Bernoulli merging for the Poisson multi-Bernoulli mixture filter

Marco Fontana*, Ángel F. García-Fernández*°, Simon Maskell*
*Dept. of Electrical Engineering and Electronics, University of Liverpool, United Kingdom
°ARIES Research Center, Universidad Antonio de Nebrija, Spain
Emails: {marco.fontana, angel.garcia-fernandez, s.maskell}@liverpool.ac.uk

***Abstract*—Under the standard multiple target tracking models and a Poisson point process birth model, the Poisson multi-Bernoulli mixture (PMBM) filter provides the closed-form recursion to computing the posterior density over the set of targets. Without approximations, the PMBM computational complexity rapidly rises in time due to the increasing number of data association hypotheses. This paper presents innovative strategies for merging Bernoulli components for the same potential target reducing the number of single-target hypotheses in the PMBM filter, aiming to lower its computational complexity while keeping its performance high. We use several measures to compute the similarity between different Bernoulli components. Simulation results show that the proposed algorithms show performance close to the PMBM filter without Bernoulli merging, as measured by the generalized optimal sub-pattern assignment (GOSPA) metric, with a significantly reduced execution time.**

***Index Terms*—Multiple target tracking, Poisson multi-Bernoulli mixtures, Bernoulli merging.**

I. Introduction

One aim of multiple target tracking (MTT) is to estimate the number of targets and their states for each observation time [1]. The estimation of the target states is based on noisy sensor measurements acquired in a dynamic scenario, where the number of targets varies due to appearance and disappearance of targets in the field of view of the sensor.

Nowadays, there are several applications which rely on this kind of algorithm, ranging from surveillance to traffic control and autonomous driving [2]. In these contexts, MTT algorithms have to deal with complex sources of uncertainty, as false alarms and missed detections, and possibly with multiple heterogeneous sensors.

In the last decades, three main techniques have been successfully investigated and implemented in real-world applications: joint probabilistic data association (JPDA) filter [3], multiple hypothesis tracking (MHT) [4]–[6], and random finite sets (RFS) [7].

In RFS-based MTT, the multi-target state is represented as a finite set of single-target states. The tracking problem is expressed as a dynamic multi-target state estimation problem, where the (multi-target) filtering density can be computed via the prediction and the update steps of the Bayesian filtering recursion. As this computation results intractable, multi-target conjugate prior densities are a popular solution to compute or approximate the filtering density. In MTT context, a multi-target density is conjugate with respect to a dynamic model if the posterior (multi-target) distribution has the same functional form as the prior [8], [9]; this conjugacy property allows the posterior to be written in terms of single target densities.

The Poisson multi-Bernoulli mixture (PMBM) [10] is a multi-target conjugate prior in which the set of targets is divided between detected and undetected targets. The latter subset is represented by the Poisson part, in the case of a Poisson RFS birth model. Moreover, the components in the mixture represent the global hypotheses, each one weighted and associated with a multi-Bernoulli density. The number of mixture components grows in time, due to increasing data association uncertainty.

The PMBM filter has been applied to deal with continuous-discrete systems [11] and in a range of applications, including mapping [12], simultaneous localization and mapping [13], and joint target sensor state tracking [14] and camera tracking [15]. In the literature, the PMBM filter has been implemented by pruning global hypotheses with low weight, which results in a bounded L1-error [16, Sec. V.D], and discarding Bernoulli components whose probability of existence is below a threshold [16], [17]. Merging of multi-Bernoulli components has been used in PMBM filtering in [16], and it also appears in the Poisson multi-Bernoulli filters, where only a single multi-Bernoulli component is propagated [10], [18].

In this paper, we propose a more flexible approach to merging multi-Bernoulli densities for a given potential target. In particular, we focus on the case in which the single target Bernoulli densities are Gaussian [17]. The proposed algorithm clusters the single-target densities using the current data association for each time instant, and merges the Bernoulli distribution within these clusters. Furthermore, the Bernoulli clusters are merged together if the distance between them is under a specific threshold. We use and evaluate the following distances between Bernoulli components: Kullback-Leibler (KL) divergence [7], L_2 norm [19], and Weighted Kullback-Leibler (WKL) divergence [20]. Reducing the number of single-target hypotheses contributes to the decrease in the number of global hypotheses, which in turn lowers the execution time.

The paper is organized as follows. Background on RFS, Bayesian filtering, and the PMBM filter is provided in Section II. Section III discusses different similarity functions to compare Bernoulli distributions, and Section IV presents the algorithm to perform a fast and efficient merging of the single-target distributions. Simulation results are presented in Section V, and conclusion are drawn in Section VI.

II. Background and Problem Formulation

In this section, we present the notation and the main concepts underlying the Bayesian filtering recursion and, more specifically, on the PMBM filter and the merging operation.

A. Bayesian Filtering Recursion

As already mentioned, a main purpose of multi-target tracking is the estimation of the states of a time-varying number of targets. In RFS-based methods, the system at time k is described by a set of states $X_k \in \mathcal{F}(\mathcal{X})$, where $\mathcal{X}$ is the single target space, and $\mathcal{F}(\mathcal{X})$ is the collection of all finite subsets of $\mathcal{X}$.

At each time step, targets survive with probability of survival $p_S(x)$, or they die with probability $1 - p_S(x)$. The evolution of a survived target state $x_k \in X_k$ can be defined by a Markov transition density $g(x_k|x_{k-1})$, where $g(\cdot|x)$ is a density on $\mathcal{X}$ for each $x \in \mathcal{X}$; i.e. a target state at the next time step only depends on their current state. Several dynamic models are commonly used in different applications, as nearly constant velocity, nearly constant acceleration, nearly coordinated turn, and many others [21]. The multi-target state at time step $k+1$, which is denoted as X_{k+1}, is the union of the surviving targets and the new targets, which are modeled by a Poisson Point Process (PPP).

The set of measurements at time k is denoted by $Z_k \in \mathcal{F}(\mathbb{R}^{n_z})$, and it includes PPP clutter and target-generated measurements with unknown origin. At each time step, existing targets are detected with probability of detection $p_D(x)$, or misdetected with probability $1 - p_D(x)$. Each detected target $x_k \in X_k$ generates a measurement which is only conditioned on its corresponding target; the single target measurement likelihood is defined as $l(z_k|x_k)$.

