Classifying Multiple Model Labeled Multi-Bernoulli Filter

Ganesh Nageswaran[1] and Rudolf Flierl
Department of Mechanical and Process Engineering
Technical University of Kaiserslautern
Kaiserslautern, Germany
2016

***Abstract*—This paper presents a method for simultaneous classification and robust tracking of traffic participants based on the labeled random finite set (RFS) tracking framework. Specifically, a method to integrate the object class information into the tracking loop of the multiple model labeled multi-Bernoulli (MMLMB) filter, using Dempster-Shafer evidence theory is presented. The multi-object state is estimated using the detections from the sensors and by propagation of multi-object density in a Bayesian fashion. Parallelly, the object class information is also predicted and updated recursively. The underlying object class information required for this could typically be obtained from different types of sensor such as radar, lidar and camera, using classical perception or more recent deep learning methods. On one hand, this enables an unified classification and tracking of traffic participants. On the other hand, it also increases the robustness of multi-object tracking, as the parameters of the tracking algorithm could be adapted using the class information. Moreover, using the Dempster-Shafer method for fusing class information from different sensor sources improves the overall performance, especially when the sensors have contradicting classification.**

***Index Terms*—multi-object tracking, random finite sets, object classification, sensor fusion, autonomous systems**

I. Introduction

Autonomous driving on a whole is a very complex task. It requires interaction between many layers of modules with different functionalities. One of the important functions is environment perception, which means sensing and understanding the surrounding of the vehicle. Environment perception itself involves further modules, in which multi-object tracking and classification plays a challenging part. Classification is the task of determining the type of the object. For example, in case of traffic participants, a classifier could determine whether the detected object is a pedestrian, bicycle or a passenger car. This could typically be achieved by extracting distinct features from sensor measurements, that could help to distinguish between different object categories. The aim of multi-object tracking on the other hand is to not only estimate the number of traffic participants present in the vehicle surrounding but also their individual states, i.e. their position, velocity etc. The multi-object tracking part and the classification part are typically handled independently. However, handling them in an unified way could have multiple advantages. For example, the classification could be further improved using the estimated object state information or the parameters of the tracking algorithm could be improved using the object class information.

The general idea of integrating classification or even other evidences into object tracking has been presented and investigated in many previous works. A generic approach of integrated evidence based existence estimation and object tracking is presented in [1]. Joint integrated probability data association (JIPDA) filter is combined with Dempster-Shafer theory (DST) of evidence, in order to predict and update the existence of an object, along with it's state. The objects are rather categorised as 'relevant' or 'not-relevant' on the application level. In the random finite set based tracking paradigm, a classification aided cardinalised probability hypothesis density (CPHD) is presented in [2], where the object class information is used for assisting data association in the update step of the filter. The mode variable in the likelihood function is replaced with a class variable and the confusion matrix from the classifier is used for class probabilities. In [3], object classification is integrated into multiple model probability hypothesis density filter (MMPHD) using DST. The elements of the transition matrix are adapted using the object class information, thereby the aim is to choose a more suitable motion model that fits that particular type of object. For example, it is beneficial to model the motion of a bicycle with a constant velocity (CV) model, whereas a passenger car with constant turn and velocity (CTRV) model. Although this improves the overall performance, as the PHD filters propagate only the first order moment of the multi-object distribution in time, they in general produce an unstable cardinality estimate i.e. the number of objects present [4].

A more accurate formulation of the multi-object Bayes filter in the forms of generalised labeled multi-Bernoulli (GLMB) filter and δ-GLMB filter, based on the labeled RFS is presented in [5]. Higher performance can be achieved with these filters due to the closed form solutions. However, this results in a higher computational complexity. The labeled multi-Bernoulli (LMB) filter presented in [6] reduces the computational complexity of δ-GLMB filter, where the multi-object posterior and prediction are approximated with LMB RFS, thereby reducing the number of hypotheses. A multiple model version of the

[1] This work was carried out by the author during his tenure as Scientific Employee at the Department of Mechanical and Process Engineering at Technical University Kaiserslautern, in cooperation with the company Knorr-Bremse System for Commercial Vehicles Private Limited in Schwieberdingen.

LMB filter is presented in [7]. It follows the same principle as the LMB filter, but the distribution of each track is given by the joint distribution from all the considered motion models. In this work, the object classification from different sources are fused using DST and integrated into the MMLMB tracking framework. Using a similar approach of adapting the motion model transition probabilities as in [3], but addressing the LMB RFS based filters is presented in this work.

The rest of the paper is organised as follows: section II begins with a brief background on the multi-object Bayes filter and labeled RFS. Then the concepts of the LMB filter and it's multiple model version, MMLMB are reviewed. Section III provides the details of the Gaussian mixture (GM) implementation of the MMLMB. A method to integrate the object classification information leading to the classifying MMLMB is introduced in section V, followed by results based on evaluation with measured sensor data.

II. Background

We begin the section with a brief background of the multi-object Bayes filter. Followed by a summary of the labeled RFS [5], based on which the LMB filter and it's multiple model version MMLMB are derived in [6] and [7] respectively.

A. Multi-Object Bayes Filter

A systematic unification of object detection, tracking and management based on finite set statistic (FISST) is proposed by Mahler [8]. A generalisation of the single-object Bayesian methods to the multi-object tracking problem is facilitated by modelling the object states and measurements as RFS, rather than individual random vectors. As in case of a single-object Bayes filter, the multi-object Bayes filter can also be realised with a recursion of prediction and update steps. But the difference is, in case of multi-object Bayes filter, the multi-object likelihood function and multi-object Markov density need to be formulated. An object state represented as a random variable x in a single-object Bayes filter is a RFS X with set of target states in multi-object Bayes filter. The multi-object Bayes filter prediction and update according to [9] are:

$$\pi_+(\mathrm{X}|\mathrm{Z}^k) = \int f_+(\mathrm{X}_+|\mathrm{X})\pi(\mathrm{X}|\mathrm{Z}^k)\delta\mathrm{X}, \tag{1}$$

$$\pi(\mathrm{X}_+|\mathrm{Z}^{k+1}) = \frac{g(\mathrm{Z}|\mathrm{X}_+)\pi_+(\mathrm{X}|\mathrm{Z}^k)}{\int g(\mathrm{Z}|\mathrm{X}_+)\pi_+(\mathrm{X}|\mathrm{Z}^k)\delta\mathrm{X}}. \tag{2}$$

in which the multi-object Markov transition density $f_+(\mathrm{X}_+|\mathrm{X})$ considering the disappearance and birth of objects is derived as

$$f_+(\mathrm{X}_+|\mathrm{X}) = \pi_B(\mathrm{X}_+)\pi_+(\emptyset|\mathrm{X})\cdot \sum_\theta \prod_{i:\theta(i)>0} \frac{p_S(x^{(i)})\cdot f_+(x_+^{\theta(i)}|x^{(i)})}{(1-p_S(x^{(i)}))\cdot\lambda_B b(x_+^{\theta(i)})}, \tag{3}$$

$$\pi(\emptyset|\mathrm{X}) = \prod_{i=1}^{|\mathrm{X}|}(1-p_S(x^{(i)})), \tag{4}$$

$$\pi_B(\mathrm{X}_+) = e^{-\lambda_B}\prod_{i=1}^{|\mathrm{X}_+|}\lambda_B b(x_+^{(i)}). \tag{5}$$

where λ_B is the expected number of new born objects distributed according to probability density $b(\cdot)$. $\pi_+(\emptyset|\mathrm{X})$ represents the probability none of the objects survive and $\pi_B(\mathrm{X}_+)$ the probability that all the objects are newly born. The multi-object likelihood function $g(\mathrm{Z}|\mathrm{X}_+)$ considering the associations and missed detections is derived as

$$g(\mathrm{Z}|\mathrm{X}_+) = \pi_C(Z)\pi(\emptyset|\mathrm{X}_+)\cdot \sum_\theta \prod_{i:\theta(i)>0} \frac{p_D(x_+^{(i)})\cdot g(z_{\theta(i)}|x_+^{(i)})}{(1-p_D(x_+^{(i)}))\cdot\lambda_c c(z_{\theta(i)})}, \tag{6}$$

$$\pi(\emptyset|\mathrm{X}_+) = \prod_{i=1}^{n}(1-p_D(x_+^{(i)})), \tag{7}$$

$$\pi_C(\mathrm{Z}) = e^{-\lambda_C}\prod_{z\in\mathrm{Z}}\lambda_c c(z). \tag{8}$$

where θ represents all the possible measurement-to-track association hypotheses, $\pi(\emptyset|\mathrm{X}_+)$ the probability that all detections are missed and $\pi_C(\mathrm{Z})$ the probability all measurements are clutter.

