $\mathbf{y}^{(i)}(\xi)$ are non-linear functions $\mathbf{y}^{(i)}(\cdot) : \mathcal{X}_0 \to \mathbb{R}^{n_z}$. For a given rectangular target estimate, computing predicted locations of measurement generating points is straightforward using some simple geometry and trigonometry.

Under the Gaussian mixture approximation, we assume that each measurement was generated by exactly one measurement generating point. Two approximation examples, corresponding to Figures 2a and 2b, are shown in Figures 2c and 2d. Note that as the number of Gaussians in the mixture approaches infinity, the approximation becomes exact.

The main motivation for the Gaussian mixture approximation is that it enables use of standard estimation techniques for non-linear Gaussian models, e.g. the Extended Kalman filter or the Unscented Kalman filter. However, this approximation introduces an association problem: given a set of $N_{z,k}$ measurements and a target estimate, predicted measurement generating points $\mathbf{y}^{(i)}(\xi)$ must be computed and associated to

the measurements. The association problem can be alleviated using some insight into how laser range sensors work. The sensors used in this work sweep the surveillance area counter clockwise, and the measurement can thus be sorted according to their bearing. By sorting the predicted measurement generating points similarly, an association is implicitly given.

The likelihood is then

$$p(\mathbf{W}|\xi) = \prod_{j=1}^{|\mathbf{W}|} \mathcal{N}\left(\mathbf{z}_k^{(j)}; \mathbf{y}^{(j)}(\xi), \sigma_r^2 I_2\right) \tag{12a}$$

$$= \mathcal{N}\left(\mathbf{z}_{\mathbf{W}}; \mathbf{y}_{\mathbf{W}}(\xi), \sigma_r^2 I_{2|W|}\right), \tag{12b}$$

where $\mathbf{z}_{\mathbf{W}}$ and $\mathbf{y}_{\mathbf{W}}(\mathbf{x})$ are vertical vectorial concatenations of the measurements and measurement generating points. In the implementation an Extended Kalman filter is used, and the Jacobian of the measurement equation is computed numerically.

The measurement models used in this work are extensions of the measurement models for rectangular targets that were presented in [10], where simulated data was used for performance evaluation. In this work the models from [10] have been adapted and improved to better handle real-world data. One very important laser sensor property that is not simulated in [10] are the reflectances properties of the cars. In real data it can often be observed that, despite that the sensor should be able to see, e.g., the entire left side of a car, only a smaller part of the left side yields measurements. The main reason for this is that the car did not reflect enough of the emitted laser light. Handling this in the measurement models is important, especially for the estimation of the heading, but also the length and width of the car.

As noted above two measurement models are used, corresponding to the two cases in Figure 2 with either one sided measurements or two sided measurements. In what follows, it is assumed that the set of measurements $\mathbf{W}$ has been sorted according to bearing.

1) One sided measurements: Let $\psi_1, \ldots, \psi_4$ be the surface normals of the estimated target shape, and define β as the angle of the vector from the first measurement to the last measurement in the set. The measurements are associated to the $i_{\min}$th side,

$$i_{\min} = \arg\min_i \ |\psi_i - \beta + \pi/2| \,. \tag{13}$$

Let $\ell_{i_{\min}}$ be the length of the side, let $\ell_{\mathbf{z}} = \left\|\mathbf{z}^{(1)} - \mathbf{z}^{(|\mathbf{W}|)}\right\|_2$ be the length of the measurement segment, and let $\rho = \ell_{\mathbf{z}}/\ell_{i_{\min}}$ be the ratio of the two lengths. If $\rho \geq 0.5$ then the entire side is seen and $|\mathbf{W}|$ measurement generating points $\mathbf{y}^{(i)}(\mathbf{x})$ are spread uniformly along the entire $i_{\min}$th side. If $\rho < 0.5$ then only parts of the side is seen, and in this case $|\mathbf{W}|$ measurement generating points are spread uniformly along a part of the side such that $100\rho\%$ of the side is covered.