In Bayesian filtering, the information about the target states is contained in the (multi-target) density. The posterior density $p_{0:k}(\cdot|Z_{1:k})$, where $Z_{1:k}$ is the observation history, contains the information about the state history to time k. It can be computed recursively starting from a prior $p_0(\cdot)$ through the usual prediction and update steps of the Bayesian filtering recursion. In the PMBM filtering recursion, the conjugate prior properties can be used to design a simplified computational process [10], [17].

B. Poisson Multi-Bernoulli Mixture Filter

The Poisson Multi-Bernoulli (PMBM) density $f_{k'|k}(\cdot)$ of the set of targets at time step k' given measurements up to time step k results from the combination of two independent RFSs: a Poisson RFS with density $f^p_{k'|k}(\cdot)$, and a multi-Bernoulli mixture (MBM) RFS with density $f^{mbm}_{k'|k}(\cdot)$, where $k' \in \{k, k+1\}$. The union of these two RFSs has been proved to be a conjugate prior with respect to the standard point target measurement model [10], [17]. The PMBM density is expressed over the union of the undetected targets set Y, and the detected target set W:

$$f_{k'|k}(X_{k'|k}) = \sum_{Y \uplus W = X} f^p_{k'|k}(Y) f^{mbm}_{k'|k}(W) \tag{1}$$

where the sum goes over all mutually disjoint sets Y and W, which have the property that their union is X.

The Poisson RFS density represents the targets that exist at the current time instant, but have not yet been detected; it is defined as

$$f^p_{k'|k}(X) = e^{-\int D_{k'|k}(x)dx} \prod_{x \in X} D_{k'|k}(x) \tag{2}$$

where $D_{k'|k}(\cdot)$ is the intensity of the Poisson RFS. In the PPP, the cardinality is Poisson distributed and targets are independent, and identically distributed. The MBM part represents the potentially detected targets, and it can be described as [10]:

$$f^{mbm}_{k'|k}(X) = \sum_{a \in \mathcal{A}_{k'|k}} w^a_{k'|k} \sum_{\uplus_{j=1}^{n_{k'|k}} X^j = X} \prod_{i=1}^{n_{k'|k}} f^{i,a^i}_{k'|k}(X^i) \tag{3}$$

where i is the index over the Bernoulli components, $a = \{a^1, \ldots, a^{n_{k'|k}}\} \in \mathcal{A}_{k'|k}$ represents a specific data association hypothesis, $a^i \in \{1, \ldots, h^i_{k'|k}\}$ is an index over the $h^i_{k'|k}$ single target hypotheses for the i-th potential target/Bernoulli component, and $n_{k'|k}$ is the number of potentially detected targets. Each set of single target hypothesis $a \in \mathcal{A}_{k'|k}$ is also called a global hypothesis, and it is associated to a weight $w^a_{k'|k}$ satisfying $\sum_{a \in \mathcal{A}_{k'|k}} w^a_{k'|k} = 1$. The same single target hypotheses can appear in several global hypotheses.

The Bernoulli density corresponding to the i-th potential target and the a_i single target hypothesis $f^{i,a^i}_{k'|k}(X)$ can describe a newly detected target, or it can represent a previously detected target or clutter; it makes it possible to efficiently model both the uncertainty regarding target existence and state. Mathematically, it can be expressed as

$$f^{i,a_i}_{k'|k}(X) = \begin{cases} 1 - r^{i,a^i}_{k'|k} & X = \emptyset \\ r^{i,a^i}_{k'|k} p^{i,a^i}_{k'|k}(x) & X = \{x\} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $r^{i,a^i}_{k'|k} \in [0,1]$ is the probability of existence and $p^{i,a^i}_{k'|k}(\cdot)$ is the state density given that it exists.

In this work, we consider the Gaussian implementation proposed in [17], where $p^{i,a^i}_{k'|k}(x) = \mathcal{N}(x; \mu^{i,a^i}_{k'|k}, P^{i,a^i}_{k'|k})$, with mean $\bar{x}^{i,a^i}_{k'|k}$ and variance $P^{i,a^i}_{k'|k}$. In this context, the MBM is entirely defined by the following parameters:

$$\{(w^a_{k'|k}, r^{i,a^i}_{k'|k}, \mu^{i,a^i}_{k'|k}, P^{i,a^i}_{k'|k})\}_{i \in \{1,\ldots,n_{k'|k}\}} \tag{5}$$

where $a \in \mathcal{A}_{k'|k}$ is defined above.

C. Merging Bernoulli components

In this paper, we will consider merging strategies at a Bernoulli level; i.e., we propose to merge Bernoulli local hypotheses for the same potential target i. That is, for a given i, we find values of a_i such that the associated Bernoullis[1] are sufficiently similar and are thus merged. Therefore, the MBM posterior is approximated as:

$$f^{mbm}_{k'|k}(X) = \sum_{a \in \mathcal{B}_{k'|k}} \widetilde{w}^a_{k'|k} \sum_{\uplus_{l=1}^{n_{k'|k}} X^l = X} \prod_{i=1}^{n_{k'|k}} \widetilde{f}^{i,a^i}_{k'|k}(X^i) \tag{6}$$

[1]We use the term Bernoullis and Bernoulli components interchangeably.

Table I: Global hypothesis matrix of Example 1 before merging algorithm

	Target $i=1$	Target $i=2$
Global hypothesis 1	$a^1=1$	$a^2=2$
Global hypothesis 2	$a^1=2$	$a^2=1$

Table II: Global hypothesis matrix of Example 1 after merging algorithm

	Target $i=1$	Target $i=2$
Global hypothesis 1	$a^1=1$	$a^2=1$
Global hypothesis 2	$a^1=2$	$a^2=1$

where $\mathcal{B}_{k'|k}$ are the considered global hypotheses, which include indices to each Bernoulli, $n_{k'|k}$ is the considered number of Bernoulli components, $a \in (a^1,\dots,a^{n_{k'|k}}) \in \mathcal{B}_{k'|k}$ represents a specific data association hypothesis, and $a^i \in \{1,\dots,h^i_{k'|k}\}$ is an index over the $h^i_{k'|k}$ single target hypotheses for the i-th Bernoulli component. The tilde on the Bernoulli distributions $\widetilde{f}^{i,a^i}_{k'|k}$ and the weights $\widetilde{w}^a_{k'|k}$ highlights that the components of the merged mixture can be different compared to those in the original mixture.