B. Labeled Random Finite Sets

In multi-object tracking paradigm, it is generally essential to have unique identities for each of the tracks. It is required for estimating object trajectories as well as to associate the track between time-steps t_{k+1} and t_k. In [5], the class of labeled RFS is proposed, where each object state is augmented with a distinct label $\ell \in \mathbb{L}$. The labels are drawn from a discrete label space $\mathbb{L} = \{\alpha_i : i \in \mathbb{N}\}$, where $\mathbb{N}$ is a set of positive integers and α_i's are distinct [5].

The multi-object labeled RFS can therefore be written as

$$\mathbf{X} = \{(x^{(1)},\ell^{(1)}),(x^{(2)},\ell^{(2)}),\dots(x^{(n)},\ell^{(n)})\} \subseteq \mathbb{X}\times\mathbb{L} \tag{9}$$

Moreover, each object should have an unique identity. Therefore a set of track labels $\mathcal{L}(\mathbf{X}) = \{\mathcal{L}(x,\ell) : (x,\ell)\in\mathbf{X}\}$ of the labeled multi-object state $\mathbf{X}$ is introduced in [5]. $\mathcal{L}(x,\ell) = \ell$ is then the projection of labeled state (x,ℓ) on the space $\mathbb{X}\times\mathbb{L}$ to the label space $\mathbb{L}$ denoted as $\mathcal{L}:\mathbb{X}\times\mathbb{L}\mapsto\mathbb{L}$. Also, each object label should be distinct, which means two different objects cannot have the same label. A label can be assigned exactly to one object only if the condition $|\mathcal{L}(\mathbb{X})| = \mathbb{X}$ and the distinct label indicator is given as

$$\Delta(\mathbf{X}) = \delta_{|\mathbf{X}|}(\mathcal{L}(|\mathbf{X}|)) \tag{10}$$

which takes the value one if the object labels are distinct, zero otherwise. More detailed derivation and properties of labeled RFS are presented in [5].

Further, augmenting each object of a multi-Bernoulli RFS with a distinct label gives the labeled multi-Bernoulli (LMB) RFS defined by the existence and spatial parameters of each object represented by it's corresponding label

$$\pi(\mathbf{X}) = \left\{(r^{(\ell)},p^{(\ell)})\right\}_{\ell\in\mathbb{L}} \tag{11}$$

The multi-object probability density of LMB RFS on $\mathbb{X} \times \mathbb{L}$ is given by [5]

$$\begin{aligned} \pi((x^{(1)}, \ell^{(1)}), \dots (x^{(n)}, \ell^{(n)})) = \delta_n(|l_1, \dots, l_n|) \\ \times \prod_{i \in \mathbb{L}} (1 - r^{(i)}) \prod_{\ell=1}^{n} \frac{1_{\mathbb{L}} r^{(\ell)} p^{(\ell)}}{1 - r^{(\ell)}} \end{aligned} \tag{12}$$

A compact notation of the LMB RFS as given in [5] and [6] using the multi-object exponential notation is

$$\pi(\mathbf{X}) = \Delta(\mathbf{X}) w(\mathcal{L}(\mathbf{X})) p^{\mathbf{X}} \tag{13}$$

where

$$\begin{aligned} w(L) &= \prod_{i \in \mathbb{L}} (1 - r^{(i)}) \prod_{\ell \in L} \frac{1_{\mathbb{L}}(\ell) r^{(\ell)}}{1 - r^{(\ell)}}, \\ p(x, \ell) &= p^{(\ell)}(x). \end{aligned} \tag{14}$$

The multiple model version of the LMB filter derived in [7] and addressed in this work is based on the LMB RFS.

C. Labeled Multi-Bernoulli Filter

An exact closed form solution of the multi-object Bayes filter is proposed in [10] on the basis of labeled RFS. Vo and Vo introduced the generalised labeled multi-Bernoulli (GLMB) filter and it is shown that the δ-GLMB density is closed under multi-object prediction and update. In case of δ-GLMB, both the prediction and update of tracks requires generating all the possible hypothesis, which means with the increase in number of objects and measurements, the computational complexity increases exponentially in both prediction and update. The LMB filter is presented in [6], where multi-object posterior and prediction in δ-GLMB are approximated by LMB RFS. This approximation reduces the total required hypotheses in the prediction step. It is shown in [6] that the LMB reduces the computational complexity by approximation of posterior, but still produces identical results of δ-GLMB filter in many scenarios.

1) Prediction: In the multi-object Bayesian recursion implementation of LMB filter, the multi-object posterior density of the previous time step is an approximated LMB RFS of the form

$$\pi(\mathbf{X}) = \Delta(\mathbf{X}) w(\mathcal{L}(\mathbf{X})) p^{\mathbf{X}} \tag{15}$$

where

$$\begin{aligned} w(L) &= \prod_{i \in \mathbb{L}} (1 - r^{(i)}) \prod_{\ell \in L} \frac{1_{\mathbb{L}}(\ell) r^{(\ell)}}{1 - r^{(\ell)}}, \\ p(x, \ell) &= p^{(\ell)}(x). \end{aligned} \tag{16}$$

Also the multi-object birth intensity is considered to be an LMB RFS of the form

$$\pi_B(\mathbf{X}) = \Delta(\mathbf{X}) w_B(\mathcal{L}(\mathbf{X})) [p_B]^{\mathbf{X}} \tag{17}$$

where

$$w_B(I) = \prod_{i \in \mathbb{B}} (1 - r_B^{(i)}) \prod_{\ell \in I} \frac{1_{\mathbb{B}}(\ell) r_B^{(\ell)}}{1 - r_B^{(\ell)}}, \tag{18}$$

$$p_B(x, \ell) = p_B^{(\ell)}(x). \tag{19}$$

with the labels $l \in \mathbb{B}$ of the new born objects distinct and doesn't have the same labels as the surviving objects i.e. $\mathbb{L} \cap \mathbb{B} = \emptyset$.

