CHALMERS
UNIVERSITY OF TECHNOLOGY

Occlusion-Aware Multiobject Tracking via Expected Probability of Detection

Citation for the original published paper (version of record):
Krejci, J., Kost, O., Xia, Y. et al (2026). Occlusion-Aware Multiobject Tracking via Expected Probability of Detection. IEEE Transactions on Aerospace and Electronic Systems, 62: 8211-8228.
http://dx.doi.org/10.1109/TAES.2026.3676431

Occlusion-Aware Multiobject Tracking via Expected Probability of Detection

JAN KREJČÍ
OLIVER KOST
University of West Bohemia, Pilsen, Czechia

YUXUAN XIA
Shanghai Jiaotong University, Shanghai, China

LENNART SVENSSON, Senior Member, IEEE
Chalmers University of Technology, Göteborg, Sweden

ONDŘEJ STRAKA, Member, IEEE
University of West Bohemia, Pilsen, Czechia

This article addresses multiobject systems, where objects may occlude one another relative to the sensor. The standard point-object model for detection-based sensors is enhanced so that the probability of detection considers the presence of all objects. A principled tracking method is derived, assigning each object an expected probability of detection, where the expectation is taken over the reduced Palm density, which means conditionally on the object's existence. The assigned probability thus considers the object's visibility relative to the sensor, under the presence of other objects. Unlike existing methods, the proposed method systematically accounts for uncertainties related to all objects in a clear and manageable way. The method is demonstrated through a visual tracking application using the multi-Bernoulli mixture filter with marks.

Received 26 November 2025; revised 2 March 2026; accepted 16 March 2026. Date of publication 23 March 2026; date of current version 28 April 2026.

DOI. No. 10.1109/TAES.2026.3676431

Refereeing of this contribution was handled by M. Ulmke.

This work was supported in part by the European Union through the project ROBOPROX under Grant CZ.02.01.01/00/22_008/0004590 and in part by the Czech Science Foundation under Grant GA 25-16919J.

Authors' addresses: Jan Krejčí, Oliver Kost, and Ondřej Straka are with the University of West Bohemia, 301 00 Pilsen, Czechia, E-mail: (jan.krejci8@icloud.com); Yuxuan Xia is with Shanghai Jiaotong University, Shanghai 200240, China; Lennart Svensson is with the Chalmers University of Technology, 412 96 Göteborg, Sweden. *(Corresponding author: Jan Krejčí.)*

I. INTRODUCTION

Multiobject tracking (MOT) refers to estimating the number and locations of multiple moving objects, given sets of noisy detections [1]. The corresponding tracking algorithms are key, e.g., in aerial and naval security [2] or autonomous driving [3]. Most sensors require a clear *line of sight* to detect objects, but *occluding* moving objects can block this visibility. When a tracking algorithm fails to predict occlusion, it may lose sight of occluded objects, which can hinder users in safety-critical situations.

Numerous methods have been proposed to handle such occlusion situations across diverse domains and sensors. The strategies for handling occlusion can vary significantly, ranging from solutions based on *ad hoc* approaches to those grounded in *physical models*. In particular, the methods can be classified into: *(i)* fully ad hoc; *(ii)* combined ad hoc and model-based; and *(iii)* fully model-based strategies. Note that by an *ad hoc* solution, we mean that the solution includes some procedure that is *inconsistent* with a principled (e.g., probabilistic) approximation that leverages the essential physical assumptions about occlusions sketched above.

A fully ad hoc strategy *(i)* directly handles track losses. Basic strategies aim to reintroduce lost object tracks when the objects become visible to the sensor again after occlusion [4], [5], [6], [7]. One can also compute a variable reflecting the current status of occlusions (e.g., visibility) and use it to prevent the method from losing the track in the first place. In the domain of computer vision (CV), this can be achieved by utilizing re-identification (re-ID) visual features [6], [7], [8], [9]. Other approaches from the CV domain include ByteTrack [10], which uses *all* bounding box (BB) outputs (i.e., detections) of a visual detection network, including the detections with low *scores*. Methods [11] or [12] employ so-called pseudo depth and use a custom detection-to-track matching. To handle track loss using the motion capture sensor, Memon et al. [13] employed the linear MOT approach [14], in which measurements appearing nearby other objects are considered as clutter.

Computation of variables reflecting occlusions may involve a model, which leads to strategy *(ii)*. A number of methods employ a customary likelihood function [15], [16], [17], [18], [19], [20], [21], [22], [23]. In CV, the works [15], [17] compute *visibility*, while the work in [18] retains *occlusion rate* variables for each object. Similarly, the authors in [19] and [20] use an *occlusion relationship* variable, an approach that shares similarity with [16]. The works [21] and [22] consider missing measurements due to occlusions as negative information. The work [23] employs particle filters that interact via a likelihood function accounting for occlusions. Instead of defining a custom likelihood, one can craft an *association cost matrix* directly [8], [10], [24], based on occlusion-related variables. Occluded regions are computed in [25] for radar tracking, where objects whose mean vector falls into the occlusion region of the other objects are programatically retained in the algorithm. A similar approach is utilized in [26] for LiDAR tracking.

The ad hoc strategies *(i)* and *(ii)* may be reasonable and lead to superior performance. However, heuristic data processing involved in these methods may lead to doubts about their application in safety-critical areas. To alleviate this, a fully model-based approach *(iii)* can be used. The method [27] proposed minimizing a custom physical *energy* that included an occlusion model, which shares similarities with [28]. A more common approach, however, is to formulate the entire physical model in probabilistic terms and use the Bayes rule to yield the corresponding optimal solution [29], [30], [31]. Note that despite adopting the Bayesian formulation, the methods [4], [5], [6] fall into *(i)* due to their occlusion-handling strategy.

To account for occlusions in the Bayesian formulation consistently, one can either *(iii.a)* adopt the merged measurement framework [32], [33], [34] or *(iii.b)* model probability of detection (PoD) as a *function* reflecting occlusions. The latter is more common in the literature. While *(iii.a)* seems reasonable for the track-before-detect paradigm used in [32], it seems unsuitable for tracking-by-detection focused on by this work, where detection can usually be assigned to one object only. Outside the Bayesian formulation, a related strategy based on object groups is taken in [35] and [36], where objects participating in an occlusion event were merged together.