After merging, the MBM posterior is approximated with another MBM posterior with a fewer number of local hypotheses per Bernoulli component, as similar Bernoullis are merged into one. For example, if $f^{1,1}_{k'|k}(X^i)$ is considered similar to $f^{1,2}_{k'|k}(X^i)$, we obtain a new $\widetilde{f}^{1,1}_{k'|k}(X^i)$ that merges the information in $f^{1,1}_{k'|k}(X^i)$ and $f^{1,2}_{k'|k}(X^i)$. Then, each local hypotheses $a=(a^1,...,a^{n_{k'|k}}) \in \mathcal{B}_{k'|k}$ such that $a^1=2$ is modified such that $a^1=1$. We proceed to illustrate this with an example.

Example 1. Consider an MBM posterior with $|\mathcal{B}_{k'|k}|=2$ global hypotheses, and the index over the Bernoulli components $i \in \{1,2\}$, i.e. there are two Bernoullis. Let us suppose that each potential target is associated with two local hypotheses, i.e. $h^1_{k'|k}=h^2_{k'|k}=2$. The filter status can be represented by the global hypothesis matrix, where columns indicate targets and rows represent global hypotheses. For each potential target i, the correspondent column indicates the index a^i of the single target hypotheses, i.e. the Bernoullis associated to that track at the time $t=k$. Table I and Table II show the global hypothesis matrix before and after merging. In this case, the algorithm merges the two Bernoulli components associated to the potential target $i=2$, so the corresponding target's status is expressed with just a single target hypothesis.

In order to better understand the benefits of the proposed algorithm, let us consider this second example.

Example 2. Consider the track of a one-dimensional target represented in Fig. 1. The target's position is estimated at each time instant $k=0,1,2,3,4$ based on the measurements received from the detector. In the figure, the stars represent the position provided by the detector, while the circles which contain the stars indicate the Bernoulli that has been updated with that measurement. As we can see, the target's status is represented by a single Bernoulli up to $k=1$. At $k=2$, the detector produces two measurements, thus the target is associated with two hypotheses, $f^{1,1}_2$ and $f^{1,2}_2$. Then, in the next time instant, the detector provides a single measurement,

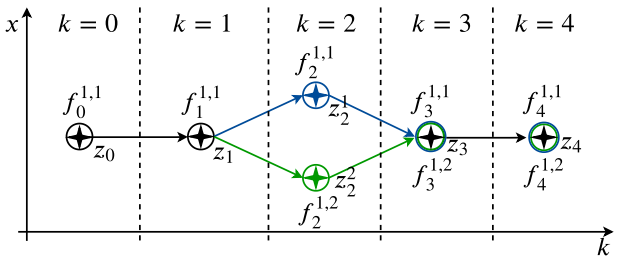

Figure 1: Example of a superposition of Bernoulli distributions. The stars indicate the measurements z_k at time k, while the circles represent the Bernoullis RFSs associated to the corresponding measurement at each time step.

which is associated to both the Bernoullis. As depicted in Fig. 1, $f^{1,1}_3$ and $f^{1,2}_3$ are quite similar, and they represent almost the same target state; the difference between them is due just to a different data association in the past, at $k=2$.

In Section IV we will present our algorithm, which aims to reduce the number of redundant single-target hypotheses taking the insight of Example 2 into account.

III. Distances and Merging of Bernoulli Distributions

In this section, we introduce some key components of the proposed merging algorithm. Firstly, we report a method to merge a subset of Bernoulli components belonging to a mixture, via moment-matching of the sufficient statistics. Then, we present a set of distances to evaluate the similarity between Bernoulli distributions in Eq. (3) through closed-form expressions.

A. Merging of Bernoulli densities

Let us consider a potential target i and the subset of global hypotheses $\mathcal{M}_{k'|k} \subseteq \mathcal{A}_{k'|k}$ in which it is supposed to exist. Each global hypothesis $a=(a^1,\dots,a^{n_{k'|k}})$, $a \in \mathcal{M}_{k'|k}$, consists of one single-target hypothesis of the target i, $f^{i,a^i}_{k|k}(X^i)$, which can appear several times through the subset $\mathcal{M}_{k'|k}$. The potential target state can be described by

$$f^i_{k'|k}(X^i) = \sum_{a \in \mathcal{M}_{k'|k}|a^i} W^{a^i}_{k'|k} f^{i,a^i}_{k|k}(X^i) \tag{7}$$

where $W^{a^i}_{k'|k}$ is the sum of the weights associated to the global hypotheses in which a specific Bernoulli component $f^{i,a^i}_{k|k}$ appears, i.e.

$$W^{a^i}_{k'|k} = \sum_{a \in \mathcal{M}_{k'|k}|a^i} w^a_{k'|k} \ . \tag{8}$$

Suppose $p^{i,a^i}_{k|k}(x)$ the single target density of $f^{i,a^i}_{k|k}(X^i)$ is Gaussian, e.g. $p^{i,a^i}_{k'|k}(x) = \mathcal{N}(x;\mu^{i,a^i}_{k'|k}, P^{i,a^i}_{k'|k})$. Assume that we aim to merge the Bernoulli mixture $f^i_{k'|k}(X^i)$ into a single Bernoulli density $\widetilde{f}^i_{k'|k}(X^i)$, with single target density

$\widetilde{p}^{i,a^i}_{k'|k}(x) = \mathcal{N}(x; \widetilde{\mu}^i_{k'|k}, \widetilde{P}^i_{k'|k})$. The approximated Bernoulli density $\widetilde{f}^i_{k'|k}\left(X^i\right)$ is expressed by [10], [18]:

$$\widetilde{f}^i_{k'|k}\left(X^i\right) = \begin{cases} 1 - \widetilde{r}^i_{k'|k} & X^i = \emptyset \\ \widetilde{r}^i_{k'|k}\mathcal{N}(x; \widetilde{\mu}^i_{k'|k}, \widetilde{P}^i_{k'|k}) & X^i = \{x\} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where

$$\widetilde{r}^i_{k'|k} = \frac{\sum_{a\in\mathcal{M}} W^a_{k'|k} r^{i,a^i}_{k'|k}}{\sum_{a\in\mathcal{M}} W^a_{k'|k}} \tag{10}$$

$$\widetilde{\mu}^i_{k'|k} = \frac{\sum_{a\in\mathcal{M}} W^a_{k'|k} r^{i,a^i}_{k'|k} \mu^{i,a^i}_{k'|k}}{\sum_{a\in\mathcal{M}} W^a_{k'|k} r^{i,a^i}_{k'|k}} \tag{11}$$

$$\begin{aligned}\widetilde{P}^i_{k'|k} = &\frac{\sum_{a\in\mathcal{M}} W^a_{k'|k} r^{i,a^i}_{k'|k} (P^{i,h^i}_{k'|k} + \mu^{i,a^i}_{k'|k}(\mu^{i,a^i}_{k'|k})^T)}{\sum_{a\in\mathcal{M}} W^a_{k'|k} r^{i,a^i}_{k'|k}} \\ &- \widetilde{\mu}^i_{k'|k}(\widetilde{\mu}^i_{k'|k})^T \ .\end{aligned} \tag{12}$$

B. Distance between Bernoulli distributions

We aim to evaluate the similarity between two Bernoulli distributions in Eq. (3), in order to detect those which can be approximated with a new Bernoulli, resulting from the merging. In this section, we revise the considered distances and provide the closed-form solutions for Gaussian single target densities.

