Fusing Radar and Scene Labeling Data for Multi-Object Vehicle Tracking

Alexander Scheel, Franz Gritschneder, Stephan Reuter, and Klaus Dietmayer*

Abstract: Scene labeling approaches which perform pixel-wise classification of images have become a very popular method for vehicle environment perception. They provide rich semantic information about objects in the surroundings which is oftentimes not available from other sensors. Yet, labeled images do not yield object-level information and object hypotheses incorporating object position and motion in a 3D world have to be retrieved through post-processing, e.g. tracking. This paper presents a vehicle tracking approach which combines the semantic information from scene labeling with precise range and Doppler data obtained from radar sensors. Thus, the respective strengths of both information sources are combined and an improved performance is achieved. By employing multi-object methods based on random finite sets, the proposed method is able to track multiple vehicles and to consider interdependencies. It is demonstrated using data from an experimental vehicle.

Keywords: information fusion, radar, scene labeling, tracking

1 Introduction

In scene labeling, each pixel in an image is classified with respect to the object type that it belongs to. Especially the increase in computational power in combination with the use of convolutional neural networks, e.g. [1], has led to tremendous progress in this field. For automated driving, scene labeling is a promising technique as it provides valuable semantic information about the environment. On the other hand, radar sensors are another widely used information source for vehicle environment perception as they—in contrast to monocular camera images—provide accurate distance and Doppler velocity measurements and work comparatively reliable in adverse lighting or weather conditions.

By fusing the data from both sources, the respective strengths can be combined. While scene labeling allows to distinguish vehicles from other objects and provides additional information about the object contour, radar measurements help to precisely determine the vehicle's motion and to locate it in a 3D world. Yet, there are several challenges that have to be overcome: Modern radar sensors provide many measurements from objects or clutter sources that have to be correctly processed. As radar measurements may originate from anywhere on the chassis and the Doppler velocities only indicate the radial portion of the actual object speed, there is considerable ambiguity in the corresponding

*The authors are with the Institute of Measurement, Control, and Microtechnology at Ulm University, Albert-Einstein-Allee 41, 89081 Ulm (e-mail: firstname.lastname@uni-ulm.de).

object state, especially when not only longitudinal but also cross traffic or turning vehicles are considered. Additionally, ambiguous associations between measurements and tightly spaced objects have to be resolved. In scene labeling, patches of classified pixels sometimes extend over several partially occluded vehicles and hence need to be handled correctly as well. Note that the two latter challenges arise if several objects are present and hence require correct treatment of the multi-object problem. Finally, the data from all sensors needs to be combined into a single and consistent representation of the environment.

This work proposes a multi-object tracking approach that makes use of finite set statistics [2] and extended object tracking methods. By using a Random Finite Set (RFS) representation of the multi-object state, the multi-object problem is tackled entirely probabilistically and object dependencies which for example occur in occlusion constellations can be considered. Data from a scene labeling module [3] and radar measurements are fused in a centralized fusion scheme and the measurement models work on the raw radar data and labeled images directly without further preprocessing routines.

In related work, Wojek et al. [4] presented a multi-object tracking approach for pedestrians and vehicles which works on monocular images only and combines information from scene labeling with classical object detections. The approach is able to consider occlusion and outputs bounding box representations of objects. Fusion of scene labeling information with other data sources has for example been presented by Nuss et al. [5] and Schneider et al. [6]. While [5] combines laser rangefinder data and labeled images for vehicle environment modeling in an occupancy grid, [6] combines stereo images with semantic data in semantic stixels. Both works hence achieve a higher abstraction level that allows for further processing but do not focus on obtaining object-level information. To the best of the authors' knowledge, however, this paper presents the first attempt to fuse radar data with semantic image information.