The prediction density in a current time step is union of the surviving tracks of the previous time step in a label space $\mathbb{L}$ and the newly born tracks in current time step in label space $\mathbb{B}$. The tracks from previous time step are considered to survive with a probability of $p_S(x, \ell)$ and evolve according to a standard Markov transition density $f_+(x|x', \ell)$ or disappear with a probability $q_S(x, \ell) = 1 - p_S(x, \ell)$. The predicted LMB distribution in label space $\mathbb{L}_+ = \mathbb{L} \cup \mathbb{B}$ is [6]

$$\pi_+ = \left\{ (r_{+,S}^{(\ell)}, p_{+,S}^{(\ell)}) \right\}_{\ell \in \mathbb{L}} \cup \left\{ (r_B^{(\ell)}, p_B^{(\ell)}) \right\}_{\ell \in \mathbb{B}}. \tag{20}$$

with

$$r_{+,S}^{(\ell)} = \eta_S(\ell) r^{(\ell)}, \tag{21}$$

$$p_{+,S}^{(\ell)} = \langle p_S(\cdot, \ell) f_+(x|\cdot, \ell), p(\cdot, \ell) \rangle / \eta_S(\ell), \tag{22}$$

$$\eta_S(\ell) = \int \langle p_S(\cdot, \ell) f_+(x|\cdot, \ell), p(\cdot, \ell) \rangle \, dx. \tag{23}$$

2) Update: It is shown in [6] that the LMB is not closed under update operation and the predicted LMB density needs to be expressed in δ-GLMB form. Therefore before the update step, the predicted LMB density needs to be represented in δ-GLMB form [10], [6] and is given by

$$\pi_+(\mathbf{X}_+) = \Delta(\mathbf{X}_+) \sum_{I_+ \in \mathcal{F}(\mathbb{L}_+)} w_+^{(I_+)} \delta_{I_+}(\mathcal{L}(\mathbf{X}_+)) \, [p_+]^{\mathbf{X}_+} \tag{24}$$

where I_+ denotes a hypothesis containing a set of track labels and for representing the LMB RFS in δ-GLMB form all the possible hypotheses corresponding to the permutations of track labels need to be generated. With the new set of measurements Z and no history of association maps available, the δ-GLMB posterior can then be given according to the δ-GLMB update [10], [6] as

$$\begin{aligned} \pi_+(\mathbf{X}_+|\mathrm{Z}) = \Delta(\mathbf{X}) \sum_{(I_+, \theta) \in \mathcal{F}(\mathbb{L}_+) \times \Theta_{I_+}} w^{(I_+, \theta)}(\mathrm{Z}) \\ \cdot \, \delta_{I_+}(\mathcal{L}(\mathbf{X})) \left[p^{(\theta)}(.|\mathrm{Z}) \right]^{\mathbf{X}} \end{aligned} \tag{25}$$

where θ denotes the measurement to track association map for the track labels in each of the hypothesis I_+. The posterior weight for each of the hypothesis being $w^{(I_+, \theta)}$, the measurement set $\mathbb{M}$, the measurement updated posterior spatial distribution of each track $p^{(\theta)}(x, \ell|\mathrm{Z})$ and the likelihoods $\eta_{\mathrm{Z}}^{(\theta)}(\ell)$.

The δ-GLMB measurement updated posterior (25) needs to be again represented in LMB RFS, so that it can be predicted again in the next time step. The posterior is approximated by the existence and spatial parameters of all the tracks in $\mathbf{X}$, with a matching PHD and mean cardinality [6]

$$\pi(\mathbf{X}|\mathrm{Z}) = \left\{ (r^{(\ell)}, p^{(\ell)}) \right\}_{\ell \in \mathbb{L}_+} \tag{26}$$

where the posterior existence probability $r^{(\ell)}$ and posterior spatial distribution $p^{(\ell)}$ of each track is given as

$$r^{(\ell)} = \sum_{(I_+,\theta)\in\mathcal{F}(\mathbb{L}_+)\times\Theta_{I_+}} w^{(I_+,\theta)}(\mathrm{Z})1_{I_+}(\ell) \tag{27}$$

$$p^{(\ell)}(x) = \frac{1}{r^{(\ell)}} \sum_{(I_+,\theta)\in\mathcal{F}(\mathbb{L}_+)\times\Theta_{I_+}} w^{(I_+,\theta)}(\mathrm{Z})1_{I_+}(\ell)p^{(\theta)}(x,\ell) \tag{28}$$

3) Multiple Model Version: In a typical multi-object tracking application scenario, many classes of objects appear in the sensor's field of view (FoV). Especially in urban scenarios, the object can be a pedestrian, bicyclist or even other vehicles. For predicting the motion of the object, using a single motion model for all the scenarios and all the object classes would be a wrong assumption. Pedestrians can be assumed to move with constant velocity, whereas bicyclists and other vehicles can also make turn and accelerate rapidly. Therefore it is beneficial to consider more than one motion model for different scenarios and object classes. Multiple model versions for the class of RFS filters based on Jump Markov System (JMS), specifically for PHD filter is proposed as in [14]. The basic idea in JMS is to append a discrete mode variable $o = \in \mathbb{O}$ to the kinematic state x, giving an augmented state of the single-object $\mathbf{x} = (x, o)$, with $x \in \mathbb{X}$. The discrete mode variable represents the different motion models considered for modelling the object motion with $\mathbb{O}$ representing the discrete space of all the possible motion models. In [7] the multiple model labeled multi-Bernoulli (MMLMB) filter is introduced, based again on JMS. It uses the same architecture as the LMB filter, but the spatial distribution of each track is given by joint distribution from all the considered motion modes. The multiple model representation for the labeled RFS case is then given by additionally including the object's label ℓ in the augmented state as [7]

$$\mathbf{X} = \{(x^{(1)},\ell^{(1)},o^{(1)}),\dots,(x^{(n)},\ell^{(n)},o^{(n)})\} \subseteq \mathbb{X}\times\mathbb{L}\times\mathbb{O} \tag{29}$$

The elements $t_{o,o'}$ of the Markov transition matrix $\chi_{o,o'}$ models the transition probability from modes o' to o. In comparison to a single model LMB filter, the difference in the MMLMB filter is that the spatial distribution of each track is given by the joint distribution, augmenting the kinematic state with the motion model as

$$p^{(\ell)}(x,o) = p^{(\ell)}(o)p^{(\ell)}(x|o)\ \forall o \in \mathbb{O} \tag{30}$$

Consequently, the multiple model representation of multi-object probability density of the LMB RFS (11) is given in as [7]

$$\pi(\mathbf{X}) = \left\{(r^{(\ell)}, p^{(\ell)}(o)p^{(\ell)}(\cdot|o)\right\}_{\ell\in\mathbb{L}} \tag{31}$$

and the birth distribution corresponding to the form in (17) as,

$$\pi_B(\mathbf{X}) = \left\{(r_B^{(\ell)}, p_B^{(\ell)}(o)p_B^{(\ell)}(\cdot|o)\right\}_{\ell\in\mathbb{B}} \tag{32}$$

where the spatial distribution of the tracks is a joint distribution defined in the space $\mathbb{X}\times\mathbb{O}$ and the labels of surviving and birth tracks are distinct such that the property $\mathbb{L}\cap\mathbb{B} = \emptyset$ holds. The prediction step of MMLMB filter follows the prediction of the single model LMB filter, but the spatial distribution of the tracks is a joint distribution conditioned on all the motion models present in the system, i.e. the objects' state are predicted with all the considered motion models. Consequently the predicted spatial distribution, normalisation constant and existence probability of a track can be given as [7]

$$p_{+,S}^{(\ell)}(x|o) = \frac{\int p_S^{(\ell)}(x')f(x|x',o')p^{(\ell)}(x'|o')dx'}{\eta_S(\ell)} \tag{33}$$

$$\eta_S(\ell) = \sum_{o'\in\mathbb{O}} p^{(\ell)}(o') \int p_S(x')p^{(\ell)}(x',o')dx' \tag{34}$$

$$r_{+,S}^{(\ell)} = \eta_S(\ell)\cdot r^{(\ell)} \tag{35}$$

Similarly, the conversion of the LMB RFS into δ-GLMB representation and the update procedure is same as that of the LMB filter, but with the spatial distribution being conditioned on the motion model. The updated multi-object posterior can then be described by the parameters [7]

$$r^{(\ell)} = \sum_{(I_+,\theta)\in\mathcal{F}(\mathbb{L}_+)\times\Theta_{I_+}} w^{(I_+,\theta)}(\mathrm{Z})1_{I_+}(\ell) \tag{36}$$

$$p^{(\ell)}(x|o) = \frac{1}{r^{(\ell)}} \sum_{(I_+,\theta)\in\mathcal{F}(\mathbb{L}_+)\times\Theta_{I_+}} w^{(I_+,\theta)}(\mathrm{Z})1_{I_+}(\ell)p^{(\ell,\theta)}(x|o) \tag{37}$$

$$p^{(\ell)}(o) = \frac{1}{r^{(\ell)}} \sum_{(I_+,\theta)\in\mathcal{F}(\mathbb{L}_+)\times\Theta_{I_+}} w^{(I_+,\theta)}(\mathrm{Z})1_{I_+}(\ell)p^{(\ell,\theta)}(o). \tag{38}$$