Drastic approximations involved in existing methods of *(iii.b)* make them fall into *(ii)*. Although the *design* of the PoD function might be sound, its integration into the method follows ad hoc workarounds. The work [37] assigns each object a PoD, assuming that it is a constant function of its mean vector, while the authors in [38], [39], [40], [41], [42], and [43] assume that it depends on an occlusion-related variable. In particular, Jovanoska et al. [39] extend [25] by using an *occlusion likelihood* that scales the PoD. The studies [41], [42], and [43] model the PoD for each object as being directly influenced by the other objects. This allowed [41] to conclude that the posterior is no longer conjugate with the prior used therein and thus intractable. As a simple workaround, the authors in [41], [42], and [43] estimated the states of other objects and substituted them into the PoD. This is arguably a sound approximation, provided that the objects have little spatial and existence uncertainties. The uncertainties, however, usually vary in time and space.

A more sound approximation given in [44] took the uncertainties into account by a weighted sum over data association hypotheses. After analyzing [44], we conclude that the notation and interpretation of the procedure are questionable. Similarly, the definition of *occlusion probability* used to scale the PoD in [38] is mathematically vague and thus questionable. Finally, the works [45], [46] designed the PoD to account for both occlusions and the existence of all the objects *explicitly*. To use it efficiently for each *heuristically* indexed object, an *expected* PoD (EPoD) was computed over its distribution. Note that ad hoc indexing is a subtle issue in the aforementioned methods [25], [38], [39], [40], [45], [46].

To the best of the authors' knowledge, none of the existing methods handle occlusions in a fully model-based and thus consistent manner *(iii)*, while also resulting in a tractable end-to-end algorithm at the same time. This article addresses this issue by developing a domain-independent variational-Bayesian solution, which yields a tractable occlusion-handling strategy. The main contributions of this article are as follows.

1) Two approximations to the multiobject posterior are derived, given a measurement model whose PoD depends on all objects. The corresponding occlusion-handling strategies involve computing the EPoD for each (*1*) state or (*2*) mark, and the expectation is taken over the *reduced Palm density* (RPD) [47].
2) RPD computation is analytically revealed for several multiobject densities used in MOT problems.
3) Tractable EPoD computation is proposed for the prior being a multi-Bernoulli mixture (MBM) with marks.
4) The MBM filter with the proposed occlusion-handling strategy is applied to visual tracking and shown to outperform comparable existing methods.

The rest of this article is organized as follows. Section II introduces background on Bayesian MOT needed for the proposed solution. Section III introduces Palm conditioning, used when a particular object is known to be present, and auxiliary variables used for convenient measurement density representation. Then, the proposed method is outlined. Section IV presents a more efficient approximation and filter implementation. The proposed method is tested in Section V. Finally, Section VII concludes this article.

II. BACKGROUND OF MODEL-BASED MOT

This section briefly reviews the *random finite set* (RFS) approach to MOT with Bayesian treatment of *finite set statistics* (FISST) [29], [30], with special regard to the so-called *standard point-object* (SPO) model. Furthermore, the occlusion-handling strategy of [41], [42], and [43] is reviewed.

A. RFS Approach to MOT

Throughout this article, the symbols $\mathcal{X}$ and $\mathcal{Z}$ denote the state and measurement spaces, respectively, and are both assumed to be *complete separable metric* and measure spaces with some reference measures (such as the Lebesgue measure in the case of $\mathcal{X}=\mathbb{R}^d$). For clarity, unnecessary mathematical details are omitted if possible.

Intuitively, an RFS Ξ on $\mathcal{X}$ is a random subset of $\mathcal{X}$ with a locally finite number of elements (analogically for $\mathcal{Z}$). That is, both the cardinality $|\Xi|$ and the elements $\mathbf{x}\in\Xi$ are random. The elements are regarded as *points*.

Throughout this article, the symbol $X_k=\{\mathbf{x}_k^1,\ldots,\mathbf{x}_k^{n_k}\}$ denotes the set of object states at time step $k=0,\ldots,K$, with $K\in\mathbb{N}$ being the final time step. That is, X_k is a realization of $\Xi_k, k=0,\ldots,K$. Whenever possible, random entities

and their realizations are denoted by the same symbol for convenience. Similarly, measurements received at each time step are modeled as RFSs $Z_k=\{\mathbf{z}_k^1,\ldots,\mathbf{z}_k^{m_k}\}$.

For $k=0,\ldots,K$, the available measurement sets $Z^k\triangleq(Z_0,\ldots,Z_k)$ are used to yield the posterior density $p(X_k|Z^k)$ of X_k via the Bayes filter [29]

$$p\left(X_k|Z^k\right)\propto p(Z_k|X_k)\cdot p(X_k|Z^{k-1}) \tag{1a}$$

$$p(X_k|Z^{k-1})=\int p(X_k|X_{k-1})p(X_{k-1}|Z^{k-1})\delta X_{k-1} \tag{1b}$$

where $p(X_0|Z^{-1})\triangleq p(X_0)$ is the initial density for $k=0$, $p(X_{k+1}|Z^k)$ is the predictive density, and the densities $p(Z_k|X_k)$ and $p(X_{k+1}|X_k)$ encapsulate the measurement and motion models (such as the SPO model), respectively. The symbol $\propto$ denotes equality up to a normalizing constant. The subscript of density functions will be used to provide additional information, such as the distribution name and/or the random variable symbol, if its omission could lead to confusion. The integral of a function $f(X)$ with finite-set inputs is taken to be the *set integral* [29]

$$\int f(X)\delta X\triangleq\sum_{n=0}^{+\infty}\frac{1}{n!}\int f(\{\mathbf{x}^1,\ldots,\mathbf{x}^n\})d\mathbf{x}^1\cdots d\mathbf{x}^n \tag{2}$$

and the densities are such that $\int p(X)\delta X=1$. Several key RFS densities are listed as follows.