2 Tracking Approach

2.1 Multi-Object Formulation and State Estimation

The aim of multi-object tracking is to recursively estimate the kinematic state of all objects in the field of view (FOV) as well as the number of objects using a sequence of noisy measurements. In multiple extended object tracking, the extent of the objects is additionally estimated based on the received measurement set which typically incorporates several measurements per object. The multi-object Bayes filter [2] provides a mathematically rigorous framework for this problem based on RFSs which naturally capture the uncertainty in the number of objects as well as in their individual states.

A labeled multi-object state is defined as the RFS $\mathbf{X}_k = \{\mathbf{x}_k^{(1)}, \ldots, \mathbf{x}_k^{(n)}\}$ where the number of present objects is denoted by n and k is the time index. Each labeled state vector $\mathbf{x}_k = [x_k^T, \ell]^T$ consists of the object's state vector x_k and its unique label or identifier ℓ. Likewise, all measurements z_k from time step k are grouped in the measurement set $Z_k = \{z_k^{(1)}, \ldots, z_k^{(m)}\}$ where m is the total number of measurements.

In the prediction step, the multi-object Bayes filter computes the prior multi-object density

$$\pi_{k|k-1}(\mathbf{X}_k|Z_{1:k-1}) = \int f_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})\pi_{k-1|k-1}(\mathbf{X}_{k-1}|Z_{1:k-1})\delta\mathbf{X}_{k-1} \tag{1}$$

using the posterior multi-object density $\pi_{k-1|k-1}(\mathbf{X}_{k-1}|Z_{1:k-1})$ at time $k-1$ and a set integral as defined in [2]. Here, $Z_{1:k-1}$ denotes the set of all measurement sets up to the previous time step $k-1$. The multi-object transition density $f_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})$ models the individual objects' motion as well as object appearance and disappearance.

The update step incorporates the new measurements at time step k by applying Bayes' theorem. This leads to the posterior multi-object density

$$\pi_{k|k}(\mathbf{X}_k|Z_{1:k}) = \frac{g_k(Z_k|\mathbf{X}_k)\pi_{k|k-1}(\mathbf{X}_k|Z_{1:k-1})}{\int g_k(Z_k|\mathbf{X}_k)\pi_{k|k-1}(\mathbf{X}_k|Z_{1:k-1})\delta\mathbf{X}_k}, \tag{2}$$

where $g_k(Z_k|\mathbf{X}_k)$ is the multi-object likelihood function which models the measurement process. To improve readability and because all explanations focus on a single filter iteration, the time index k is dropped in the remainder of the paper.

In general, the multi-object Bayes filter is computationally intractable. However, the family of labeled multi-object distributions [7] allows for a closed form solution. Two distributions from this family are used in this work and briefly outlined in the following.

The multi-object density of the Labeled Multi-Bernoulli (LMB) distribution is

$$\boldsymbol{\pi}(\mathbf{X}) = \Delta(\mathbf{X})w(\mathcal{L}(\mathbf{X}))\prod_{\mathbf{x}\in\mathbf{X}} p(\mathbf{x}), \quad \text{where } w(L) = \prod_{i\in\mathbb{L}}\left(1-r^{(i)}\right)\prod_{l\in L}\frac{1_{\mathbb{L}}(\ell)r^{(\ell)}}{1-r^{(\ell)}}. \tag{3}$$

Here, the distinct label indicator $\Delta(\mathbf{X})$, which equals 1 if and only if all labels in $\mathbf{X}$ are unique and is zero otherwise, ensures that only meaningful multi-object states receive non-zero density values and the label projection function $\mathcal{L}(\mathbf{X}) = \{\ell \mid [x^T, \ell]^T \in \mathbf{X}\}$ is used to retrieve all labels from a multi-object state. Moreover, the inclusion function $1_{\mathbb{L}}(\ell)$ is 1 if and only if $\ell \in \mathbb{L}$. Intuitively, the LMB distribution consists of independent object hypotheses with corresponding probability of existence $r^{(\ell)}$ and state distribution $p(\mathbf{x})$. Due to the independence assumption, however, it is not able to represent influences of objects on each other.