III. Gaussian Mixture Implementation of MMLMB Filter

In case of Gaussian mixture (GM) implementation, the distribution of each track is approximated by a sum of Gaussian components. Given the mean of the Gaussian state $\hat{x}$ for a track ℓ, the covariance $\underline{\mathrm{P}}$, transition matrix $\underline{\mathrm{F}}$, the noise matrix $\underline{\mathrm{Q}}$ and assuming the survival probability p_S of the track to be independent of it's state, the normalisation constant and the corresponding predicted spatial density of a track ℓ can be given of the form

$$\begin{aligned}
\eta_S(\ell) &= \sum_{o'\in\mathbb{O}} p^{(\ell)}(o') \int p_S \sum_{j=1}^{J^{(\ell)}(o')} w^{(\ell,j)}(o') \\
&\quad \cdot \mathcal{N}\left(x; \underline{\mathrm{F}}(o)\hat{x}^{(\ell,j)}, \underline{\mathrm{F}}(o)\underline{\mathrm{P}}^{(\ell,j)}\underline{\mathrm{F}}(o)^\top + \underline{\mathrm{Q}}(o)\right)dx \\
&= p_S \sum_{o'\in\mathbb{O}} p^{(\ell)}(o') \sum_{j=1}^{J^{(\ell)}(o')} w^{(\ell,j)}(o') \\
&\quad \cdot \int \mathcal{N}\left(x; \underline{\mathrm{F}}(o)\hat{x}^{(\ell,j)}, \underline{\mathrm{F}}(o)\underline{\mathrm{P}}^{(\ell,j)}\underline{\mathrm{F}}(o)^\top + \underline{\mathrm{Q}}(o)\right)dx \\
&= p_S
\end{aligned} \tag{39}$$

$$p_{+,S}^{(\ell)}(x,o) = \sum_{o' \in \mathbb{O}} t_{o,o'} p^{(\ell)}(o') \sum_{j=1}^{J^{(\ell)}} w^{(\ell,j)} \cdot \mathcal{N}\left(x; \underline{\mathrm{F}}(o)\hat{x}^{(\ell,j)}, \underline{\mathrm{F}}(o)\underline{\mathrm{P}}^{(\ell,j)}\underline{\mathrm{F}}(o)^\top + \underline{\mathrm{Q}}(o)\right) \tag{40}$$

which can then be factorised into the probability of motion model and spatial distribution parts as

$$p_{+,S}^{(\ell)}(o) = \sum_{o' \in \mathbb{O}} t_{o,o'} p^{(\ell)}(o') \tag{41}$$

$$p_{+,S}^{(\ell)}(x|o) = \sum_{j=1}^{J^{(\ell)}} w^{(\ell,j)} \tag{42}$$

$$\cdot \mathcal{N}\left(x; \underline{\mathrm{F}}(o)\hat{x}^{(\ell,j)}, \underline{\mathrm{F}}(o)\underline{\mathrm{P}}^{(\ell,j)}\underline{\mathrm{F}}(o)^\top + \underline{\mathrm{Q}}(o)\right) \tag{43}$$

Similarly for the update, given the mean of Gaussian measurement $z_\theta(\ell)$ for the track ℓ conditioned on mode o, the innovation matrix $\underline{\mathrm{S}}$ and assuming state independent detection probability p_D, the update likelihood of the measurement in case the object being detected $(\theta(\ell) > 0)$ is of the form

$$\begin{aligned}\eta_{\mathrm{Z}}^{(\theta)}(\ell|o) &= \int \frac{p_D}{\kappa(z_{\theta(\ell)})} \sum_{j=1}^{J_+^{(\ell)}(o)} w_+^{(\ell,j)}(o) \\ &\quad \cdot \mathcal{N}\left(z_{\theta(\ell)}; z_+^{(\ell,j)}, \underline{\mathrm{S}}^{(\ell,j)}\right) \mathcal{N}\left(x; \hat{x}^{(\ell,j,\theta)}, \underline{\mathrm{P}}^{(\ell,j)}\right) dx \\ &= \frac{p_D}{\kappa(z_{\theta(\ell)})} \sum_{j=1}^{J_+^{(\ell)}(o)} w_+^{(\ell,j)}(o) \mathcal{N}\left(z_{\theta(\ell)}; z_+^{(\ell,j)}, \underline{\mathrm{S}}^{(\ell,j)}\right).\end{aligned} \tag{44}$$

The updated posterior spatial density of the track ℓ conditioned on mode o is then given for the case $\theta(\ell) > 0$ as,

$$p^{(\ell,\theta)}(x|o) = \tag{45}$$

$$\frac{\frac{p_D}{\kappa(z_{\theta(\ell)})} \sum_{j=1}^{J_+^{(\ell)}(o)} w_+^{(\ell,j)}(o) \cdot C}{\frac{p_D}{\kappa(z_{\theta(\ell)})} \sum_{j=1}^{J_+^{(\ell)}(o)} w_+^{(\ell,j)}(o) \mathcal{N}\left(z_{\theta(\ell)}; z_+^{(\ell,j)}, \underline{\mathrm{S}}^{(\ell,j)}\right)} \tag{46}$$

$$= \sum_{j=1}^{J_+^{(\ell)}(o)} w_Z^{(\ell,j,\theta)}(o) \mathcal{N}\left(x; \hat{x}^{(\ell,j,\theta)}, \underline{\mathrm{P}}^{(\ell,j)}\right) \tag{47}$$

with the term C and posterior weights of the Gaussian components given as

$$C = \mathcal{N}\left(z_{\theta(\ell)}; z_+^{(\ell,j)}, \underline{\mathrm{S}}^{(\ell,j)}\right) \mathcal{N}\left(x; \hat{x}^{(\ell,j,\theta)}, \underline{\mathrm{P}}^{(\ell,j)}\right) \tag{48}$$

$$w_Z^{(\ell,j,\theta)}(o) = \frac{\frac{p_D}{\kappa(z_{\theta(\ell)})} w_+^{(\ell,j)}(o) \mathcal{N}\left(z_{\theta(\ell)}; z_+^{(\ell,j)}, \underline{\mathrm{S}}^{(\ell,j)}\right)}{\eta_{\mathrm{Z}}^{(\theta)}(\ell|o)} \tag{49}$$

In case the object is not detected $(\theta(\ell) = 0)$, the likelihood is of the form

$$\eta_{\mathrm{Z}}^{(\theta)}(\ell|o) = \int (1 - p_D) \sum_{j=1}^{J_+^{(\ell)}(o)} w_+^{(\ell,j)}(o) \tag{50}$$

$$\cdot \mathcal{N}\left(x; \hat{x}^{(\ell,j,\theta)}, \underline{\mathrm{P}}^{(\ell,j)}\right) dx \tag{51}$$

$$= (1 - p_D) \sum_{j=1}^{J_+^{(\ell)}(o)} w_+^{(\ell,j)}(o) \tag{52}$$

$$= (1 - p_D), \tag{53}$$

and the corresponding posterior spatial density is same as the prediction, given as

$$p^{(\ell,\theta)}(x|o) = \sum_{j=1}^{J_+^{(\ell)}(o)} w^{(\ell,j,\theta)}(o) \mathcal{N}\left(x; \hat{x}^{(\ell,j,\theta)}, \underline{\mathrm{P}}^{(\ell,j)}\right) \tag{54}$$

where

$$w^{(\ell,j,\theta)}(o) = w_+^{(\ell,j)}, \tag{55}$$

$$\hat{x}^{(\ell,j,\theta)} = \hat{x}_+^{(\ell,j)}, \tag{56}$$

$$\underline{\mathrm{P}}^{(\ell,j)} = \underline{\mathrm{P}}_+^{(\ell,j)}. \tag{57}$$