Given two multi-target densities $f(X)$ and $g(X)$, the Kullback-Leibler (KL) divergence of $g(X)$ from $f(X)$ is defined as [22]:

$$\mathrm{D}_{KL}(f(X) \,\|g(X)) = \int f\left(X\right) \log \frac{f\left(X\right)}{g\left(X\right)} \delta X \tag{13}$$

where the integral in Eq. (13) is a set integral, defined in [7, Section 3.3].

The closed-form of the KL divergence between two Bernoulli distributions is presented in Lemma 3; the proof is available in the Appendix.

Lemma 3 (Kullback-Leibler divergence)**.** *Let $f_1(X)$ and $f_2(X)$ be two Bernoulli RFS distributions with Gaussian single target densities. The i-th Bernoulli RFS has probability of existence r_i, mean $\bar{x}_i$, and covariance matrix P_i. If $r_2 \notin \{0, 1\}$, the Kullback-Leibler (KL) divergence of f_2 from f_1 exists and it is a finite value, given by:*

$$\begin{aligned}&\mathrm{D}_{KL}(f_1(X) \,\|f_2(X)) \\ &\quad = (1 - r_1) \log \frac{1 - r_1}{1 - r_2} + r_1 \log \frac{r_1}{r_2} \\ &\qquad + \frac{r^1}{2}\left[\mathrm{tr}\left(\left(P_2\right)^{-1} P_1\right) - \log\left(\frac{|P_1|}{|P_2|}\right) - n_x \right. \\ &\qquad \left. + \left(\overline{x}_2 - \overline{x}_1\right)^T \left(P_2\right)^{-1} \left(\overline{x}_2 - \overline{x}_1\right)\right]\end{aligned} \tag{14}$$

If $r_1 = r_2 \in \{0, 1\}$, the KL divergence is:

$$\begin{aligned}&\mathrm{D}_{KL}(f_1(X) \,\|f_2(X)) \\ &\quad = \frac{r^1}{2}\left[\mathrm{tr}\left(\left(P_2\right)^{-1} P_1\right) - \log\left(\frac{|P_1|}{|P_2|}\right) - n_x \right. \\ &\qquad \left. + \left(\overline{x}_2 - \overline{x}_1\right)^T \left(P_2\right)^{-1} \left(\overline{x}_2 - \overline{x}_1\right)\right] .\end{aligned} \tag{15}$$

Another distance used for Gaussian mixture reduction is based on L_2 norm, or Integrated Squared Difference (ISD) [23]. The derivation for two Bernoulli RFS distributions is presented in the next lemma.

Lemma 4 (L_2 norm)**.** *Let $f_1(X)$ and $f_2(X)$ be two Bernoulli RFS distributions with Gaussian single target densities. The i-th Bernoulli RFS has probability of existence $r_i \in [0, 1]$, mean $\bar{x}_i$, and covariance matrix P_i. The L_2 norm (or Integrated Squared Difference, ISD) is defined as [19]:*

$$\begin{aligned}&D_{L_2}(f_1(X) \,\|f_2(X)) \\ &\quad = \int \frac{1}{K^{|X|}} (f_1(X) - f_2(X))^2 \delta X \\ &\quad = (r_2 - r_1)^2 + K\Big[r_1^2 \mathcal{N}\left(\overline{x}_1; \overline{x}_1, 2P_1\right) \\ &\qquad + r_2^2 \mathcal{N}\left(\overline{x}_2; \overline{x}_2, 2P_2\right) - 2r_1 r_2 \mathcal{N}\left(\overline{x}_1; \overline{x}_2, P_1 + P_2\right)\Big]\end{aligned} \tag{16}$$

where $|X|$ denotes the cardinality of the set X, and K represents the units of hypervolume of the state space $\mathcal{X}$.

It should be noted that the factor $1/K^{|X|}$ should be introduced in the L_2 norm definition due to how the set integral is defined and its relation to the measure theoretic integral [19]. If the state is assumed unitless, it is not necessary to introduce it. In addition, the units of $N(\cdot; \overline{x}, P)$ are $1/K$ so there is a multiplication by K to have consistent units.

Some mixture reduction techniques rely on distance metrics based not only on the parameters of the components, but also on the weights of those in the mixture [20], [24]. Unfortunately, it is not always possible to derive a closed-form equation to compute this kind of similarities. In [20], the author provides an upper bound for the Kullback-Leibler divergence between two Gaussian components in a mixture. Analogously, we define a new divergence between two Bernoullis RFS distributions based on Eq. (13).

Definition 5 (Weighted Kullback-Leibler divergence)**.** Let $f_1(X)$ and $f_2(X)$ be two Bernoulli RFS distributions in a mixture $f(X) = w_1 f_1(X) + w_2 f_2(X)$, with Gaussian single target densities. Suppose w_1 the weight associated with $f_1(X)$, and w_2 the weight associated with $f_2(X)$. The Weighted Kullback-Leibler (WKL) divergence between two weighted components $w_i f_i$ in a mixture f is defined as

$$\begin{aligned}&D_{WKL}(w_1 f_1(X) || w_2 f_2(X)) \\ &\quad = w_1 \mathrm{D}_{KL}\left(f_1(X) \,\|f_{12}(X)\right) + w_2 \mathrm{D}_{KL}\left(f_2(X) \,\|f_{12}(X)\right)\end{aligned} \tag{17}$$

where $f_{12}(X)$ is the resulting Bernoulli RFS distribution from the merge of $f_1$and f_2, which has been presented in Eq. (9).