At an early stage it was considered to include more measurement models to handle the whole-side/part-of-side ambiguity. However, this approach increases the computational complexity and the above threshold was considered the better choice because it achieved the same result at lower computational complexity. The measurement $\mathbf{z}^{(i)}$ to measurement generating point $\mathbf{y}^{(i)}$ association is simply $i \leftrightarrow i$, for $i = 1, \ldots, N_z$.

2) Two sided measurements: In the two sided model the measurement set $\mathbf{W}$ is split in two parts $\mathbf{W}_1 = \left\{\mathbf{z}^{(j)}\right\}_{i=1}^{n}$ and $\mathbf{W}_2 = \left\{\mathbf{z}^{(j)}\right\}_{i=n+1}^{|\mathbf{W}|}$. This is performed by least squares fitting lines to the two segments for each $n \in [2, \ldots, |\mathbf{W}|-1]$. The n for which the least squares error is smallest defines the split. For each of the two segments the one sided measurement model is applied, where the two segments are associated to the $i^1_{\min}$th and $i^2_{\min}$th sides. Note that care is taken to ensure that $i^1_{\min} \neq i^2_{\min}$, and to ensure that the two sides are adjacent.

D. MM-PHD filter for extended targets

The multiple motion and measurement models were integrated into a version of the Gaussian Mixture (GM) MM-PHD filter presented in [6]. The PHD has the following GM representation

$$D_{k|k}(\xi) = \sum_{j=1}^{J_{k|k}(o)} w_{k|k}^{(j)}(o) \mathcal{N}\left(\mathbf{x}; \Theta_{k|k}^{(j)}(o)\right). \tag{14}$$

The birth PHD is also assumed to be a GM,

$$D_k^b(\xi) = \pi_k(o) \sum_{j=1}^{J_k(o)} w_{b,k}^{(j)}(o) \mathcal{N}\left(\mathbf{x}; \Theta_{b,k}^{(j)}(o)\right). \tag{15}$$

The Gaussian parameters $\Theta_{b,k}^{(j)}(o)$ are set such that they represent the locations where targets are likely to appear, as well as likely heading and size. The prediction and correction updates are given in [6, Section 3] and are not repeated here.

For simplicity the probability of detection is assumed constant in [6]. Laser range sensors are susceptible to occlusion and therefore the probability of detection cannot be assumed to be constant, it is instead modeled as a function of the kinematical state. For the probability of detection for a target we assume the following to hold for all components j,

$$P_D(\mathbf{x}_k|o_k)\mathcal{N}\left(\mathbf{x}_k; \Theta_{k|k-1}^{(j)}(o_k)\right) \tag{16}$$

$$\approx P_D\left(\Theta_{k|k}^{(j)}(o_k)\middle| o_k\right)\mathcal{N}\left(\mathbf{x}_k; \Theta_{k|k}^{(j)}(o_k)\right)$$

The probability of detection for the jth component is abbreviated as $P_D^{(j)} = P_D\left(\Theta_{k|k}^{(j)}(o_k)\middle| o_k\right)$. Trivially this assumption holds if the probability of detection is independent of the kinematic state. In general the assumption holds when the probability of detection does not vary much in the uncertainty zone of the kinematic state space, determined by $P_{k|k}^{(j)}(o_k)$. This is true either when $P_D(\mathbf{x}_k|o_k)$ is a sufficiently smooth function w.r.t $\mathbf{x}_k$, or when the uncertainty zone is sufficiently small. In the next section we give an occlusion model that is used to estimate the probability of detection for the target estimates.