Consequently, the normalisation constant considering the measurement likelihood for track ℓ averaged over all the motion modes $o \in \mathbb{O}$ if $\theta(\ell) > 0$ is given as

$$\eta_{\mathrm{Z}}^{(\theta)}(\ell) = \sum_{o \in \mathbb{O}} p_+^{(\ell)}(o) \frac{p_D}{\kappa(z_{\theta(\ell)})} \sum_{j=1}^{J_+^{(\ell)}(o)} w_+^{(\ell,j)}(o) \cdot \mathcal{N}\left(z_{\theta(\ell)}; z_+^{(\ell,j)}, \underline{\mathrm{S}}^{(\ell,j)}\right) \tag{58}$$

and $(1 - p_D)$, if $\theta(\ell) = 0$. The normalisation constant is then used for calculating the updated weight $w^{I_+,\theta}(\mathrm{Z})$ of the hypothesis component, and to represent the joint probability density $p^{(\theta)}(x, \ell|\mathrm{Z})$ of the track ℓ, considering all the motion models.

IV. Dempster-Shafer Theory

The Dempster-Shafer theory (DST) of evidence is a generalised probability theory, which defines the occurrence of an event as combination of evidences from many sources, rather than individual probability constituents. The DST provides a framework for reasoning uncertainty with the help of belief functions. The set of all the states or hypotheses under consideration is represented by a finite set of n elements, called the Frame of Discernment (FOD)

$$\Omega = \{e_1, e_2, \ldots, e_n\} \tag{59}$$

The occurrence of an event is then calculated from the power set $\mathbb{P}(\Omega)$ of the order 2^Ω, consisting all the elements and subsets of the elements in FOD, including the null-set. A belief for each element in the power set, expressing the evidence from different sources, is given as a basic belief mass (BBM) $m(E_i) \geq 0$. Every $E \in \Omega$ with $m(E) > 0$ is called the focal hypotheses. The masses of all the subsets of power set equals to 1 and is defined by the basic belief assignment (BBA) function given as:

$$\begin{aligned} m(\emptyset) &= 0 \\ \sum_{E\in\mathbb{P}(\Omega)} m(E) &= 1 \end{aligned} \tag{60}$$

Unlike the probability theory based approaches where the masses can be given only to the elements of Ω, in DST the masses can be given to any subsets of Ω. Moreover, the BBA in (60) can support a set E, without supporting any of it's subsets, which represents the limited knowledge capacity of DST. The mass function $m(E)$ is called a categorical mass function in case there is only one focal set, such that $m(E) = 1$ with $E \subseteq \Omega$. If $E = \Omega$ in the above case, then the condition $m(\Omega)$ represents the state of total ignorance, which means none of the individual subsets of Ω are supported with an evidence.

The degree of belief $bel(E)$ is defined in the transferable belief model (TBM) in [11] as the lower bound of the probability interval, which is the sum of all masses that support E.

$$\begin{aligned} bel &: 2^{\Omega} \mapsto [0,1] \\ bel(E) &= \sum_{\emptyset\neq X\subseteq E} m(X) \end{aligned} \tag{61}$$

Moreover, a concept on reliability of an evidence is proposed, in which the masses for evidences from less reliable sources are multiplied with a discounting factor, to represent a lesser weight for those unreliable sources. The degree of plausibility $pl(E)$ is given in [11] as the total amount of belief that supports E.

$$\begin{aligned} pl &: 2^{\Omega} \mapsto [0,1] \\ pl(E) &= \sum_{X\cap E\neq\emptyset} m(X) \end{aligned} \tag{62}$$

The degree of plausibility represents the upper bound of the uncertainty $u(E)$ and therefore it is given as

$$bel(E) \leq u(E) \leq pl(E) \tag{63}$$

Dempster's rule of combination provides a basis for the combining the masses of different pieces of evidence. The combination for the case $E = \emptyset$ is given as:

$$(m_1 \oplus m_2)(E) = 0 \tag{64}$$

and for the case $E \neq \emptyset$ as

$$(m_1 \oplus m_2)(E) = \frac{\sum\limits_{X\cap Y=E} m(X)m(Y)}{1 - \sum\limits_{X\cap Y=\emptyset} m(X)m(Y)} \; \forall E \in 2^{\Omega} \tag{65}$$

The method of discounting in the TBM enables to model the reliability of BBA from a particular source. If the belief masses in BBA are considered to be coming from imprecise source, they should be used only with a certain degree of confidence. Consequently, discounting of a BBA with a parameter d depending on it's source can be given as

$$m^d(E) = \begin{cases} d \cdot m(E), & \text{if } E \neq \emptyset \\ 1 - d + d \cdot m(E), & \text{if } E = \Omega. \end{cases} \tag{66}$$

In the TBM in [11], two levels are defined as the credal level and the pignistic level. The formulation of belief functions and combination of evidences from many sources are all said to be in the so called credal level. Then the actual decisions based on the belief functions in the credal level are made in the so called pignistic level. Further, the pignistic transformation is given as:

$$BetP_m(E) = \sum_{B\subseteq\Omega, B\neq\emptyset} \frac{|E\cap B|}{|B|} m(B) \tag{67}$$

V. Classifying MMLMB Filter

Considering the Gaussian mixture implementation of the MMLMB filter (GM-MMLMB) described in III, every track is basically augmented with the motion mode o along with it's label ℓ and state x. With the aim of integrating object class information into the GM-MMLMB filter, we augment each track component additionally with the classification information. Thereby, a track ℓ which has state x, motion mode o and class-BBA $m_C^{(\ell)}$, can be represented as $\left(x, \ell, o, m_C^{(\ell)}\right)$ and the i^{th} Gaussian mixture component (GMC) of it as

$$\left(w^{(i)}(o), \mathcal{N}(x; \hat{x}^{(i)}, \underline{\mathrm{P}}^{(i)}), m_C^{(\ell)}\right) \tag{68}$$

where it holds the class-BBA of it's parent track. The class-BBA $m_C^{(\ell)}$ of a track represents the certainty values of different object classes that are in question. The focal elements of the FOD Ω used in the DST are then the considered object classes. Therefore, the surviving tracks and the birth tracks contain their class-BBA along with their joint probability density. The scheme for integrating the classification information alongside the multi-object filter recursion with the GM-MMLMB filter is presented in Figure 1. Following [1] and [3], the prediction and update steps are divided into state and class levels, with interactions in between both the levels. The class-BBA information from time k is used by the state level for adapting parameters of multiple model object state prediction into time $k+1$ and the measurement updated object state information is used for the posterior class-BBA at time $k+1$. Consequently, the interactions denote that the class-BBA is used for two purposes as in [3]: to adapt the mode transition probabilities and to classify the object based on both measurement and track features.

1) Prediction: In the track prediction step of the GM-MMLMB filter, mean and covariance of each of the track's GMC is predicted, for each of the considered motion mode $o \in \mathbb{O}$, thereby new GMCs are generated. The weight of the predicted GMC is obtained by multiplying the weight of it's parent GMC with the transition and model probabilities of the parent track. The state prediction step follows the prediction equations of the GM-MMLMB filter, as in (40).