The Generalized Labeled Multi-Bernoulli (GLMB) distribution overcomes this limitation by introducing an index $c \in \mathbb{C}$ and allowing for mixtures of different hypotheses with weights $w^{(c)}(L) > 0$,

$$\boldsymbol{\pi}(\mathbf{X}) = \Delta(\mathbf{X})\sum_{c\in\mathbb{C}} w^{(c)}(\mathcal{L}(\mathbf{X}))\prod_{\mathbf{x}\in\mathbf{X}} p^{(c)}(\mathbf{x}). \tag{4}$$

Note that all weights have to sum to one. Since the LMB distribution is a special case of the GLMB distribution, it can be easily converted to this form. However, the opposite direction, i.e. from GLMB to LMB, constitutes an approximation as information on the existence of objects is lost [8].

2.2 Tracking Procedure and Data Fusion

The proposed tracking approach represents all vehicles in the environment using LMB and GLMB distributions. Before prediction, the posterior distribution of the last time step is given in LMB form. Then, the multi-object state is predicted to the next time step using the standard prediction equations of the LMB filter [8]. Each object is predicted independently using its state model. Moreover, the persistence probability $p_S(\mathbf{x})$ models

the probability of an existing object to survive to the next step and thus governs the disappearance of objects. New object hypotheses are initialized using radar measurements from previous steps that indicated moving objects and have not considerably contributed to updating already existing objects.

Due to the independent prediction of the vehicles, the prior multi-object distribution allows for physically infeasible constellations with overlapping vehicles. It is therefore transformed to the GLMB form, which allows for modeling dependencies between objects, and it is ensured that only feasible object constellations are created. See [9] for a detailed description with corresponding equations.

Since the GLMB distribution is a conjugate prior for the extended object measurement models used in this work, the posterior distribution is again in GLMB form with updated parameters. Afterwards, the posterior GLMB distribution is approximated by a simpler LMB distribution. This avoids the combinatorial complexity of the GLMB filter at the cost of loosing information on object dependencies.

For fusing the scene labeling data with the radar measurements, the filter uses a centralized fusion scheme. That is, an entire filter recursion using the measurement model of the respective sensor is conducted each time new measurements arrive. Measurements are buffered and processed in correct order to avoid out-of-sequence problems.

Vehicles are assumed to follow a constant turn rate and velocity model. Hence, the kinematic state vector is given by $\xi = [x_R, y_R, \varphi, v, \omega]^T$. It is defined in the coordinate system of the ego-vehicle and the entries x_R and y_R denote the position of the center of the rear axle, φ represents the orientation, v is the speed of the vehicle, and ω denotes the yaw rate. The extent of a vehicle is described by its width a and length b which are combined in the extent vector $\zeta = [a, b]^T$. A constant value of 1.5 m is assumed for the height. Also, note that the position of the rear axle is fixed at 77% of the vehicle length since empirical studies showed that this value is suitable for most vehicles. The composed vehicle state is given by $x = [\xi^T, \zeta^T]^T$ and the state density is modeled by a Rao-Blackwellized particle distribution due to the non-linearity of the applied measurement models. Similar to [10], Rao-Blackwellization is used to reduce computational demands by only representing the kinematic portion of the state vector using particles. Each particle holds a separate distribution over the extent, which is updated analytically. Here, the extent is modeled by a discrete probability distribution with varying elements.

3 Measurement Models

For the two sensor types, different formulations of the multi-object likelihood function $g(Z|\mathbf{X})$ are used. While the radar measurement model uses a detection-type model, a separable likelihood approach is used for the scene labeling data.

3.1 Radar Measurement Model

The employed radar measurement model was first presented in [11] and is based on the general mutli-object formulation from [12]. It precisely models the relationship between an object's state and the received range, azimuth, and Doppler measurements. Additionally, it evaluates several possible association and clustering hypotheses. Hence, it is able to handle cross-traffic and turning vehicles as well as complicated association problems, e.g.

in the presence of clutter measurements from rotating wheels. See [11] for more details and the corresponding equations.