1) *Bernoulli RFS:* A Bernoulli RFS is either empty or a singleton. The corresponding density function is

$$p_{\text{Ber}}(X)=\begin{cases}1-r, & \text{if } X=\emptyset\\ r\cdot p_{\text{sp}}(\mathbf{x}), & \text{if } X=\{\mathbf{x}\}\\ 0, & \text{otherwise}\end{cases} \tag{3}$$

where $r\in[0,1]$ is the existence probability and $p_{\text{sp}}(\mathbf{x})$ is the spatial density of $\mathbf{x}$ conditioned on its existence.

2) *Multi-Bernoulli RFS:* A multi-Bernoulli (MB) RFS is a union of multiple independent Bernoulli RFSs. Denoting the density of the ℓth Bernoulli RFS with $p_{\text{Ber}}^{\ell}(X)$, $\ell=1,\ldots,N$, the resulting MB RFS density is given by

$$p_{\text{MB}}(X)=\sum_{\uplus_{\ell=1}^{N}Y^{\ell}=X}\prod_{\ell=1}^{N}p_{\text{Ber}}^{\ell}(Y^{\ell}) \tag{4}$$

where the symbol $\uplus$ means that the sum is taken over all mutually *disjoint* sets $Y^1,\ldots,Y^N$ whose union is X.

3) *Poisson RFS:* A Poisson RFS describes an independent and identically distributed set of points whose number is Poisson distributed. The corresponding density function is

$$p_{\text{Pois}}(X)=\mathrm{e}^{-\rho}\prod_{\mathbf{x}\in X}\rho\cdot p_{\text{sp}}(\mathbf{x}) \tag{5}$$

where $\rho\geq 0$ is the expected number of points.

4) *Poisson Multi-Bernoulli RFS:* The union of independent Poisson-distributed RFS X^0 and MB-distributed RFS X^1 is the Poisson multi-Bernoulli (PMB) RFS. The corresponding density function is

$$p_{\text{PMB}}(X)=\sum_{Y^0\uplus Y^1=X}p_{\text{Pois}}\left(Y^0\right)\cdot p_{\text{MB}}\left(Y^1\right) \tag{6}$$

where the summation is analogical to that in (4).

5) *Marked RFS:* As a special RFS case, an RFS $X\subset\mathcal{X}$ with $\mathcal{X}=\mathcal{X}^{\mathcal{S}}\times\mathcal{M}$ being a joint space consisting of a complete separable metric space $\mathcal{X}^{\mathcal{S}}$ and a set of *marks* $\mathcal{M}$ is called a *marked RFS*, whenever the points without marks (i.e., the corresponding marginal random set on $\mathcal{X}^{\mathcal{S}}$) again forms an RFS; cf. [48, Sect. 1.8]. In this article, the *space of marks* $\mathcal{M}$ is assumed to be discrete. The same applies to $\mathcal{Z}$.

Points with marks will be denoted as $\mathbf{x}=(x,m)\in\mathcal{X}$. To extract marks from a set X of marked points, define

$$\mathcal{L}(X)\triangleq\{m\in\mathcal{M}:\exists x\in\mathcal{X}^{\mathcal{S}}\text{ s.t. }(x,m)\in X\}. \tag{7}$$

Note that marks are not necessarily unique, i.e., $|\mathcal{L}(X)|$ might not be equal to $|X|$. For a set X, let

$$X^m\triangleq\{\mathbf{x}\in X:\mathcal{L}(\{\mathbf{x}\})=\{m\}\} \tag{8}$$

denote the subset of points of X with the same mark $m\in\mathcal{M}$. Integral of a function $f(\mathbf{x})$ on $\mathcal{X}=\mathcal{X}^{\mathcal{S}}\times\mathcal{M}$ is

$$\int f(\mathbf{x})d\mathbf{x}=\sum_{m\in\mathcal{M}}\int f(x,m)dx \tag{9}$$

and the set integral (2) of a function $f(X)$ with finite-set inputs thus takes the form

$$\int f(X)\delta X=\sum_{n=1}^{+\infty}\frac{1}{n!}\sum_{m^1,\ldots,m^n\in\mathcal{M}}\int f(\{(x^1,m^1),\ldots,(x^n,m^n)\})dx^1\cdots dx^n. \tag{10}$$

Denoting the density of a marked RFS with $p_{\text{marked}}(\cdot)$, the corresponding set of points without marks has the marginal density for any $\{x^1,\ldots,x^n\}\subset\mathcal{X}^{\mathcal{S}}$ given by

$$p_{\text{unmarked}}(\{x^1,\ldots,x^n\})=\sum_{m^1,\ldots,m^n\in\mathcal{M}}p_{\text{marked}}(\{(x^1,m^1),\ldots,(x^n,m^n)\}). \tag{11}$$

A significant portion of the MOT literature is concerned with the case when points must always have distinct marks. In such a case, the marks are referred to as *labels* [49], [50], [51] and are used to distinguish among different objects to form trajectories. Labels typically apply to a set of states only, while in this article, both the sets of states and measurements will be considered as having marks.

B. MOT Without the Presence of Occlusions

A widely adopted model for tracking objects without any interactions (including occlusions) is given by the SPO model. It should be emphasized that the terminology *point* object refers to the assumption that each point (i.e., object) $\mathbf{x}_k\in X_k$ is *resolvable* as an actual *point* for the sensor. That is, each object is either detected or not, whereas there is at most one detection per object.

In the SPO measurement model, each object $\mathbf{x}_k\in X_k$ is either detected with the state-dependent PoD $P_D(\mathbf{x}_k)$ and independently generates a measurement according to the

single-object likelihood function $L(\mathbf{z}_k|\mathbf{x}_k)$, or it is undetected with the probability $1-P_D(\mathbf{x}_k)$. Moreover, independent clutter detections described with the RFS C_k may be produced by a sensor. The set of all measurements at time step k is thus [29, p. 411]

$$Z_k = \left(\bigcup_{\mathbf{x}_k \in X_k} D(\mathbf{x}_k) \right) \cup C_k \tag{12}$$

where $D(\mathbf{x}_k)$ is a conditional Bernoulli RFS with density

$$p(D|\mathbf{x}_k) = \begin{cases} 1 - P_D(\mathbf{x}_k), & \text{if } D = \emptyset \\ P_D(\mathbf{x}_k) L(\mathbf{z}_k|\mathbf{x}_k), & \text{if } D = \{\mathbf{z}_k\} \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

The clutter RFS C_k is modeled as Poisson. The density $p(Z_k|X_k)$ resulting from (12) is thus a PMB (6) [29, Sect. 12]. The SPO motion model is omitted in this article for brevity and can be found, e.g., in [29, p. 467].