TABLE II
OCCLUSION MODEL

Input: Point of interest x, y, set of kinematic state estimates $\{\hat{\mathbf{x}}^{(i)}\}_{i=1}^{N_t}$ with weights $w^{(i)}$.
Intialize probability of detection $P_D \leftarrow P_D^0$.
Compute range and bearing to point of interest: r, α.
for $i = 1, \ldots, N_t$ **do**
    Range/bearing to corners of estimate:
$$\hat{r}_j^{(i)}, \hat{\alpha}_j^{(i)}, \; j = 1, 2, 3, 4$$
    Minimum/maximum bearings:
$$j_- = \arg\min_j \hat{\alpha}_j^{(i)}, \; j_+ = \arg\max_j \hat{\alpha}_j^{(i)}$$
    Mean range:
$$r_-^+ = 0.5(\hat{r}_{j_+}^{(i)} + \hat{r}_{j_-}^{(i)})$$
    if $r \geq r_-^+ \wedge \hat{\alpha}_{j_-}^{(i)} \leq \alpha \leq \hat{\alpha}_{j_+}^{(i)}$ **then**
        Decrease P_D when point of interest is located behind estimate:
$$P_D \leftarrow P_D - (1 - g_- - g_+)\, w^{(i)}$$
$$g_- = \exp(-((\alpha - \hat{\alpha}_{j_-}^{(i)})/\sigma_\alpha)^2)$$
$$g_+ = \exp(-((\alpha - \hat{\alpha}_{j_+}^{(i)})/\sigma_\alpha)^2)$$
    end if
end for
Enforce minimum probability of detection: $P_D \leftarrow \max\{P_D, \; P_D^{\min}\}$
Output: Probability of detection P_D at point of interest.

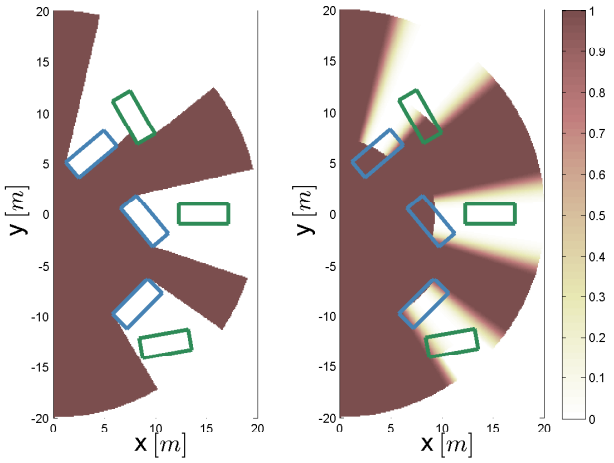

Fig. 3. Target occlusion. (a) Ground truth probability of detection for six rectangular targets. (b) Modeled probability of detection for six rectangular target estimates. For the occluded cars (green), the probability of detection is estimated to 99% (top), 1% (middle), and 50% (bottom).

III. OCCLUSION MODEL

An inherent property of many sensor types, e.g. laser, radar and video, is that the sensors are subject to occlusions. Simply put, if two targets are located along approximately the same sensor to target bearing, the target that is more distant from the sensor will be either fully or partially occluded. In an urban scenario with multiple targets, it is likely that one or more targets will be occluded to some degree. It is therefore important that a multiple extended target filter is capable of handling occlusions.

In previous work occlusion has successfully been handled by modeling the probability of detection as non-homogeneous, see e.g. [11], [23], [24]. The occlusion model that was used in this work is based on similar ideas. In Table II pseudo code is given for the probability of detection at a point of interest inside the surveillance area. The basic intuition in the model is to lower the probability of detection if the point is located behind a target estimate. The decrease is proportional to the weight of the estimate, i.e. proportional to how certain the filter is that there is actually a target there. Furthermore Gaussian kernels g_- and g_+ are used to give a smooth transition from visible area to occluded area. In the implementation we use $\sigma_\alpha = 2.5$ degrees. Note that the edges are smoothed only on the "inside" of the occluded area, this gives a conservative estimate of occlusion.

The probability of detection for the jth component $P_D^{(j)}$ is computed by first discretizing the rectangular shape into points separated by 10 cm. Using the method in Table II the probability of detection is then computed for each point along the shape, and $P_D^{(j)}$ is computed as the average of the ten points with highest probability of detection.