Lemma 6. *Suppose $w_1 + w_2 = 1$. The Weighed Kullback-Leibler divergence is an upper bound of the Kullback-Leibler divergence, that is:*

$$D_{KL}(w_1 f_1(X) + \bar{w}_2 f_2(X)||f_{12}(X)) \leq D_{WKL}(w_1 f_1(X))||(w_2 f_2(X)) \ .$$

The proof of Lemma 6 is available in Appendix.

IV. Merging Algorithm

In this section, we present an algorithm to efficiently merge the Bernoulli components mentioned in Subsection II-C. The pseudo-code of the proposed algorithm is presented in Alg. 1.

Let us consider the Bernoulli $f^{i,a^i}_{k|k}$ associated to the i-th potential target through all the $N_{k|k} = |\mathcal{A}_{k|k}|$ global hypotheses, with $a^i \in \{1, \dots, h^i_{k|k}\}$ index over the $h^i_{k|k}$ single target hypotheses. Furthermore, we consider the measurement set at time step k as $Z_k = \{z^1_k, \dots, z^{m_k}_k\}$. Each Bernoulli $f^{i,a^i}_{k|k}$ has been updated with one of the m_k measurements, or with a misdetection; thus, $h^i_{k|k} = h^i_{k|k-1}(m_k + 1)$ [10]. This process leads to an exponential increase of the number of single target hypotheses, especially in high-clutter scenarios, where m_k is higher.

Now let us consider the $h^i_{k|k-1}$ single target hypotheses for a particular potential target which have been updated with the measurement z^j_k, $j \in \{1, \dots, m_k\}$. These Bernoulli components differ just for the data association history, but they usually represent quite similar target states at time k. Thus, it is possible to summarize the information carried by each single target hypotheses into a single Bernoulli component, defined as the merge of the $h^i_{k|k-1}$ Bernoullis components associated with z^j_k, and denoted as $\widetilde{f}^{i,j}_{k|k}$. The merging of the Bernoullis is performed through moment matching of the sufficient statistics, presented in Eq. (9-12) [10], [18].

The output of this procedure is a set of mixture components S, which consist of m_k single target hypotheses resulting from the merging algorithm, $h^i_{k|k-1}$ Bernoullis associated with a misdetection hypothesis, and their relative weights in the mixture. More specifically, the weight $\widetilde{W}^j_{k|k}$ associated to the merged Bernoulli component $\widetilde{f}^{i,j}_{k|k}$ is defined as $\widetilde{W}^j_{k|k} = \sum_{a^i \in \{1,\dots,h^i_{k|k}\}} W^{a^i}_{k|k}$, see Eq. (8), as we assume that all the $h^i_{k|k}$ hypotheses have been updated with the measurement z^j.

At this point, the algorithm called "Compare and Merge" computes the distance between all the Bernoulli components of the i-th potential target; if the distance is below a certain threshold, the Bernoullis are merged as well.

The Compare and Merge algorithm performs a greedy merge procedure, which is similar to what has been proposed by Runnals in [20], but with some key differences. Firstly, the procedure considers the distances between all the elements of the Bernoulli set, and merges the two most similar Bernoullis at each iteration, i.e. those which show the minimum distance. The merging is performed only if the distance is below a specified threshold Γ. Otherwise, the algorithm breaks the cycle and returns the current set of Bernoullis. Secondly, we use several distances, presented in Section III, to compute the similarity between pairs of Bernoulli components.

The pseudo-code of the Compare and Merge algorithm is presented in Alg. 2. For simplicity, we denote the Bernoulli components by f, and we assume to compute the distance between two mixture components $(w_i, f_i)_{i\in\{1,2\}}$. We compute the distance taking into consideration their ranking according to their weight, in order to uniquely define the algorithm for non-symmetric divergences, e.g., the KLD. Even though we take the weights into account to perform the Bernoulli merging, the corresponding global hypothesis weights are not modified, unless there are duplicate global hypotheses.

Algorithm 1 Merging algorithm

Input: $\{(w^a_{k|k}, \{f^{i,a^i}_{k|k}(X^i)\}_{i\in\{1,\dots,n_{k|k}\}})\}_{a\in\mathcal{A}_{k|k}}$
Output: $\{(\widetilde{w}^a_{k|k}, \{\widetilde{f}^{i,a^i}_{k|k}(X^i)\}_{i\in\{1,\dots,n_{k|k}\}})\}_{a\in\mathcal{B}_{k|k}}$
for $i \in \{1, \dots, n_{k|k}\}$ **do**
  $S \leftarrow h^i_{k|k-1}$ Bernoullis and relative weights associated with a misdetection hypothesis
  for $\forall\ z^j_k,\ j \in \{1, \dots, m_k\}$ associated with the target i at time k **do**
    $\widetilde{f}^{i,j}_{k|k} \leftarrow$ Merge the $h^i_{k|k-1}$ Bernoullis with data association z^j_k, see Eq. (9)
    $S \leftarrow S \cup \{(\widetilde{W}^j_{k|k}, \widetilde{f}^{i,j}_{k|k})\}$
  end for
  $\{\widetilde{f}^{i,a^i}_{k|k}\} \leftarrow$ Run the Alg. 2 on S
  Reindex the entries of the global hypothesis matrix corresponding to merged Bernoullis, see Example 1
end for
$\{\widetilde{w}^a_{k|k}\}_{a\in\mathcal{B}_{k|k}} \leftarrow$ Delete the duplicate rows in the global hypothesis matrix and sum their global hypotheses weights

Algorithm 2 Compare and Merge algorithm

Input: $\{S\}$ **Output:** $\{S\}$
$c \leftarrow 1$
while c **do**
  for $\forall$ unordered pair $\{(W, f_W), (w, f_w)\} \in S$ **do**
    Compute $D((W, f_W)||(w, f_w))$, $W \geq w$
▷ D is one of the distances in Eq. (14-15), (16), and (17)
  end for
  $(w_1, f_1), (w_2, f_2) \leftarrow$ pair of components with minimum distance
  if $D((w_1, f_1)||(w_2, f_2)) < \Gamma$ **then**
    $(w_M, f_M) \leftarrow$ merge f_1 and f_2, see Eq. (9)
    $S \leftarrow S \cup \{(w_M, f_M)\}$
    $S \leftarrow S \backslash \{(w_1, f_1), (w_2, f_2)\}$
  else
    $c \leftarrow 0$
    return S
  end if
end while

V. Results

We proceed to assess the accuracy and computational time of the proposed Bernoulli merging strategies. The simulations of the algorithm presented in this work are based on the

scenario used in [17]. It consists of four targets, all born at time step 1 and alive throughout the simulation, except one which dies at time step 40 (the blue one in Fig. 2). The scenario is considered challenging, as the targets all get close at time step 40, when the blue one dies.