Assuming that the birth and the initial density are both MB (or mixtures thereof), the corresponding model-based filter is the MBM filter [52], [53]. If object states for the MBM filter are uniquely marked, the labeled MBM (LMBM) filter [52] results without the need to alter the filtering recursion. The so-called generalized labeled multi-Bernoulli (GLMB) [51], also referred to as MBM_{01} in [54], can be understood as a special case of the LMBM filter, where the existence probability of each Bernoulli is either 0 or 1, resulting in an exponential increase of global hypotheses in the prediction step. Approximations can be employed to yield, e.g., the labeled multi-Bernoulli (LMB) [55] and other filters [56]. It is also possible to assume that the birth and the initial distribution are Poisson; see, e.g., [29, Sect. 16.3], [54], [56].

Particular filtering recursions are omitted and can be found in the literature. Note that *Gaussian mixture* (GM) implementations often assume constant PoD $P_D(\mathbf{x}_k) = P_D$.

C. MOT in the Presence of Occlusions

In this section, the SPO model is generalized to account for occlusions, following [41], [42], and [43]. In particular, the structure of the SPO measurement model (12) and (13) is adjusted to allow the possibility that detections are less likely to occur for occluded objects. The corresponding model is referred to as SPO with *dependence* (SPO-D).

Both the PoD and the single-object likelihood function for each $\mathbf{x}_k \in X_k$ may also depend on the rest of the objects $X_k \backslash \{\mathbf{x}_k\}$. Since we are modeling *point objects*, it is reasonable to assume that measurements, if received, are not influenced by the presence of other objects. Therefore, the dependence of the likelihood function $L(\mathbf{z}_k|\mathbf{x}_k)$ on the other objects $X_k \backslash \{\mathbf{x}_k\}$ is dropped. That is, (12) is generalized for the SPO-D model to

$$Z_k = \left(\bigcup_{\mathbf{x}_k \in X_k} D(\mathbf{x}_k,\ X_k \backslash \{\mathbf{x}_k\}) \right) \cup C_k \tag{14}$$

where $D(\mathbf{x}_k,\ X_k \backslash \{\mathbf{x}_k\})$ is a Bernoulli RFS with

$$p(D \mid \mathbf{x}_k, X_k \backslash \{\mathbf{x}_k\}) = \begin{cases} 1 - P_D(\mathbf{x}_k, X_k \backslash \{\mathbf{x}_k\}), & \text{if } D = \emptyset \\ P_D(\mathbf{x}_k, X_k \backslash \{\mathbf{x}_k\}) \cdot L(\mathbf{z}_k|\mathbf{x}_k), & \text{if } D = \{\mathbf{z}_k\} \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

where $P_D(\mathbf{x}_k, X_k \backslash \{\mathbf{x}_k\})$ is the PoD that depends on the given object $\mathbf{x}_k$ and the set of *other* objects $X_k \backslash \{\mathbf{x}_k\}$. Examples of $P_D(\mathbf{x}_k, X_k \backslash \{\mathbf{x}_k\})$ expressions can be found in [41], [42], and [43] and will be discussed in detail later. Clutter C_k is still modeled as Poisson. It is easy to see that the density $p(Z_k|X_k)$ resulting from (14) is still a PMB (6), since X_k is given. The SPO-D motion model is taken to be the SPO motion model.

Unless the prior is Bernoulli, the posterior corresponding to the SPO-D model contains *interactions* among objects induced by the dependence of $D(\mathbf{x}_k, X_k \backslash \{\mathbf{x}_k\})$ on $X_k \backslash \{\mathbf{x}_k\}$. This was observed by Ong et al. [41] for the GLMB prior. To achieve tractability of the posterior, approximations must be developed. The works [41], [42], and [43] used an ad hoc approximation within the labeled framework, which could be stated as follows:

E: estimate the set of all objects $\hat{X}_k$ based on the prior;
S: substitute $\hat{X}_k$ into the SPO-D model, i.e., define

$$P_D(m) := P_D\left((\hat{x}_k, m),\ \hat{X}_k \backslash \{(\hat{x}_k, m)\}\right) \tag{16}$$

where $\hat{x}_k$ is the estimate corresponding to label m;
O: obey the SPO model parameterized with (16) to achieve tractable filtering recursion.

This occlusion-handling strategy will be referred to as ESO strategy for short. Since the estimation step E is based on existence probabilities, the ESO strategy considers (*I*) no spatial uncertainties and (*II*) only a limited amount of existence uncertainties corresponding to the objects. Note that the computational complexity of ESO depends on the chosen estimator and on the PoD function implementation.

Other strategies that take the uncertainties into account are either insufficiently described [38], [44] or assume that the PoD depends on the uncertainties explicitly [45], [46], which fails to follow physical modeling. A strategy is proposed in the following section that considers uncertainties in a mathematically sound manner.

III. PROPOSED APPROXIMATION

The proposed occlusion-handling strategy modifies the steps Eand S, while it keeps the appealing step Oin use. It serves as an approximation of the posterior, which will be detailed later. Specifically, (16) is replaced by calculating a conditional *expectation* of $P_D(\mathbf{x}_k, X_k \backslash \{\mathbf{x}_k\})$, i.e., EPoD. Note that Lamard et al. [45] computed an EPoD differently, and that EPoD can be found in other contexts [57], [58].

To theoretically develop the strategy outlined above, we first introduce *Palm conditioning* (see Section III-A) and *auxiliary variables* (see Section III-B), which are essential for establishing the approximation in Section III-C.