A simple example with six targets is given in Figure 3, where three of the cars are at least partially occluded (green rectangles). The top example is mostly occluded, but a considerable part of the car can be seen and $P_D^{(j)} = P_D^0 = 99\%$. The middle example is fully occluded and $P_D^{(j)} = P_D^{\min} = 1\%$. The bottom example has one of its corners on the edge of the visible/occluded are, and $P_D^{(j)} = 50\%$. Remaining three cars in Figure 3 (blue rectangles) are not occluded, and $P_D^{(j)} = 99\%$.

IV. EXPERIMENTS

A. Datasets

Three different datasets are used: one single target dataset, and two multiple target datasets. In the multiple target datasets measurements of the stationary background were removed beforehand using background subtraction, however the clutter measurements were not removed. The laser range sensor that was used to collect the first two datasets is an Ibeo Lux sensor, the third dataset was collected using a Sick LMS sensor. The Ibeo sensor has sample time 0.08 seconds, the Sick sensor has 0.20 seconds.

1) D1: In the single target dataset (D1) the sensor was mounted on a car that drove behind another car on a rural road. Both the ego-vehicle and the leading vehicle were equipped with a differential GPS sensor and an additional IMU sensor (GPS+IMU), and the ground truth data was used to extract the target generated measurements (i.e. there are no measurements of background or clutter). In this data only the rear side of the leading vehicle is visible to the sensor, neither the left nor the right side is ever fully visible.

2) D2: In the first multi target dataset (D2) the sensor was stationary next to a roundabout. Multiple cars drove through the roundabout, for one of these cars GPS+IMU data is available. The GPS+IMU equipped target circled the roundabout three times. Because of stationary structures in the center of the roundabout the targets were occluded from sensor view in parts of the roundabout. In this dataset all the targets move

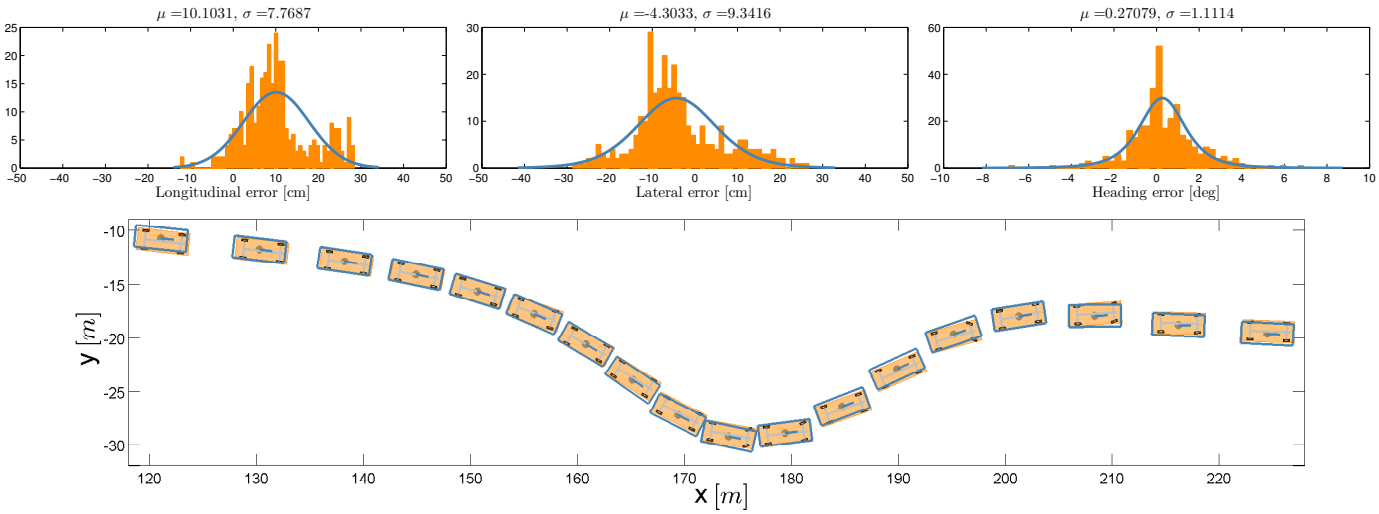

Fig. 4. Results for dataset D1. Top row: Histograms with estimation errors for longitudinal position, lateral position, and heading φ. Also shown are distributions fitted to the error data, with means μ and standard deviations σ. Bottom row: example tracking results, plotted every ninth time step for increased clarity. Estimates in orange, ground truth in blue.