3.2 Scene Labeling Measurement Model

For updating the multi-object state using the scene labeling data, a separable likelihood approach which has been proposed for multi-object tracking in image sequences in [13] and [14] is used. It treats each pixel in an image as an independent measurement and concatenates the pixel values in the measurement vector $z = [z_1, \ldots, z_m]$. For the application to labeled images in this work, each pixel z_i is a binary variable which takes the value 1 if the pixel is labeled with the vehicle class or 0 if it belongs to any other class. Here, the traditional measurement set is now replaced by a vector, since the number of measurements m is fixed. Furthermore, the measurement model makes four assumptions:

1. The indices of pixels that belong to an object with labeled state $\mathbf{x}$ can be computed and are given by $T(\mathbf{x})$.

2. $T(\mathbf{x}) \cap T(\mathbf{x}') = \emptyset$ holds for all $\mathbf{x} \neq \mathbf{x}'$, since objects are assumed to stay separated in the measurement space.

3. The measurement likelihood of a single pixel given $\mathbf{x} = [x^T, \ell]^T$ is

$$p(z_i|\mathbf{x}) = \begin{cases} \varphi_i(z_i|x), & i \in T(\mathbf{x}) \\ \psi_i(z_i), & i \notin T(\mathbf{x}) \end{cases}, \tag{5}$$

   where $\varphi_i(z_i|x)$ is the likelihood for a pixel that belongs to an object and $\psi_i(z_i)$ is the background likelihood.

4. The measurements, i.e. pixel values, are (conditionally) independent.

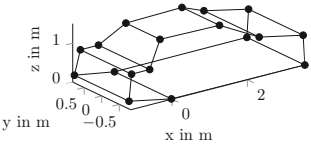
This results in the multi-object likelihood

$$g(z|\mathbf{X}) = f(z) \prod_{\mathbf{x}\in\mathbf{X}} g(z|\mathbf{x}) \;\text{ with }\; g(z|\mathbf{x}) = \prod_{i\in T(\mathbf{x})} \frac{\varphi_i(z_i|x)}{\psi_i(z_i)},\; f(z) = \prod_{i=1}^{m} \psi_i(z_i). \tag{6}$$

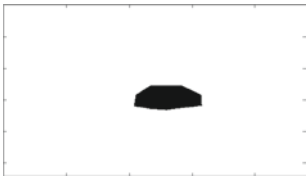
The single object likelihood function $g(z|\mathbf{x})$ is defined to be 1 if $T(\mathbf{x}) = \emptyset$. For the closed-form update equations in presence of a GLMB prior, please refer to [15]. In this work, the likelihood functions $\varphi_i(z_i|x)$ and $\psi_i(z_i)$ are Bernoulli distributions with parameters p_φ and p_ψ. These parameters specify the probability that a pixel is labeled as vehicle if it actually belongs to a vehicle and the probability that it is labeled as vehicle if it is actually part of the background.

To determine all pixels that belong to a vehicle $T(\mathbf{x})$, an image mask is constructed from a coarse 3D model of a vehicle. This is done individually for each particle in the state distribution. Then, $g(z|\mathbf{x})$ is evaluated for all pixels indicated by the mask.

The 3D vehicle model roughly describes the shape of a sedan. It is composed of 18 points and can be scaled to account for different vehicle lengths and widths. See Fig. 1a for an example. For each particle, the object model is projected onto the labeled image and the image mask is created by computing the convex hull around the points in pixel



(a) 3D vehicle model scaled at 5 m length, 1.8 m width, and 1.5 m height



(b) Exemplary vehicle mask for a vehicle at 10 m distance

Figure 1: Mask generation for the scene labeling model

coordinates. All pixels inside the hull are attributed to the vehicle. An exemplary image mask is depicted in Fig. 1b. Note that the convex hull is used because of its computational efficiency despite the resulting mask inaccuracies in certain configurations. As the goal of the model is only to give a rough approximation and vehicles vary considerably in their shape, this simplification is tolerable. Also, note that vehicles are assumed to drive on an even ground plane and the model does not consider unevenness of the road. If necessary, adding a state variable for the z-position is straightforward.