A. Palm Conditioning

The *Palm distribution* is the conditional probability law of the RFS Ξ, given that a particular point $\mathbf{x} \in \mathcal{X}$ is known to be an element of Ξ, i.e., given the event $\mathbf{x} \in \Xi$. The *reduced* Palm distribution is then the probability law of the *other* points, i.e., of the subset $\Theta = \Xi \backslash \{\mathbf{x}\}$, given $\mathbf{x} \in \Xi$. Its corresponding density function, called the RPD, is needed in this article. In the FISST notation, the RPD evaluated at $O := X \backslash \{\mathbf{x}\}$ can be computed simply as (see [47, p. 511])

$$p_{\Theta|\mathbf{x}}(O|\mathbf{x}) \overset{\text{abbr.}}{=} p_{\Xi\backslash\{\mathbf{x}\}|\mathbf{x}\in\Xi}(O|\mathbf{x}) = \frac{p_\Xi(O \cup \{\mathbf{x}\})}{D_\Xi(\mathbf{x})} \tag{17}$$

where $D_\Xi(\mathbf{x})$ is the probability hypothesis density (PHD) corresponding to Ξ as

$$D_\Xi(\mathbf{x}) = \int p_\Xi(X \cup \{\mathbf{x}\}) \delta X. \tag{18}$$

Palm conditioning springs from the random counting measure treatment of RFSs, which is an alternative to FISST. The rigorous construction of (17) within FISST is detailed in Appendix A. Several RPD examples follow.

1) *RPD Corresponding to a Bernoulli RFS:* It follows trivially that the RPD for a Bernoulli RFS is equal to one if $O = \emptyset$ and zero otherwise. It also follows from (17) using the Bernoulli PHD $D_{\text{Ber}}(\mathbf{x}) = r \cdot p_{\text{sp}}(\mathbf{x})$.

2) *RPD Corresponding to a Poisson RFS:* According to the *Slivnyak–Mecke* theorem [47, p. 281], Palm conditioning within a Poisson process introduces no knowledge of the distribution of the other points. Indeed, the RPD (17) for the Poisson RFS (5) is

$$p_{\text{Pois}:\Theta|\mathbf{x}}(O|\mathbf{x}) = \frac{\mathrm{e}^{-\rho} \prod_{\mathbf{o}\in O\cup\{\mathbf{x}\}} \rho \cdot p_{\text{sp}}(\mathbf{o})}{\rho \cdot p_{\text{sp}}(\mathbf{x})} = p_{\text{Pois}}(O). \tag{19}$$

3) *RPD Corresponding to a Poisson Multi-Bernoulli RFS:* Consider an RFS Ξ with a PMB density (6). A point $\mathbf{x} \in \Xi$ might have come from either the Poisson component or one of the Bernoulli components. Conditioning on "$\mathbf{x} \in \Xi$" must thus explore either possibility, leading to a mixture of PMB densities. First, the numerator of (17) is

$$\begin{aligned} p_{\text{PMB}}(O \cup \{\mathbf{x}\}) &= \sum_{O^0 \uplus O^1 = O \cup \{\mathbf{x}\}} p_{\text{Pois}}(O^0) \cdot p_{\text{MB}}(O^1) \\ &= \sum_{Y^0 \uplus Y^1 = O} p_{\text{Pois}}(Y^0) \left(\rho \cdot p_{\text{sp}}(\mathbf{x}) \cdot p_{\text{MB}}(Y^1) \right. \\ &\quad \left. + \sum_{\ell=1}^{N} r^\ell \cdot p_{\text{sp}}^\ell(\mathbf{x}) \cdot p_{\text{MB}\backslash\ell}(Y^1) \right) \end{aligned} \tag{20}$$

where r^ℓ is the existence probability and $p_{\text{sp}}^\ell(\mathbf{x})$ is the spatial probability density corresponding to the ℓth Bernoulli component, respectively; $\ell = 1, \ldots, N$, The symbol MB$\backslash\ell$ is a shorthand notation for the MB resulting from omitting the ℓth component

$$p_{\text{MB}\backslash\ell}(X) = \sum_{\uplus_{m=1, m\neq\ell}^{N} Y^m = Y} \prod_{\substack{m=1 \\ m\neq\ell}}^{N} p_{\text{Ber}}^m(Y^m). \tag{21}$$

Dividing (20) with the PHD

$$D_{\text{PMB}}(\mathbf{x}) = \rho \cdot p_{\text{sp}}(\mathbf{x}) + \sum_{\ell=1}^{N} r^\ell \cdot p_{\text{sp}}^\ell(\mathbf{x}) \tag{22}$$

[cf. (17)] yields the desired RPD as

$$\begin{aligned} p_{\text{PMB}:\Theta|\mathbf{x}}(O|\mathbf{x}) = \sum_{O^0 \uplus O^1 = O} p_{\text{Pois}}(O^0) \left(w^0 \cdot p_{\text{MB}}(O^0) \right. \\ \left. + \sum_{\ell=1}^{N} w^\ell \cdot p_{\text{MB}\backslash\{\ell\}}(O^1) \right) \end{aligned} \tag{23}$$

which is a PMB mixture, where the weights are

$$w^0 = \frac{\rho \cdot p_{\text{sp}}(\mathbf{x})}{\rho \cdot p_{\text{sp}}(\mathbf{x}) + \sum_{\ell=1}^{N} r^\ell \cdot p_{\text{sp}}^\ell(\mathbf{x})} \tag{24a}$$

$$w^\ell = \frac{r^\ell \cdot p_{\text{sp}}^\ell(\mathbf{x})}{\rho \cdot p_{\text{sp}}(\mathbf{x}) + \sum_{\ell=1}^{N} r^\ell \cdot p_{\text{sp}}^\ell(\mathbf{x})}, \qquad \ell = 1, \ldots, N. \tag{24b}$$

Note that the RPD corresponding to MB RFS is a special case of (23) and (24) when setting $\rho = 0$.

If a point $\mathbf{x} \in \Xi$ contains the information from which PMB component it came from, both the PMB density and its RPD get simplified significantly. We aim to simplify the PMB densities related to the SPO and SPO-D measurement likelihoods for convenience. For this reason, a special class of marked RFSs is introduced, whose marks are made to correspond with the different density components. Such marks effectively behave like auxiliary variables that facilitate the development of efficient approximations [59].