such that at least three of the four sides are visible at some point, and there are also instances of full and partial target occlusion.

3) D3: In the second multi target dataset (D3) the sensor was stationary next to a T-intersection. Four cars passed through the surveillance area, however there were no occlusions. For this dataset there is unfortunately no ground truth data available.

B. Performance evaluation

For D1 and D2 we compare the tracking results to the GPS+IMU ground truth. Instead of showing the x− and y−position errors, we show the longitudinal and lateral position errors. This corresponds to the error in position along the heading of the vehicle and the error in position orthogonal to the heading of the vehicle. Gaussian distributions were fitted to the error histograms, and the means and standard deviations are given. For D3 there is unfortunately no ground truth data. To validate the estimation results we plot them on top of an aerial image of the location.

C. Results

1) D1: For this data set the estimate was initialized using ground truth data, which means that there is little convergence error. The estimation errors and example tracking results are shown in Figure 4. For this dataset the errors are quite small: longitudinal position 10 ± 8cm; lateral position -4 ± 9cm; heading 0.27 ± 1.11 degrees.

The length and width errors are both very small (below one cm), and the error histograms are therefore not visualized. Mainly the very low errors are due to the initialization using ground truth. The full length is never observed in the sensor data, and thus the initial length estimate is not updated. The full width is observed in each time step, however since the initial estimate fits the data very well, only very small updates are made to the width. Subsequently, both length and width errors are very small.

In reality it naturally not possible to initialize target estimates using ground truth, because the ground truth is inherently unknown in target tracking. Cars do come in a limited range of sizes though. Assuming that the initialized target estimate is for a car, in most practical cases it is possible to initialize the length and width reasonably close to the ground truth. In case several target types are present in the scene, e.g. bicycles and pedestrians in addition to the cars, some kind of object classification capability is necessary.

2) D2: This dataset contained measurements from multiple cars, however since one of the cars was equipped with GPS+IMU we will emphasize the results for this car. For D2 the estimates were not initialized using GPS+IMU, and the initialization errors are thus larger than for D1. The largest errors can be observed just after the targets have entered the view of the sensor, i.e. before the estimates have converged. Typically the estimate converges in about five time steps (0.4 seconds), at most it takes up to ten time steps (0.8 seconds).

Estimation errors and example tracking results are shown in Figure 5. For D1 the errors are generally larger than for D2: longitudinal position 28 ± 9cm; lateral position 4cm $\pm$13cm; heading 0.16 ± 1.18 degrees; length -18 ± 18cm; width -4 ± 6cm. For longitudinal position and length the errors are significantly non-zero. The longitudinal position appears to have a positive bias, which in part is due to the length error being transferred to the longitudinal position.

However, when the sensor data and the GPS+IMU ground truth was carefully investigated, another reason for the longitudinal position bias was also found. The bias is partly caused by the fact that the rear of the car's ground truth rectangle is not quite aligned with the measurements that originated from