Assumption 2 of the separable likelihood model requires objects be separated in the measurement space. That is, objects in the multi-object state may not share pixels. In practice, however, this is the case in occlusion situations and individually computed vehicle masks could overlap. From a theoretical point of view, a violation of Assumption 2 results in a cardinality estimation error, i.e, in the estimated number of objects. Due to the representation of the multi-object state in GLMB form during update, however, such object dependencies can be considered. Therefore, the standard update is augmented by an occlusion handling similar to [16], where occlusions in laser scans were considered in a separable likelihood model.

In the prediction step, the prior GLMB distribution is constructed such that it is composed of several realizations of the multi-object state with a definite declaration of present objects. For each multi-object state hypothesis, the vehicle constellation is analyzed and the masks are adapted by subtracting all masks from vehicles that are closer to the camera. Note however, that this step is simplified by only considering the masks of the prior vehicle mean state to avoid computing all possible particle combinations.

4 Evaluation

The proposed tracking approach was implemented in MATLAB and is tested on data from an experimental vehicle. It is equipped with two short range radar sensors mounted in the corners of the front bumper with an opening angle of about 170° and a range of 43 meters as well as a wide angle mono camera behind the windshield (opening angle approx. 115°). Three scenarios are presented: A highly dynamic scenario with a single vehicle and precise ground truth values to examine the tracking accuracy, a scenario with two leading vehicles to demonstrate fusion benefits for resolving closely spaced targets, as well as another scenario for testing the capability to handle occlusion.

State	Radar only	Fusion
x_R in m	0.284	0.199
y_R in m	0.231	0.142
φ in °	5.029	2.755
v in m/s	0.365	0.274
ω in °/s	6.507	5.037
a in m	0.325	0.327
b in m	0.589	0.539

Table 1: RMSE values for the horizontal eight scenario

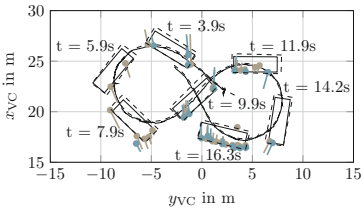


Figure 2: Horizontal eight scenario: estimated track and trajectroy (solid), ground truth (dashed), and radar measurements (beige - left sensor, blue - right sensor)

4.1 Tracking Accuracy

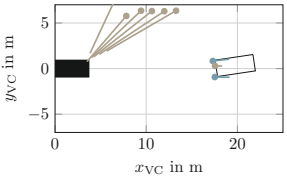
In the first scenario, the target vehicle drives a horizontal eight in front of the stationary ego-vehicle. This scenario is challenging due to the changing aspect angles and highly dynamic motion including fast turns with yaw rates up to 60°/s. Table 1 lists root mean squared error (RMSE) values for the scenario when tracking with radar only and when additionally fusing the semantic information from the image. The values are averaged over 20 Monte Carlo (MC) runs to reduce random effects due to the particle implementation. All estimation errors are decreased when fusing the scene labeling data, except for the width, where the error almost remains identical. Especially the estimate of vehicle orientation is considerably improved which is most likely due to the additional information from the object contour. An excerpt of the scenario is depicted in Fig. 2.