B. Auxiliary Variables

First, define a spatial density *auxiliary supported on* $\mu \in \mathcal{M}$ to be a spatial density on the joint space $\mathcal{X} = \mathcal{X}^{\mathcal{S}} \times \mathcal{M}$, for which

$$p_{\text{sp}}(x, m) = p_{\text{sp}}(x|m) \cdot \delta_\mu[m], \qquad (x, m) \in \mathcal{X} \tag{25}$$

where $p_{\text{sp}}(x|m)$ is a spatial density on $\mathcal{X}^{\mathcal{S}}$ conditioned on $m \in \mathcal{M}$, and $\delta_\mu[m]$ is the Kronecker delta. Moreover, an *auxiliary Bernoulli (respectively, Poisson) RFS supported on* $\mu \in \mathcal{M}$ is a Bernoulli (respectively, Poisson) RFS whose spatial density is auxiliary supported on μ. The corresponding density functions are denoted with $p_{\text{A-Pois}}^\mu(X)$ and $p_{\text{A-Ber}}^\mu(X)$, for the Poisson and Bernoulli cases, respectively.

The union $X = X^0 \uplus X^1 \uplus \cdots \uplus X^N$ of independent RFSs where X^0 is auxiliary Poisson RFS supported on $\mu^0 \in \mathcal{M}$, and X^ℓ is auxiliary Bernoulli RFS supported on $\mu^\ell \in \mathcal{M}$, $\ell = 1, \ldots, N$, with $\mu^0, \mu^1, \ldots, \mu^N$ being distinct, is the *auxiliary* PMB (A-PMB) RFS. The corresponding density function defined on the joint space is

$$\begin{aligned} &p_{\text{A-PMB}}(X) \\ &= 1_M(\mathcal{L}(X)) \cdot p_{\text{A-Pois}}^{\mu^0}\left(X^{\mu^0}\right) \prod_{\ell=1}^{N} p_{\text{A-Ber}}^{\mu^\ell}\left(X^{\mu^\ell}\right) \end{aligned} \tag{26}$$

where $M = \{\mu^\ell\}_{\ell=1}^N$, the symbol X^{μ^ℓ} is as defined in (8), and $1_A(B)$ is the *inclusion function* [50], which is equal to one if $B \subseteq A$ and zero otherwise. The density corresponding to the set of points without marks (11) is the standard PMB density (6) (see [59]).

To define auxiliary measurement likelihoods, let $\mathcal{Z}=\mathcal{Z}^{\mathcal{S}}\times(\mathcal{M} \uplus \{0\})$ be the joint space of measurements, where 0 will be used to determine clutter. Consider that actual measurements received are subsets $\tilde{Z}_k$ of $\mathcal{Z}^{\mathcal{S}}$. Note that the SPO and SPO-D measurement models from Sections II-B and II-C, respectively, are considered to be defined for Z_k on $\mathcal{Z}$. Moreover, it is assumed that the set of states on the joint space $\mathcal{X}=\mathcal{X}^{\mathcal{S}}\times\mathcal{M}$ forms a labeled RFS, i.e., elements of X_k have distinct marks.

For each given $\mathbf{x}_k\in\mathcal{X}$, specialize the definition of the single-object likelihood $L(\cdot\,|\mathbf{x})$ on $\mathcal{Z}$ to be the spatial density auxiliary supported on $\mathcal{L}(\mathbf{x}_k)\in\mathcal{M}$, as

$$L(z, m|\mathbf{x}_k) = L(z|\mathbf{x}_k)\cdot\delta_{\mathcal{L}(\mathbf{x}_k)}[m], \qquad (z, m)\in\mathcal{Z} \tag{27}$$

where $L(z|\mathbf{x}_k)$ is the likelihood of the actual $z\in\mathcal{Z}^{\mathcal{S}}$. For the SPO measurement model, the corresponding likelihood is the A-PMB density, further called *auxiliary likelihood*

$$\begin{aligned} p_{\text{A-SPO}_k}(Z_k|X_k,\ P_D) &= 1_{\mathcal{L}(X_k)\uplus\{0\}}\left(\mathcal{L}(Z_k)\right) \\ &\quad\times p_{C_k}(Z_k^0)\prod_{\mathbf{x}_k\in X_k} p_{D(\mathbf{x}_k)}(Z_k^{\mathcal{L}(\mathbf{x}_k)}|\mathbf{x}_k,\ P_D) \end{aligned} \tag{28}$$

where Z_k^m (8) is the subset of measurements with the mark m and where $p_{C_k}(\cdot)$ is a Poisson density of clutter. The density $p_{D(\mathbf{x}_k)}(\cdot\,|\mathbf{x}_k,\ P_D)$ is defined in (13), where the dependence on $P_D(\cdot,\ \cdot)$ is coined explicitly. The SPO-D auxiliary likelihood $p_{\text{A-SPO-D}_k}(Z_k|X_k)$ is a straightforward generalization of (28) and is thus omitted. Since (28) is an auxiliary PMB, the density corresponding to the set of points without marks (i.e., the actual measurements) becomes the likelihood of the standard form (6). Note that the number of Bernoulli components in (28) is $|X_k|$ and that marks in X_k must be distinct so that (28) is A-PMB.

C. SPO Approximation With Best PoD: General Case

Consider that the true MOT model obeys the SPO-D measurement model. A top-down approximation of the (unnormalized) posterior is established below, using an auxiliary likelihood corresponding to a much simpler SPO model structure and an optimal parameter $P_D(\mathbf{x}_k)$.

DEFINITION 1 (KULLBACK–LEIBLER DIVERGENCE (KLD)) Given two densities $p(\cdot)$ and $q(\cdot)$ defined on a (measure) space $\mathcal{Y}$, the expression

$$D_{\text{KL}}(p\|q) = \int_{\mathcal{Y}} p(y)\log\left(\frac{p(y)}{q(y)}\right)dy \tag{29}$$

is a statistical quantification of the discrepancy between the two densities, commonly known as the KLD.