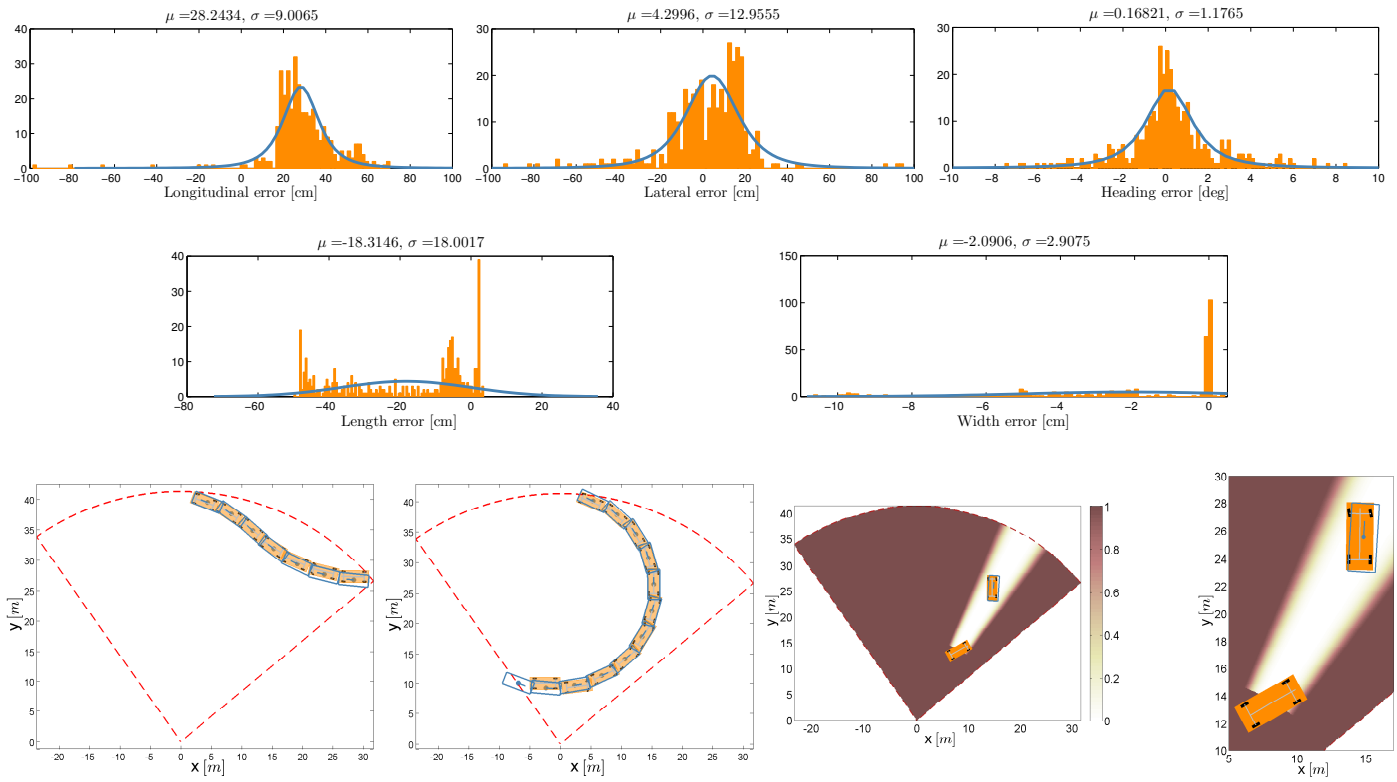

Fig. 5. Results for dataset D2. Top and middle row: Histograms with estimation errors for longitudinal position, lateral position, heading φ, length and width. Also shown are distributions fitted to the error data, with means μ and standard deviations σ. Bottom row, left: two example tracking results, plotted every ninth time step for increased clarity. Estimate in orange, ground truth in blue. Bottom row, right: tracking snapshot where the GPS+IMU equipped car is occluded by another vehicle. The occlusion model correctly estimates a low probability of detection (9%) for the car. As can be seen in the zoomed detail (far right), the difference between the estimate (orange) and the ground truth (blue) is very small.

the rear of the car. Because the measurement model assumes the measurements to originate from points on the surface of the car, the estimate is aligned with the measurements. This introduces a small bias for the longitudinal position.

Noteworthy is that, despite the fact that the GPS+IMU equipped car was fully occluded for several time steps, the estimation errors did not increase during these time steps. This can be attributed to the use of an occlusion model and the use of motion models that can simulate the motion of a car with high accuracy.