In this simulation, target motion follows a nearly constant velocity model. The target state is described in a two-dimensional Cartesian coordinate system by $s_k = [p_{x,k}, v_{x,k}, p_{y,k}, v_{y,k}]^T$, where the first two components represent position and velocity of the target on the x-axis, and the last two those on the y-axis. The targets are born according to a Poisson Point Process of intensity 3 at the first time step, and 0.005 at the next time steps. The intensity is Gaussian with mean $[100, 0, 100, 0]^T$, and covariance $\mathrm{diag}([150^2, 1, 150^2, 1])$, which covers the considered area of $[0, 300] \times [0, 300]$. The parameters of the linear and Gaussian motion and measurement models are

$$F = I_2 \otimes \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}, \quad Q = qI_2 \otimes \begin{pmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{pmatrix}$$
$$H = I_2 \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad R = I_2$$

where $\otimes$ is the Kronecker product, $q = 0.01$, and $T = 1$ is the sampling period.

In this simulation, we set the probability of survival of the targets $p_S = 0.99$, and the probability of detection $p_D = 0.9$. The clutter model is Poisson, uniformly distributed in the area of interest, with a mean number of clutter measurements per scan equal to λ.

The standard PMBM filter implementation has maximum number of global hypotheses $N_h = 200$, threshold for pruning the Poisson components $\Gamma_p = 10^{-5}$, threshold for pruning global hypotheses $\Gamma_{mbm} = 10^{-4}$, and threshold for pruning Bernoulli components $\Gamma_b = 10^{-5}$. The ellipsoidal gating is performed with a threshold equal to 20, while the estimation is performed selecting the global hypothesis with the highest weight and reporting Bernoullis whose existence probability is above 0.4 [17, Sec. VI.A].

In order to evaluate the performance of the algorithm, we used the GOSPA metric ($\alpha = 2$, $c = 10$, $p = 2$) [25], which allows to decompose the total error into three components: localization error, missed target error and false target error.

A. Distances comparison

The GOSPA metric computed on 100 Monte Carlo runs yields the results presented in Fig. 3 and in Table III, jointly with the approximated execution times[2] and the mean number of global hypotheses. We compared different distances, described in Section III, to the performance of the standard PMBM implementation [17] and the (track-oriented) Poisson Multi Bernoulli (PMB) filter proposed in [10]. The threshold values applied are reported in the first column of Table III, jointly with the specific metric used.

As we can see, the merged PMBM filter with KLD and L_2 norm shows a GOSPA error quite similar to the standard PMBM implementation, while the reduced number of global hypotheses allows to speed up the tracker, halving the execution times. On the other hand, WKLD performs considerably worse than KLD and L_2 norm; this result is probably due to a suboptimal selection of the Bernoulli components in Alg. 2.

[2]MATLAB(R) implementation on Intel(R) Core(TM) i7-3770 @ 3.40 GHz.

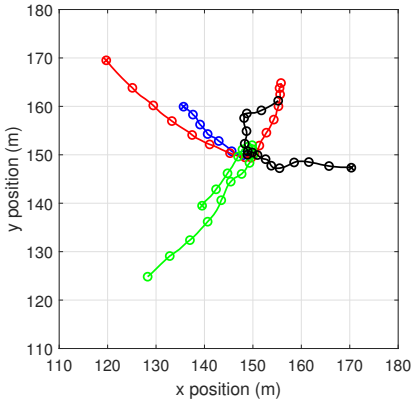

Figure 2: Scenario proposed in [17]. Each one of the four targets is depicted with a different color. The targets are born at time step $k = 1$; they survive for 80 time steps, except the blue target that dies at time step $k = 40$. The target positions at $k = 1$ are indicated by a cross, and the circles show the target positions every five time steps.

In addition, we tested the algorithm with $\Gamma = 0$, i.e. merging only a Bernoullis that have been updated with the same measurement z_k. If we compare these results with those obtained for KLD and L_2 norm, we can note a minimal increase of the execution time and a small growth of the number of global hypothesis computed, while the performance is essentially equal.

The PMB filter is considerably faster than the proposed merged PMBM, as it merges all the single-target hypotheses reducing the number of global hypotheses to one. Nevertheless, the PMB root mean square (RMS) GOSPA error is higher than the KLD and L_2 norm error, as highlighted in Fig. 3.

Table III and Fig. 3 refer to a simulation based on the scenario depicted in Fig. 2, with clutter intensity $\lambda = 10$. In Fig. 5, we also compared the KLD and L_2 norm in scenarios with clutter intensity $\lambda \in \{5, 20, 30\}$; as we can see, the performance and the execution times are similar between the two distances for all the considered scenarios.

B. Threshold analysis

In Fig. 4 and 5 are reported two graphs, in which it is possible to compare the performance of the merging algorithm for different values of the threshold Γ, see Alg. 2. We considered only KLD and L_2 norm; WKLD threshold analysis is not reported due to poor performance results.

Fig. 4 shows the mean number of global hypotheses in different scenarios for KLD and L_2 norm. As we can see, the two distances present very similar results for the chosen ranges of threshold; that is, the complexity of the filter can be considered comparable in the proposed intervals. Nevertheless, the standard deviation in the number of global hypotheses through the Monte Carlo runs is considerably high, thus making difficult a fair comparison between different scenarios; for example, in Fig. 4 the mean number of global hypotheses

Table III: Performance parameters with clutter intensity $\lambda = 10$ on 100 Monte Carlo runs.