Let $p(\cdot)$ be the true joint density on the right-hand side of (1a). That is, omitting the conditioning on Z^{k-1}

$$p(Z_k, X_k) := p_{\text{A-SPO-D}_k}(Z_k|X_k)p(X_k|Z^{k-1}). \tag{30}$$

Let $q(\cdot)$ be a joint density $q_{\mathcal{P}}(\cdot)$ resulting from using the simpler likelihood (28) with parameter $\mathcal{P}:\mathcal{X}\to[0,\,1]$ as

$$q_{\mathcal{P}}(Z_k, X_k) := p_{\text{A-SPO}_k}(Z_k|X_k,\ \mathcal{P})p(X_k|Z^{k-1}). \tag{31}$$

Denote the space of joint densities of the form (31) with

$$Q \triangleq \{q_{\mathcal{P}}(\cdot,\ \cdot) \text{ of the form (31)}:\quad \mathcal{P}:\mathcal{X}\to[0,\,1]\}. \tag{32}$$

The following proposition reveals the optimal $\mathcal{P}^*$, for which the joint densities are closest in the sense of KLD.

PROPOSITION 1 The best-fitting joint density

$$q^* = \arg\min_{q_{\mathcal{P}}\in Q} D_{\text{KL}}(p\|q_{\mathcal{P}}) \tag{33}$$

with $p(\cdot)$ of the form (30) is $q_{\mathcal{P}^*}(Z_k, X_k)$, where

$$\mathcal{P}^*(\mathbf{x}_k) = \text{E}_{\Theta_k|\mathbf{x}_k}\left[P_D(\mathbf{x}_k,\ O_k)\right] \tag{34}$$

is the EPoD. The expectation is taken over the RPD (17). The proof is given in Appendix B.

Using $P_D(\mathbf{x}_k) := \mathcal{P}^*(\mathbf{x}_k)$ (34) for the SPO model, we conclude that the density $p_{\text{A-SPO}_k}(Z_k|X_k,\ \mathcal{P}^*)$ best matches $p_{\text{A-SPO-D}_k}(Z_k|X_k)$ for any X_k distributed according to the prior density. By considering the auxiliary likelihoods without marks using (11), we obtain the best SPO approximation for the intractable SPO-D model.

IV. SPO APPROXIMATION WITH BEST POD: MARKS ONLY

As indicated before, implementations using GMs often assume a constant PoD. To address occlusions, however, the PoD cannot be the same constant for all objects. In the following, we deal with an efficient *mark-only* case of Proposition 1 to find a suitable constant for each object. Assume that $\mathcal{P}$ in (31) is a function of only the mark, i.e., $\mathcal{P}:\mathcal{M}\to[0,\,1]$. Redefining Q (32) as

$$Q \triangleq \{q_{\mathcal{P}}(\cdot,\ \cdot) \text{ of the form (31)}:\quad \mathcal{P}:\mathcal{M}\to[0,\,1]\} \tag{35}$$

and using it with (33) yields the optimal

$$\mathcal{P}^*(m) = \text{E}_{O_k,x_k|(\cdot,m)}\left[P_D\left((x_k, m)\right), O_k\right] \tag{36a}$$

$$\equiv \iint P_D\left((x_k, m)\right), O_k) \times p_{\Theta_k|\mathbf{x}_k}(O_k|(x_k, m))\,D(x_k|m)\delta O_k dx_k \tag{36b}$$

where the single-object conditional and marginal densities for the given mark $m\in\mathcal{M}$ are given by the PHD as

$$D(x_k|m) = \frac{D_{\Xi}(x_k, m)}{D(m)}, \quad D(m) = \int D_{\Xi}(x_k, m)dx_k. \tag{37}$$

As a special case of Proposition 1, the result (36) provides the best-fitting PoD as a function of only the mark. Its practical use is explained in the following, while the proof is given in Appendix B (see Corollary 3).

Recall that marks in X_k must be distinct so that the auxiliary SPO and SPO-D likelihoods are A-PMB densities. It is apparent that this assumption is unnecessary for computing (36), while X_k does not even need to be marked for the calculation of (34). For such a case, however, the optimality (Proposition 1 and thus Corollary 3) was not established in

this article. To give a provably optimal end-to-end solution below, we remain consistent with the assumption on X_k and use the MBM filter for convenience.

A. MBMs in MOT

Assuming that the birth is MB, the MBM density is a conjugate prior for the SPO model (see Section II-B), giving rise to the MBM filter [53]. We equip each Bernoulli with a unique mark to compute the EPoD for each mark instead of each state. Following Section III-B directly, the marks are called auxiliary variables in this article and they are equivalent to labels. In particular, we deal with the special case of A-PMBM densities where the Poisson part is missing, which are addressed as auxiliary MB (A-MB) densities. Note that the resulting algorithm is a special case of the MBM filter that is equivalent to the LMBM filter. For notational simplicity, the time index k is dropped in the following.

1) *Auxiliary MB RFS:* The A-MB RFS results by dropping the Poisson component from the A-PMB RFS (see Section III-B), and it can be regarded as the LMB RFS [30, pp. 453–458]. The corresponding density can be written as

$$p_{\text{A-MB}}(X) = 1_M(\mathcal{L}(X)) \cdot \prod_{\ell=1}^{N} p_{\text{A-Ber}}^{\mu^\ell}\left(X^{\mu^\ell}\right). \tag{38}$$

If $m \in M=\{\mu^1, \ldots, \mu^N\}$, its RPD is a special case of (23)

$$p_{\text{A-MB}:\Theta|\mathbf{x}}(O|(x,m)) = p_{\text{A-MB}\backslash m}(O) \tag{39}$$

and zero otherwise, with $p_{\text{A-MB}\backslash m}(O)$ given in (21)