3) D3: The sensor data is shown Figure 6 (left), where the detections are color coded to show the different time steps. The estimates of the center of mass positions are shown in Figure 6 (middle), and the estimated headings are shown in Figure 6 (right). The first target (color coded as blue) entered the surveillance area to the left, the other three targets (color coded orange, red and green) entered the surveillance area at the top. Note that the cars had driven a couple of meters into the surveillance area before any measurements were received by the sensor, hence the tracks do not start at the edge of the surveillance area.

For the three targets entering at the top there were quite few (about 2 to 4) measurements per time step, causing the initial target estimates to be quite uncertain. This can be seen in Figure 6 (right), where the heading estimate for the green target changes rapidly in the beginning. However, as soon as the targets generate more measurements the estimates become much more certain.

V. Conclusion and Future Work

In this paper we presented a multiple model approach for tracking cars in urban environments. The car shape was approximated by a rectangle, and two measurement models were used corresponding to measurements from one or two sides. For motion modeling a single track steering model was used. The models were integrated into an extended target MM-PHD filter and were evaluated using real world laser range data acquired using both stationary and moving sensors.

In a comparison to ground truth data it was shown that quite small estimation errors could be achieved. Noteworthy is that the lateral position error is on average only 5 cm, and the heading error is on average less than one degree. The longitudinal position error is about 30 cm for one of the dataset, but significantly lower for another dataset. An occlusion model was integrated into the MM-PHD filter and it was shown that cars can be tracked with high accuracy as they move through parts of the surveillance area that are occluded by other cars.

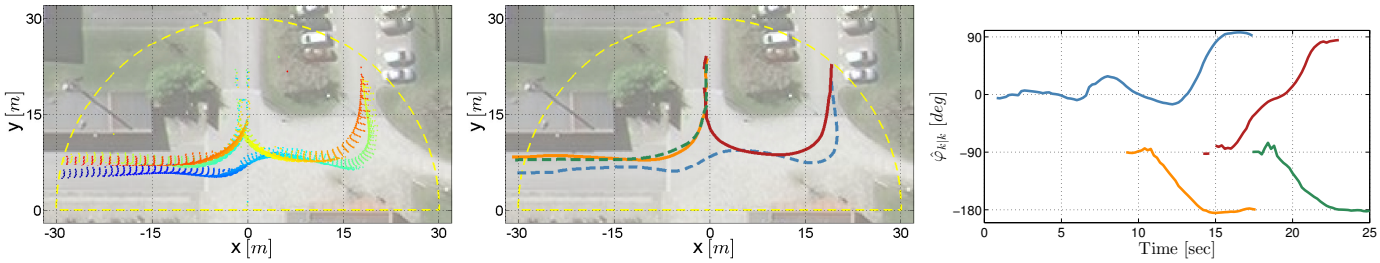

Fig. 6. Results for dataset D3. Left: Measurements overlaid on aerial image, color coded according to time (blue–early, green–intermediate, red–late). Middle: Positions of center of mass. Right: Heading estimates.

Future work includes integrating the presented work such that multiple target types can be tracked simultaneously, e.g. cars, bicycles and pedestrians in an urban environment.

Acknowledgment

Karl Granström would like to thank the project Collaborative Unmanned Aircraft Systems (CUAS), funded by the Swedish Foundation for Strategic Research (SSF), for financial support. Stephan Reuter would like to thank the Transregional Collaborative Research Center SFB/TRR 62 Companion-Technology for Cognitive Technical Systems funded by the German Research Foundation (DFG) for financial support.

References

[1] Y. Bar-Shalom and T. E. Fortmann, *Tracking and data association*, ser. Mathematics in Science and Engineering. San Diego, CA, USA: Academic Press Professional, Inc., 1987, vol. 179.