	RMS GOSPA Error				Execution time (s)	Mean number of
	Total	Localization	False Target	Missed Target	(approx.)	Global Hypotheses
Standard PMBM	2.83	1.78	1.09	1.91	8.3	135.38
PMB	3.09	1.79	1.7	1.86	1.5	1
Merged PMBM, $\Gamma = 0$	2.83	1.77	1.12	1.89	3.8	31.62
Merged PMBM, KLD, $\Gamma = 0.25$	2.83	1.77	1.11	1.91	3.5	23.97
Merged PMBM, L_2 norm, $\Gamma = 0.05$	2.82	1.77	1.11	1.9	3.6	24.04
Merged PMBM, WKLD, $\Gamma = 0.75$	3.19	1.82	1.01	2.42	2.7	9

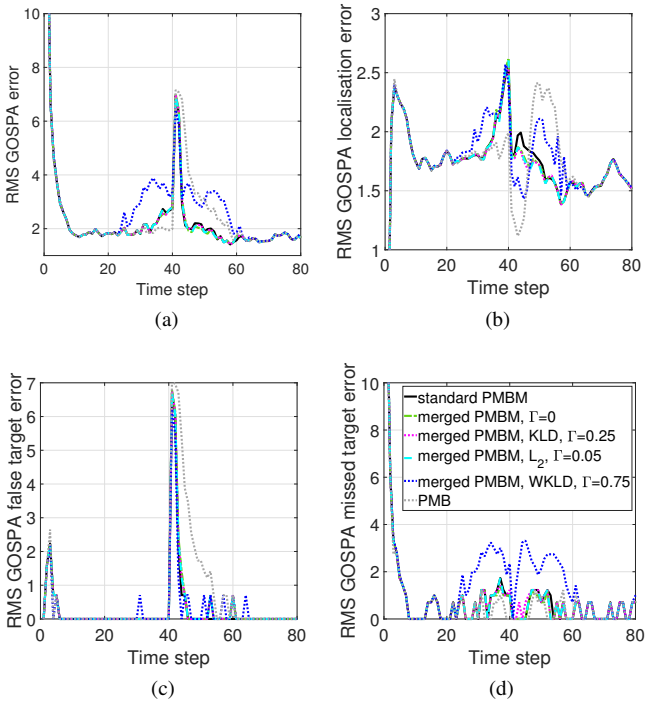

Figure 3: GOSPA metric results with clutter intensity $\lambda = 10$ on 100 Monte Carlo runs. The results of the merged PMBMs, except the one based on the WKLD, show a performance essentially equal to the standard PMBM filter.

for $\lambda = 20$ results counter-intuitively slightly higher than in the $\lambda = 30$ scenario, though both results have a high variance.

Starting from this assumption of equal complexity, we aim to compare the results obtained with both distances in different scenarios, in which the clutter rate is equal to $\lambda \in \{5, 10, 20, 30\}$. Fig. 5 presents the results of the simulation, where the performance expressed through GOSPA metric are compared with the execution times to run the algorithm. The markers and the lines plotted with cold colors represent the performance of the KLD, while the ones drawn with warm colors are related to the L_2 norm; each type of maker depicts the simulation for a specific scenario with clutter rate λ.

As we can see, the value of the threshold Γ affects both performance and execution times in the same way, and it allows to choose the desired trade-off between result accuracy and execution speed. Setting the threshold Γ to infinity, we can minimize the computation times at the cost of a higher GOSPA error, as pointed out by the PMB filter in Fig. 5.

Furthermore, we can notice that the merged PMBM filter

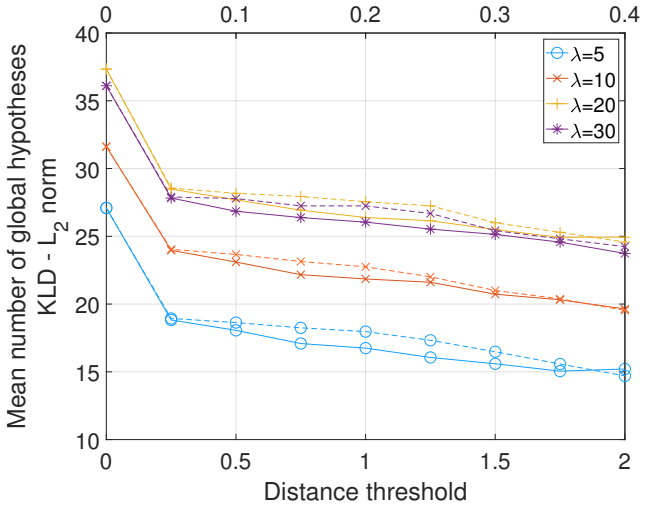

Figure 4: Mean number of global hypotheses of the merged PMBM filter with KLD (solid lines, lower x-axis) and L_2 norm (dashed lines, upper x-axis) for different distances thresholds in four scenarios: $\lambda = 5$ (circles), $\lambda = 10$ (crosses), $\lambda = 20$ (plus signs), $\lambda = 30$ (asterisk).

shows better performance in high-clutter scenarios, as the increase in GOSPA error is lower than in the PMB filter.

VI. Conclusions

In this paper we presented a novel approach to reduce the computational complexity of the PMBM filter. We proposed several distances to measure the similarity between the Bernoullis components, and we provided the closed-form equations for an efficient computation. We presented a merging algorithm, which clusters the Bernoulli components based on the last measurement associated, and merges the most similar components through moment matching. Finally, the obtained reduction of the number of single target hypotheses decreased the number of global hypotheses, leading to a faster algorithm with computational time halved, and performance close the the PMBM filter without Bernoulli merging. The resulting algorithm is also directly applicable to the MBM filter [17].

Appendix

Lemma 3

Proof:

$$\mathrm{D}_{KL}\left(f_1(X) \,\|\, f_2(X)\right) = \int f_1(X) \log \frac{f_1(X)}{f_2(X)} \delta X \tag{18}$$

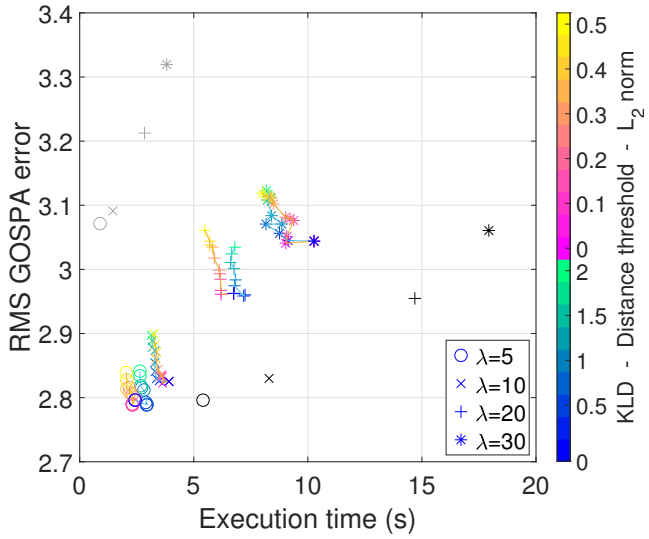

Figure 5: Comparison between performance and execution times of the standard and the merged PMBM filter for different distances thresholds. The values used in the simulation are those analyzed in Fig. 4. The markers indicate the same in four scenarios: $\lambda = 5$ (circles), $\lambda = 10$ (crosses), $\lambda = 20$ (plus signs), $\lambda = 30$ (asterisk). The color of the marker represents the value of the distance threshold used in the simulation of the merged PMBM, as well as the type of distance used (cold colors for KLD, warm colors for L_2 norm). The performance of the standard PMBM filter and the PMB filter are reported in black and grey, respectively.