4.2 Multi-Object Scenario

The second scenario comprises two vehicles passing the ego-vehicle on the left and right and then driving side by side with little distance. When both vehicles are close to each other, the radar sensors intermittently yield merged measurements since the sensors cannot resolve the two objects with almost identical speed and distance. Due to this effect and the proximity of the resolved measurements, radar-only tracking sometimes mistakes the two vehicles as one. See Fig. 3 for an example. This effect can be tackled by explicitly considering target resolution constraints in the radar measurement model or by reducing approximations made in the update step, for instance, by incorporating more clustering hypotheses or refraining from the LMB approximation of the posterior distribution.

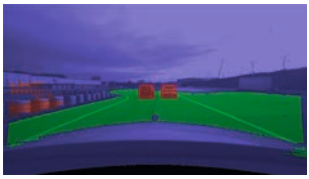
Another method, however, is to assist the radar-only tracking with additional scene labeling information. Figure 3a shows that both vehicles are clearly distinguishable in the camera image. A centralized fusion of both data sources thus enhances the estimation result and allows to continuously keep track of both vehicles. A projection of the fusion result into the labeled image is shown in Fig. 4a and Fig. 4b compares the estimated number of objects from radar-only tracking with the fusion result as well as the ground truth. The estimates have been averaged over 10 MC runs. Note that the fusion result is constantly close to the ground truth and does not show pronounced deviations.

(a) Track (red box) project into labeled image (blue - background, green - road, red - vehicle)

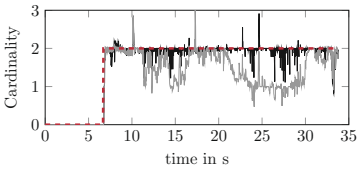


(b) Ego-vehicle (filled black), track (black rectangle), and radar data (beige - left sensor, blue - right sensor)

Figure 3: Radar-only results for two vehicles driving closely to each other



(a) Tracking results project into labeled image



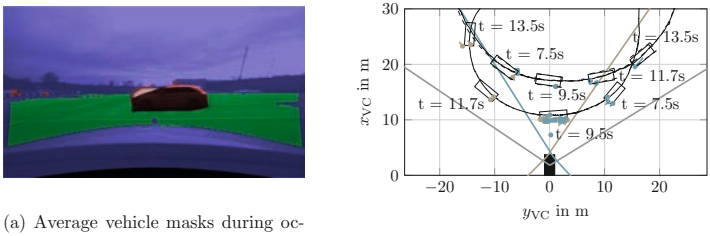
(b) Cardinality estimate: truth (dashed red), radar-only (gray), and fusion (black)

Figure 4: Estimation results for the multi-object scenario

4.3 Occlusion Scenario

To demonstrate occlusion, two vehicles are driving a semi-circular arch in the third scenario. Once both vehicles are in front of the stationary ego-vehicle, the front vehicle occludes the second one. Figure 5a illustrates the masks that are used for updating the multi-object hypothesis which contains both vehicles in an exemplary time step. Note that the illustrated masks are the average from all particle masks. Only the visible portion of the occluded vehicle is used to update the distribution. Hence, the filter avoids cardinality errors and overconfidence which occur if pixels are falsely used multiple times to update different objects. Also, this example demonstrates that the filter does not require preceding segmentation of labeled pixels into objects and is able to inherently cope with labeled pixel patches that encompass multiple objects. The estimation result including ground truth for the occluded vehicle is depicted in Fig. 5b.

At this point it is important to mention that occlusion handling as well as tracking accuracy obviously depend on a certain similarity between the tracked vehicles and the 3D model. Since the model is rather coarse and the convex hull additionally distorts the shape, the masks are suitable for most vehicles. This has been observed in supplementary experiments. Yet, the estimation errors understandably grow with increasing shape dissimilarity, e.g. for large utility vehicles.