[2] K. Gilholm, S. Godsill, S. Maskell, and D. Salmond, "Poisson models for extended target and group tracking," in *Proceedings of Signal and Data Processing of Small Targets*, vol. 5913. San Diego, CA, USA: SPIE, Aug. 2005, pp. 230–241.

[3] K. Gilholm and D. Salmond, "Spatial distribution model for tracking extended objects," *IEE Proceedings of Radar, Sonar and Navigation*, vol. 152, no. 5, pp. 364–371, Oct. 2005.

[4] M. Baum, F. Faion, and U. D. Hanebeck, "Modeling the Target Extent with Multiplicative Noise," in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 2406–2412.

[5] Y. Boers, H. Driessen, J. Torstensson, M. Trieb, R. Karlsson, and F. Gustafsson, "A track before detect algorithm for tracking extended targets," *IEE Proceedings Radar, Sonar and Navigation*, vol. 153, no. 4, pp. 345–351, Aug. 2006.

[6] K. Granström and C. Lundquist, "On the Use of Multiple Measurement Models for Extended Target Tracking," in *Proceedings of the International Conference on Information Fusion*, Instanbul, Turkey, Jul. 2013.

[7] N. Petrov, L. Mihaylova, A. Gning, and D. Angelova, "A novel sequential monte carlo approach for extended object tracking based on border parametrisation," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 306–313.

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

[9] M. Baum, B. Noack, and U. D. Hanebeck, "Extended Object and Group Tracking with Elliptic Random Hypersurface Models," in *Proceedings of the International Conference on Information Fusion*, Edinburgh, UK, Jul. 2010.

[10] K. Granström, C. Lundquist, and U. Orguner, "Tracking Rectangular and Elliptical Extended Targets Using Laser Measurements," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 592–599.

[11] S. Reuter and K. Dietmayer, "Pedestrian tracking using random finite sets," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 1101–1108.

[12] J. Degerman, J. Wintenby, and D. Svensson, "Extended target tracking using principal components," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 330–337.

[13] J. Lan and X.-R. Li, "Tracking of extended object or target group using random matrix – part I: New model and approach," in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 2177–2184.

[14] S. Reuter, B. Wilking, and K. Dietmayer, "Methods to model the motion of extended objects in multi-object Bayes filters," in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 527–534.

[15] C. Lundquist, K. Granström, and U. Orguner, "Estimating the Shape of Targets with a PHD Filter," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 49–56.

[16] M. Baum and U. D. Hanebeck, "Shape Tracking of Extended Objects and Group Targets with Star-Convex RHMs," in *Proceedings of the International Conference on Information Fusion*, Chicago, IL, USA, Jul. 2011, pp. 338–345.

[17] J. Lan and X.-R. Li, "Tracking of extended object or target group using random matrix – part II: Irregular object," in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 2185–2192.

[18] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, 2007.

[19] ——, "Multitarget Bayes filtering via first-order multi target moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

[20] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[21] R. Mahler, "PHD filters for nonstandard targets, I: Extended targets," in *Proceedings of the International Conference on Information Fusion*, Seattle, WA, USA, Jul. 2009, pp. 915–921.

[22] K. Granström, C. Lundquist, and U. Orguner, "A Gaussian mixture PHD filter for extended target tracking," in *Proceedings of the International Conference on Information Fusion*, Edinburgh, UK, Jul. 2010.

[23] ——, "Extended Target Tracking using a Gaussian Mixture PHD filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 3268–3286, Oct. 2012.

[24] K. Granström and U. Orguner, "A PHD filter for tracking multiple extended targets using random matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5657–5671, Nov. 2012.

[25] H. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780–783, aug 1988.

[26] R. Mahler, "On Multitarget Jump-Markov Filters," in *Proceedings of the International Conference on Information Fusion*, Singapore, Jul. 2012, pp. 149–156.

[27] S. Pasha, B.-N. Vo, H. Tuan, and W.-K. Ma, "A Gaussian mixture PHD filter for jump Markov system models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 919–936, Jul. 2009.

[28] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.