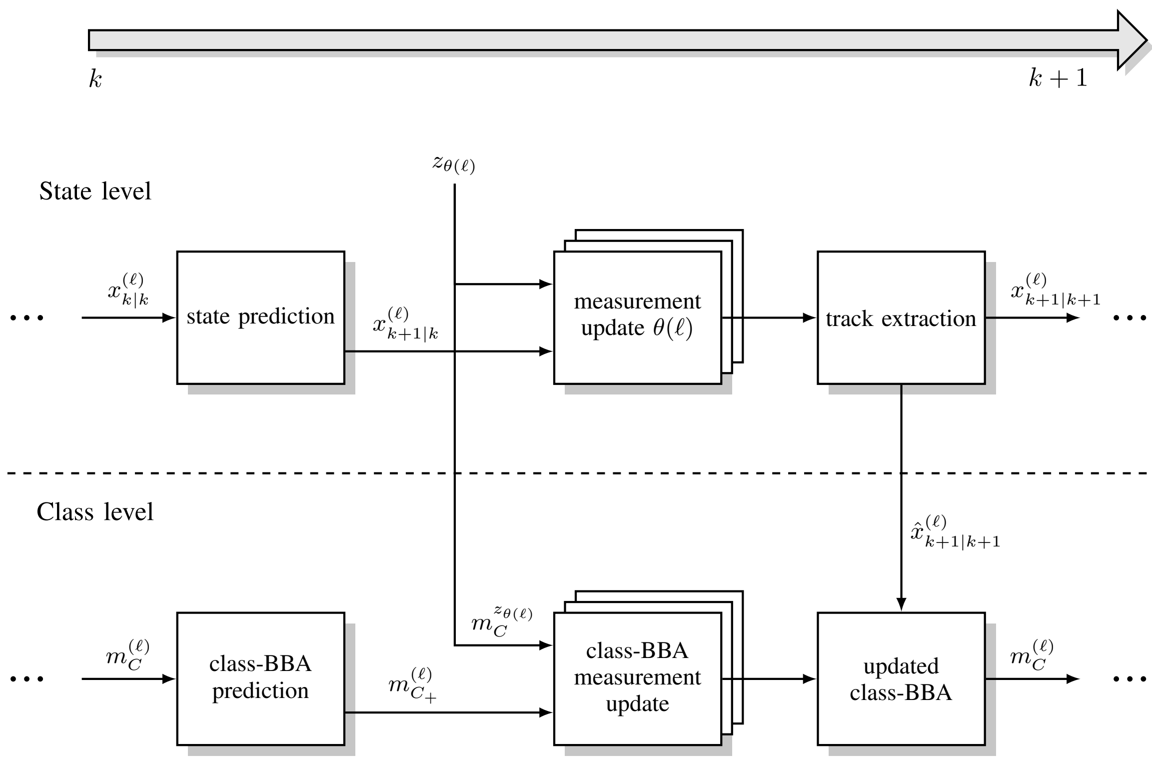


Fig. 1: Integrating class information into tracking framework. The state of the track is predicted and updated with GM-MMLMB filter equations on the upper state level, where the class-BBA influences the motion model transition. In the lower path, the class-BBA is predicted and updated. Object class is then derived by fusing BBAs from both the levels.

The difference here is, instead of a fixed value, the model transition probabilities $t_{o,o'}$ used for the track prediction in GM-MMLMB filter are adapted in every time step, according to the posterior class-BBA of the track. Therefore, the class-BBA $m_C^{(\ell)}$ of the track should represent the likelihood of a motion model $o_i \in \mathbb{O}$, that is more suitable for the class of the object. In order to obtain the probability for each of the classes in the FOD Ω, the class-BBA needs to be transformed from it's credal level to the pignistic level as in (67). As in case of the general Markov transition matrix, with each row covering all the motion models, the transformed probabilities in each row must sum to one. The Markov transition matrix $\chi_{o,o'}$ used for track ℓ is therefore recalculated with each update and is given as

$$\chi^{(\ell)}_{o,o'} = \begin{bmatrix} BetP_{m_C^{(\ell)}}(E^{(\ell)}_{o_1}) & \dots & BetP_{m_C^{(\ell)}}(E^{(\ell)}_{o_N}) \\ \vdots & \ddots & \vdots \\ BetP_{m_C^{(\ell)}}(E^{(\ell)}_{o_1}) & \dots & BetP_{m_C^{(\ell)}}(E^{(\ell)}_{o_N}) \end{bmatrix} \tag{69}$$

As illustrated in [12], the class hypotheses $E^{(\ell)}_{o_i}$ in the above transition probability matrix should cover at least one of the object classes which follow the same motion model $o_i (i \in \{1, \dots, N_o\})$ and all the considered classes need to be covered by the hypotheses $\left\{E^{(\ell)}_{o_i}, \dots, E^{(\ell)}_{o_j}\right\}$ with $E^{(\ell)}_{o_i} \cap E^{(\ell)}_{o_j} = \emptyset$ for $i \neq j$.

The class-BBA of the track on the other side is predicted from time k to $k+1$ independent of the state of the track, by discounting with a parameter d as

$$m^{(\ell)}_{C_+}(E) = \left(m^{(\ell)}_C(E)\right)^d \tag{70}$$

2) Update: The update step of the GM-MMLMB filter considers the various measurement to track association hypotheses, where an hypothesis is given by the mapping $\theta : I_+ \mapsto \{0, 1, \dots, |\mathrm{Z}|\}$, with I_+ representing a predicted set of track labels and Z the measurement set. The update of a track ℓ therefore involves innovations with all the measurements available at time $k+1$. The likelihoods of assignment of a measurement $z_{\theta(\ell)}$ to a track ℓ (given by the case $\theta(\ell) > 0$) and also association of the track to a missed detection (given by the case $\theta(\ell) = 0$) are considered. For track innovation, each measurement along with it's Gaussian distribution also contains the measurement based class-BBA $m_C^{z_{\theta(\ell)}}$ of the object, delivered by the classifier which can be given as

$$z^{(i)} = \left(\mathcal{N}(z; z^{(i)}_+, \underline{\mathrm{R}}^{(i)}), m_C^{z^{(i)}}\right), \; i \in \{1, \dots, |Z|\} \tag{71}$$

On the state level, the measurement $z_{\theta(\ell)}$ updates the spatial distribution of a track ℓ for an association $\theta(\ell)$ by innovation of each of the predicted GMCs of that track. The equations used for state and weight update of GMCs are same as (45) and (49). Additionally, on the class level, the predicted class-BBA $m^{(\ell)}_{C_+}(E)$ is updated by the measurement based class-BBA for an association $\theta(\ell)$ according to the DST rule of combination in (65) as

$$m_C^{(\ell,\theta)}(E) = m_{C_+}^{(\ell)} \oplus m_C^{z_{\theta(\ell)},\alpha} = \frac{\sum\limits_{S_+ \cap S_z = E} m_{C_+}^{(\ell)}(S_+) m_C^{z_{\theta(\ell)},\alpha}(S_z)}{1 - \sum\limits_{S_+ \cap S_z = \emptyset} m_{C_+}^{(\ell)}(S_+) m_C^{z_{\theta(\ell)},\alpha}(S_z)} \; \forall E \in 2^{\Omega} \tag{72}$$

The measurement based class-BBA is discounted with a factor α before being fused with the predicted class-BBA of the track. This discounting facilitates the consideration of confidence in classification output by the classifier. Each update hypothesis therefore has a corresponding updated class-BBA for the track. Further, the weights of the hypotheses $w^{(I_+,\theta)}(\mathrm{Z})$ are updated according to update step of GM-MMLMB described in section III. On the state level, the posterior existence probability, posterior spatial density and model probabilities of a track are calculated according to the equations (36) to (38), which means the posterior of a track is from various hypotheses which include that track in them. On the class level, the class-BBA of a track ℓ is obtained by combining the updated class-BBA from the hypotheses that contains that track. Before combining, the class-BBA from an hypothesis is discounted with the normalised updated weight of that hypothesis and the combination of BBAs is given as

$$m_C^{(\ell)} = \left(m_C^{(\ell,\theta)}\right)^{\widetilde{w}_1^{(I_+,\theta)}(\mathrm{Z})} \oplus \left(m_C^{(\ell,\theta)}\right)^{\widetilde{w}_2^{(I_+,\theta)}(\mathrm{Z})} \cdots \oplus \left(m_C^{(\ell,\theta)}\right)^{\widetilde{w}_N^{(I_+,\theta)}(\mathrm{Z})}, \tag{73}$$

where $(I_+,\theta) \in \mathcal{F}(\mathbb{L}_+) \times \Theta_{I_+}$ and $\ell \in I_+$. The track extraction step follows that of the standard LMB filter family, based on MAP estimate of cardinality distribution or track existence probabilities. The state estimates of a track can be approximated by the mean of it's GMC which has the highest weight. Therefore, this estimated state of the track from the state level can also be used to further update the class-BBA of the track, apart from the update by measurement based class-BBA. One of the track features which can be used to classify the object is for example, it's velocity. The class-BBA can then be constructed based on the velocity of the track and be fused in the same way as the measurement based class-BBA in order to get a posterior class-BBA of the track on the class level.