(a) Average vehicle masks during occlusion: front vehicle (darkly tinted) and rear vehicle (lightly tinted)

(b) Exemplary track estimates with trajectories, FOVs, and measurements

Figure 5: Estimation results for the occlusion scenario

5 Conclusion

This paper proposes a new method to fuse radar and scene labeling data for vehicle tracking. It uses an RFS-based multi-object tracking approach in combination with extended object models that make full use of all available data. That is, no data preprocessing such as segmentation is required and the filter works on the radar targets as well as labeled images directly. Also, the multi-object formulation allows for considering occlusion situations and dealing with large patches of labeled pixels from several objects internally. As demonstrated in experiments, considerable improvement of the tracking results with respect to accuracy and the cardinality estimate is achievable by fusing the scene labeling information, despite the rather simple image processing based on a 3D vehicle model and image masks.

Acknowledgment

The authors would like to thank Markus Thom for providing the scene labeling module and for the insightful discussions.

References

[1] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning Hierarchical Features for Scene Labeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, 2013.

[2] Mahler, Ronald P. S., *Statistical multisource-multitarget information fusion*, ser. Artech House information warfare library. Boston: Artech House, 2007.

[3] M. Thom and F. Gritschneder, "Rapid Exact Signal Scanning With Deep Convolutional Neural Networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 5, pp. 1234–1250, 2017.

[4] C. Wojek, S. Walk, S. Roth, K. Schindler, and B. Schiele, "Monocular Visual Scene Understanding: Understanding Multi-Object Traffic Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, 2013.

[5] D. Nuss, M. Thom, A. Danzer, and K. Dietmayer, "Fusion of Laser and Monocular Camera Data in Object Grid Maps for Vehicle Environment Perception," in *Proceedings of the 2014 17th International Conference on Information Fusion*, 2014.

[6] L. Schneider, M. Cordts, T. Rehfeld, D. Pfeiffer, M. Enzweiler, U. Franke, M. Pollefeys, and S. Roth, "Semantic Stixels: Depth is Not Enough," in *Proceedings of the 2016 IEEE Intelligent Vehicles Symposium*, 2016, pp. 110–117.

[7] B.-T. Vo and B.-N. Vo, "Labeled Random Finite Sets and Multi-Object Conjugate Priors," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3460–3474, 2013.

[8] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, "The Labeled Multi-Bernoulli Filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3246–3260, 2014.

[9] A. Scheel, S. Reuter, and K. Dietmayer, "Using Separable Likelihoods for Laser-Based Vehicle Tracking with a Labeled Multi-Bernoulli Filter," in *Proceedings of the 2016 19th International Conference on Information Fusion*, 2016.

[10] A. Petrovskaya and S. Thrun, "Model Based Vehicle Detection and Tracking for Autonomous Urban Driving," *Autonomous Robots*, vol. 26, pp. 123–139, 2009.

[11] A. Scheel, C. Knill, S. Reuter, and K. Dietmayer, "Multi-Sensor Multi-Object Tracking of Vehicles Using High-Resolution Radars," in *Proceedings of the 2016 IEEE Intelligent Vehicles Symposium*, 2016, pp. 558–565.

[12] M. Beard, S. Reuter, K. Granström, B.-T. Vo, B.-N. Vo, and A. Scheel, "Multiple Extended Target Tracking with Labelled Random Finite Sets," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1638–1653, 2016.

[13] B.-N. Vo, B.-T. Vo, N.-T. Pham, and D. Suter, "Joint Detection and Estimation of Multiple Objects from Image Observations," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5129–5241, 2010.

[14] R. Hoseinnezhad, B.-N. Vo, and B.-T. Vo, "Visual Tracking in Background Subtracted Image Sequences via Multi-Bernoulli Filtering," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 392–397, 2013.

[15] F. Papi, B.-N. Vo, B.-T. Vo, C. Fantacci, and M. Beard, "Generalized Labeled Multi-Bernoulli Approximation of Multi-Object Densities," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5487–5497, 2015.

[16] A. Scheel, S. Reuter, and K. Dietmayer, "Vehicle Tracking Using Extended Object Methods: An Approach for Fusing Radar and Laser," in *Proceedings of the IEEE International Conference 2017*, 2017, [accepted].