$$= (1-r_1)\log\frac{1-r_1}{1-r_2} + r_1\int \mathcal{N}(x;\overline{x}_1,P_1)\log\frac{r_1\mathcal{N}(x;\overline{x}_1,P_1)}{r_2\mathcal{N}(x;\overline{x}_2,P_2)}dx \quad (19)$$

$$= (1-r_1)\log\frac{1-r_1}{1-r_2} + r_1\log\frac{r_1}{r_2} + r_1\int \mathcal{N}(x;\overline{x}_1,P_1)\log\frac{\mathcal{N}(x;\overline{x}_1,P_1)}{\mathcal{N}(x;\overline{x}_2,P_2)}dx \quad (20)$$

$$= (1-r_1)\log\frac{1-r_1}{1-r_2} + r_1\log\frac{r_1}{r_2} + \frac{r_1}{2}\left[\mathrm{tr}\left(\left(P_2\right)^{-1}P_1\right) - \log\left(\frac{|P_1|}{|P_2|}\right) - n_x + \left(\overline{x}_2-\overline{x}_1\right)^T\left(P_2\right)^{-1}\left(\overline{x}_2-\overline{x}_1\right)\right]. \quad (21)$$

■

Lemma 6

Proof: The KL divergence upper bound can be proven applying the log sum inequality to each term on the left hand side [22, Theorem 2.7.2]:

$$\left(w_1f_1(X)+w_2f_2(X)\right)\log\frac{w_1f_1(X)+w_2f_2(X)}{f_{12}(X)} = \left(w_1f_1(X)+w_2f_2(X)\right)\log\frac{w_1f_1(X)+w_2f_2(X)}{w_1f_{12}(X)+w_2f_{12}(X)} \leq w_1f_1(X)\log\frac{f_1(X)}{f_{12}(X)} + w_2f_2(X)\log\frac{f_2(X)}{f_{12}(X)} \quad (22)$$

as $0 \leq w_{1,2} \leq 1$ and the logarithm is a monotonic function. Taking the integral of both sides of Inequality (22), we prove Lemma 6. ■

REFERENCES

[1] B.-N. Vo, M. Mallick, Y. Bar-Shalom, S. Coraluppi, R. Osborne, R. Mahler, and B.-T. Vo, "Multitarget tracking," *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2015.

[2] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1991.

[3] T. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association," *IEEE Journal of Oceanic Engineering*, vol. 8, no. 3, 1983.

[4] T. Kurien, "Issues in the design of practical multitarget tracking algorithms," in *Multitarget-Multisensor Tracking: Advanced Applications*. Artech House, 1990.

[5] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, dec 1979.

[6] E. Brekke and M. Chitre, "Relationship between finite set statistics and the multiple hypothesis tracker," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1902–1917, Aug. 2018.

[7] R. P. S. Mahler, *Advances in Statistical Multisource-Multitarget Information Fusion*. Artech House, 2014.

[8] C. P. Robert, *The Bayesian Choice*. New York, NY, USA: Springer, 2007.

[9] B. Vo and B. Vo, "Labeled Random Finite Sets and Multi-Object Conjugate Priors," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3460–3475, Jul. 2013.

[10] J. L. Williams, "Marginal multi-Bernoulli filters: RFS derivation of MHT, JIPDA, and association-based MeMBer," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 3, pp. 1664–1687, 2015.

[11] Á. F. García-Fernández and S. Maskell, "Continuous-Discrete Multiple Target Filtering: PMBM, PHD and CPHD Filter Implementations," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1300–1314, 2020.

[12] M. Fatemi, K. Granström, L. Svensson, F. J. R. Ruiz, and L. Hammarstrand, "Poisson Multi-Bernoulli Mapping Using Gibbs Sampling," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2814–2827, Jun. 2017.

[13] E. Brekke, B. Kalyan, and M. Chitre, "A novel formulation of the Bayes recursion for single-cluster filtering," in *Proc. IEEE Aerospace Conf*, Mar. 2014, pp. 1–16.

[14] M. Fröhle, C. Lindberg, K. Granström, and H. Wymeersch, "Multisensor Poisson Multi-Bernoulli Filter for Joint Target–Sensor State Tracking," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 4, pp. 609–621, 2019.

[15] S. Scheidegger, J. Benjaminsson, E. Rosenberg, A. Krishnan, and K. Granström, "Mono-Camera 3D Multi-Object Tracking Using Deep Learning Detections and PMBM Filtering," in *Proc. IEEE Intelligent Vehicles Symp. (IV)*, Jun. 2018, pp. 433–440.

[16] K. Granstrom, M. Fatemi, and L. Svensson, "Poisson multi-Bernoulli mixture conjugate prior for multiple extended target filtering," *IEEE Transactions on Aerospace and Electronic Systems*, p. 1, 2019.

[17] Á. F. García-Fernández, J. L. Williams, K. Granström, and L. Svensson, "Poisson multi-Bernoulli mixture filter: direct derivation and implementation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 4, pp. 1883–1901, 2018.

[18] Y. Xia, K. Granström, L. Svensson, and M. Fatemi, "Extended target Poisson multi-Bernoulli filter," *arXiv preprint arXiv:1801.01353*, 2018.

[19] H. G. Hoang, B. Vo, B. Vo, and R. Mahler, "The Cauchy-Schwarz Divergence for Poisson Point Processes," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4475–4485, Aug. 2015.

[20] A. R. Runnalls, "Kullback-Leibler Approach to Gaussian Mixture Reduction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 989–999, Jul. 2007.

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

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[23] J. L. Williams and P. S. Maybeck, "Cost-function-based hypothesis control techniques for multiple hypothesis tracking," *Mathematical and Computer Modelling*, vol. 43, no. 9-10, pp. 976 – 989, 2006.

[24] D. J. Salmond, "Mixture Reduction Algorithms for Point and Extended Object Tracking in Clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 2, pp. 667–686, Apr. 2009.

[25] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, "Generalized optimal sub-pattern assignment metric," in *Proc. 20th Int. Conf. Information Fusion (Fusion)*, Jul. 2017, pp. 1–8.