The updated class-BBA from the above steps is further used for estimating the class of the track. For this the class-BBA of the track needs to be transformed to the pignistic level according to (67). The object class and the corresponding class probability can then be derived as

$$C^{(\ell)} = \max_{e=\Omega} BetP_{m_C^{(\ell)}}(e) \tag{74}$$

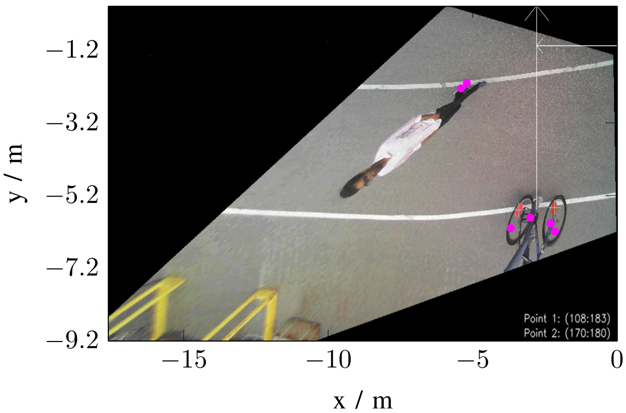
$$p_C^{(\ell)} = BetP_{m_C^{(\ell)}}(C^{(\ell)}). \tag{75}$$

VI. Results

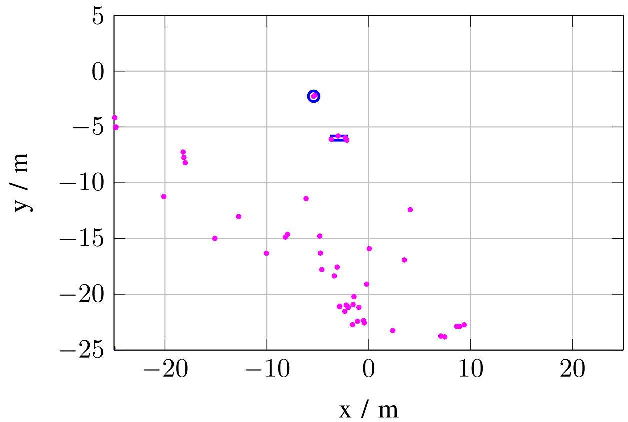
The presented classifying GM-MMLMB filter is evaluated based on recordings from the test vehicle, equipped with fish-eye lens camera, 77 GHz radar sensor and data logging solutions. In the conducted test, the ego-vehicle moves at a speed of 2.8 m/s, starting from a stationary position at the beginning of the test. The objects that move in the sensor FoV are a pedestrian and a bicyclist. The radar detections are processed, clustered and classified according to the methods presented in [13]. Figure 2 illustrates the radar detections of the pedestrian and bicyclist projected on to the calibrated camera image for the time instance $k = 16$ and the corresponding cluster. The pedestrian starts from the rear end of the truck and the bicyclist has a longitudinal and lateral separation of more than 3 m from the pedestrian. The survival probability of the tracks $p_s = 0.99$ and the detection probability of the sensor $p_D = 0.98$ are assumed to be state independent. The clutter measurements follow Poisson distribution with intensity $\kappa(z) = 1.3 \cdot 10^{-3}$, with average rate of $\lambda_c = 2$ clutter measurements per scan. Two motion models, CV and CTRV are defined with standard deviation of the process noise in longitudinal and lateral direction as $\sigma_a = 1$ m/s^2. The standard deviation of the turn rate noise for CTRV is $\sigma_\omega = \pi/180$ rad/s. Further, the model probabilities for the CV and CTRV models are initialised as $[p(o)] = [0.5\ 0.5]$. The class BBAs are discounted with a factor of 0.95, based on the class confidence estimates of the classifier. Additionally the track management parameters are defined such that all the tracks with existence probabilities $r < 0.1$ are pruned. For reducing the computational complexity all the Gaussian components of the track which have weights $w < 0.01$ are pruned and Gaussian components with a Mahalanobis distance $d_{MHD} < 0.2$ are merged.

The pedestrian is initially detected by the radar for certain time steps, but exists the FoV at around time step $k = 37$ as the vehicle moves forward faster than the pedestrian. The pedestrian starts walking faster and catches up with the ego-vehicle at $k = 83$ and starts walking again with almost constant velocity of 1.85 m/s. The bicyclist is within the sensor FoV for the complete test sequence and is tracked continuously. The OSPA distance and cardinality estimate are shown in Figure 3. Reference positions for the bicycle are available for the majority of the sequence. The reference for the pedestrian is not available. However based on the defined trajectory, at the time instances when the pedestrian is detected, the longitudinal positions are approximated by considering a nearly constant relative speed of 0.93 m/s with respect to the ego-vehicle and the lateral position at 1.5 m. The time instances when the reference values from the camera were not available are illustrated by gray background in the plot.

Peaks and higher deviations in OSPA distances are due to the change in cardinality and approximated reference for the pedestrian. The estimated positions of the objects as they start moving are depicted in Figure 4. The strong deviation



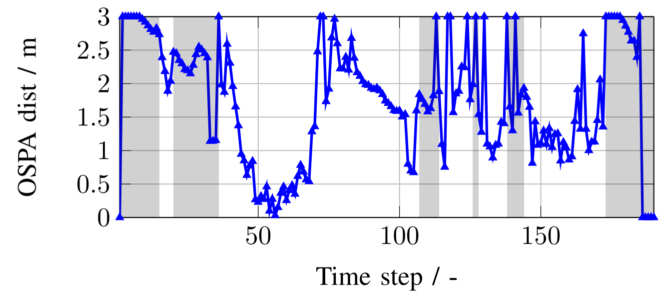
(a) Background subtracted detections projected on to the ground truth camera image.



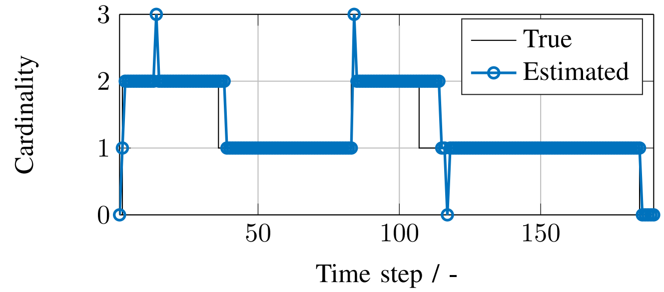
(b) Radar detections and clusters

Fig. 2: Figure (a) depicts the detections of the pedestrian and the bicyclist projected on to the camera image depicted by magenta points corresponding to test scenario. Green cross illustrates the marker position detected from the camera. The right top corner of the image is the origin of the radar. All the detections, including also the ones from background objects are shown in Figure (b). The origin of the radar in Figure (b) is (0,0). The cluster of bicyclist and pedestrian are illustrated as blue bounding box and circle, respectively.

of the bicyclist track at $k = 72$ is due to the imprecise clustering of detections. The reflections from a metal container in the background at this time point is not separable from the bicyclist detection resulting in a wandering reference point. The RMSE values of the bicyclist using the camera detection as reference is shown in Figure 5. The reference velocity is derived from the camera reference positions by smoothing. The ground truth point, when available, is transformed to the object reference point for calculating the RMSE. Time instances when the ground truth were not available are again indicated by a grey background in the plot. Apart from the higher values due to approximated ground truth and imprecise clustering at $k = 72$, the RMSE are low with a mean position error of 0.896 m and mean velocity error of 0.5644 m/s. As

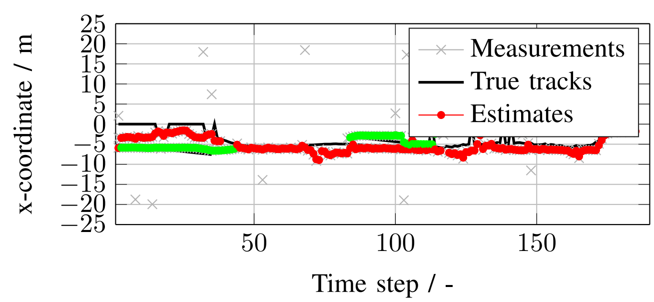


(a) OSPA

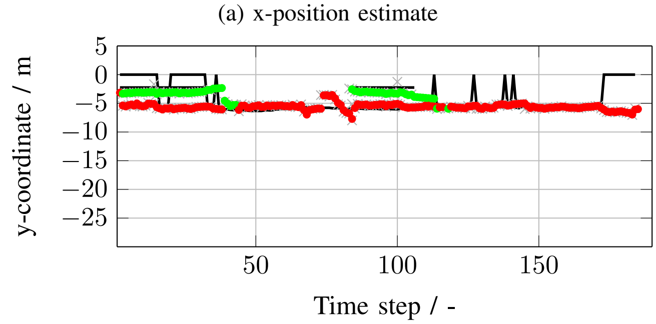


(b) Cardinality

Fig. 3: OSPA distance and cardinality estimates from the test scenario.

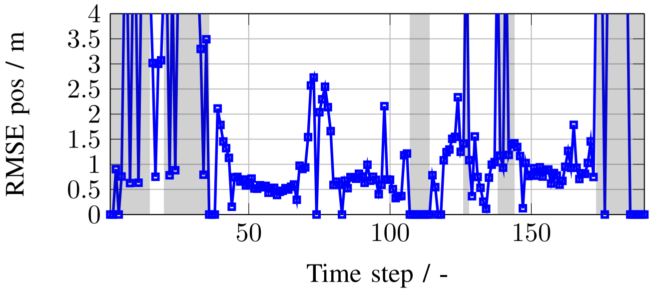


(a) x-position estimate

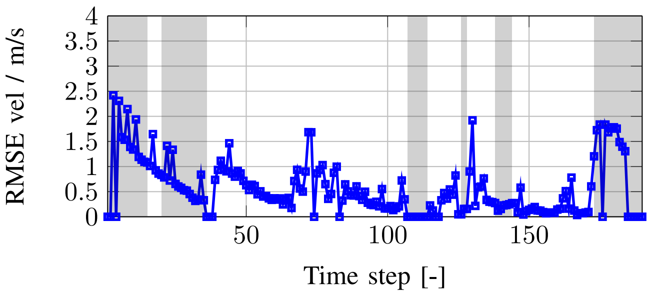


(b) y-position estimate

Fig. 4: Position estimates corresponding to the test scenario.



(a) RMS position error



(b) RMS velocity error

Fig. 5: RMS position and velocity errors of bicyclist corresponding to test scenario.

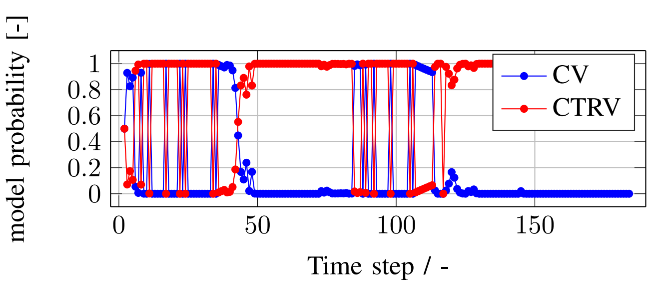


Fig. 6: Model probability of the bicyclist track corresponding to test scenario 2.

the pedestrian most of the time remains towards the rear end of the truck, generates only one or two detection points and it's class BBA is majorly from the velocity part. Consequently, the pedestrian track is smoother than the bicyclist track as it is most of the time tracked as a point target in this scenario. The model probabilities of the bicyclist track are depicted in Figure 6.

VII. Conclusion

This paper presented an approach to integrate the object classification information, typically derived from classical scene understanding or deep learning approaches, into the MMLMB based multi-object tracking framework. Compared to similar solutions proposed for the PHD filters, integrating classification information into the MMLMB filter is expected to increase the robustness and performance further. Equations for the GM implementation of the MMLMB are detailed and a method to predict and update the state and class of the track is presented. The performance of the classifying MMLMB filter based on clinical test scenario, but with real sensor data is analysed.

Moreover the classifying GM-MMLMB filter was compared to an existing implementation of a classical tracking algorithm based on linear KF and greedy nearest neighbour data association. The state estimates from the classifying GM-MMLMB shows a significant performance improvement in terms of RMSE and cardinality. This is majorly due to the switch to the suitable motion model based on object class and the ability to keep multiple hypotheses in the GM implementation. However, detailed performance evaluations and comparison to the MMLMB without classification information based on multiple scenarios and classification based on bigger multimodal data sets are some future work to do.

References

[1] M. Munz, "Generisches Sensorfusionsframework zur gleichzeitigen Zustands-und Existenzschätzung für die Fahrzeugumfelderkennung," PhD dissertation,Ulm University, 2011.

[2] R. Georgescu, and P. Willet, "Classification aided cardinalized probability hypothesis density filter," In: Signal Processing, Sensor Fusion, and Target Recognition XXI, 2012.

[3] D. Meissner, S. Reuter, E. Strigel, and K. Dietmayer, "Road User Tracking Using a Dempster-Shafer based Classifying Multiple-Model PHD Filter," In: 16th International Conference on Information Fusion, Istanbul, Turkey, 2013.

[4] O. Erdinc, P. Willett and Y. Bar-Shalom, "Probability hypothesis density filter for multitarget multisensor tracking." In: Proceedings of the 8th International Conference on Information Fusion, pages 1–8, 2005.

[5] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate prior", IEEE Transactions on Signal Processing, vol. 61,no. 13, pp. 3460–3475, 2013.

[6] 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.

[7] S. Reuter, A. Scheel and K. Dietmayer, "The Multiple Model Labeled Multi-Bernoulli Filter," In: Proceedings of the 18th International Conference on Information Fusion, Washington DC, 2015.

[8] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments", In: IEEE Transactions on Aerospace and Electronic Systems, vol. 39, no. 4, pp. 1152 - 1178, 2003.

[9] R. Mahler, "Statistical Multisource-Multitarget Information Fusion", Artech House Inc., Norwood, 2007.

[10] B.-T. Vo and B.-N. Vo, "Tracking, identification, and classification with random finite sets," In: Proceedings of the SPIE 8745, Signal Processing, Sensor Fusion, and Target Recognition XXII. International Society for Optics and Photonics, 2013.

[11] P. Smets and R. Kennes, "The transferable belief model," In: Artificial Intelligence, vol. 66, no. 2, pages 191–234, 1994.

[12] D. Meissner, "Intersection-Based Road User Tracking Using a Classifying- Multiple-Model PHD Filter," PhD dissertation, Ulm University, 2015.

[13] G. Nageswaran, "High-resolution radar based object classification for automotive applications," online publication, KLUEDO, TU Kaiserslautern, 2016.

[14] R. Mahler, "On multitarget jump-Markov filters," In: Proceedings of the 15th International Conference on Information Fusion, pages 149–156, 2012.