2) *Auxiliary MBM RFS:* A mixture of A-MB-distributed RFSs is referred to as the auxiliary MBM (A-MBM) RFS, which can be viewed as the LMBM RFS [52]. In MOT, each mixture term corresponds to a *global data association hypothesis*. Assigning a subindex h to each term/hypothesis, the density of the A-MBM RFS becomes

$$p_{\text{A-MBM}}(X) = \sum_{h=1}^{\mathcal{H}} w_h \cdot p_{\text{A-MB},h}(X) \tag{40}$$

where $\mathcal{H}$ is the number of mixture terms, and w_h is the weight of the hypothesis h, $h=1, \ldots, \mathcal{H}$, such that $\sum_{h=1}^{\mathcal{H}} w_h = 1$. Let $M_h \triangleq \{\mu_h^\ell\}_{\ell=1}^{N_h}$ be the set of auxiliary variables forming the support corresponding to the A-MB hypothesis h, and extend the existence probability r_h^m to be zero if $m \notin M_h$. The RPD for the A-MBM RFS is then

$$p_{\text{A-MBM}:\Theta|\mathbf{x}}(O|(x,m)) = \frac{1}{D_{\text{A-MBM}}(x,m)} \sum_{h=1}^{\mathcal{H}} w_h \cdot r_h^m \cdot p_{\text{sp},h}(x|m) \cdot p_{\text{A-MB},h\backslash m}(O) \tag{41}$$

where the PHD of the A-MBM RFS is

$$D_{\text{A-MBM}}(x,m) = \sum_{h=1}^{\mathcal{H}} w_h \cdot r_h^m \cdot p_{\text{sp},h}(x|m). \tag{42}$$

Employing the PHD conditioning (37) yields

$$D_{\text{A-MBM}}(x|m) = \frac{\sum_{h=1}^{\mathcal{H}} w_h \cdot r_h^m \cdot p_{\text{sp},h}(x|m)}{D_{\text{A-MBM}}(m)} \tag{43a}$$

$$D_{\text{A-MBM}}(m) = \sum_{h=1}^{\mathcal{H}} w_h \cdot r_h^m. \tag{43b}$$

The conditional density for the expectation (36) becomes

$$p_{\text{A-MBM}:\Theta|\mathbf{x}}(O|(x,m))\, D_{\text{A-MBM}}(x|m) = \frac{1}{D_{\text{A-MBM}}(m)} \sum_{h=1}^{\mathcal{H}} w_h \cdot r_h^m \cdot p_{\text{sp},h}(x|m) \cdot p_{\text{A-MB},h\backslash m}(O). \tag{44}$$

B. Optimal Parameter for MBM Filtering

Once the A-MBM prior density is obtained in the form (40), the best-fitting approximation (36) becomes

$$P_D(m) := \frac{1}{\sum_{h=1}^{\mathcal{H}} w_h \cdot r_h^m} \sum_{h=1}^{\mathcal{H}} w_h \cdot r_h^m \cdot \bar{P}_D(m,h) \tag{45}$$

where the EPoD corresponding to the hypothesis h is

$$\bar{P}_D(m,h) \triangleq \iint P_D\left((x,m),O\right)\, p_{\text{A-MB},h\backslash m}(O) p_{\text{sp},h}(x|m)\, \delta O dx. \tag{46}$$

Since $P_D(\cdot,\cdot)$ cannot generally be written as a sum or a product, the computation of (46) is challenging. We propose to use directly the set integral form of (46), i.e.,

$$\bar{P}_D(m,h) = \sum_{n=0}^{|M_h\backslash\{m\}|} \sum_{A\subseteq(M_h\backslash\{m\}),\text{ s.t. }|A|=n} \omega_h^m(A) \int \underbrace{\int \cdots \int}_{n\text{-times}} P_D\left((x,m),\{(o^\mu,\mu)\}_{\mu\in A}\right) \times p_{\text{sp},h}(x|m) \left(\prod_{\mu\in A} p_{\text{sp},h}(o^\mu|\mu)\right) dx \prod_{\mu\in A} do^\mu \tag{47}$$

if $m \in M_h$ and zero otherwise, where

$$\omega_h^m(A) \triangleq \left(\prod_{\mu\in(M_h\backslash\{m\})\backslash A} (1-r_h^\mu)\right)\left(\prod_{\mu\in A} r_h^\mu\right) \tag{48}$$

is a set function, cf. [30, p. 454]. Note that (47) generally has $\sum_{n=0}^{|M_h|-1} \binom{|M_h|-1}{n} = 2^{|M_h|-1}$ terms and the complexity of their evaluation ranges greatly among them.

In practice, the number of other objects with high existence probability that are likely to occlude the object with mark m is often rather small. In such a case, the sums in (47) involve many insignificant terms. Next, we propose a reduction and efficient approximation of (47).

1) *Simplification of the Hypothesis h :* First, reduce the set $M_h\backslash\{m\}$ by discarding Bernoulli components with

insignificant existence probabilities. Further reduction can be based on the following axiom.

AXIOM 1 (OBJECTS OCCLUDE INDIVIDUALLY) PoD is a locally integrable function satisfying the implication

$$P_D(\mathbf{x},\{\mathbf{o}\}) = P_D(\mathbf{x},\emptyset) \Rightarrow P_D(\mathbf{x}, O\cup\{\mathbf{o}\}) = P_D(\mathbf{x},O) \quad (49)$$

for any $\mathbf{x}, \mathbf{o}\in\mathcal{X}$ and any finite $O\subset\mathcal{X}$. In essence, if an object $\mathbf{o}$ does not influence the PoD of the object $\mathbf{x}$ (i.e., does not occlude $\mathbf{x}$) on its own, then it never will, regardless of the presence of any other object.

COROLLARY 1 Let $\mathbf{x}$ and $\mathbf{o}$ be random variables on $\mathcal{X}$ with continuous (spatial) density functions; then

$$\begin{aligned}&\mathrm{E}_{\mathbf{x},\mathbf{o}}[P_D(\mathbf{x},\{\mathbf{o}\})] = \mathrm{E}_{\mathbf{x}}[P_D(\mathbf{x},\emptyset)]\\ &\quad\Rightarrow \mathrm{E}_{\mathbf{x},\mathbf{o}}[P_D(\mathbf{x},O\cup\{\mathbf{o}\})] = \mathrm{E}_{\mathbf{x}}[P_D(\mathbf{x},O)] \end{aligned} \quad (50)$$

for any finite $O\subset\mathcal{X}$, where $\mathcal{X}$ is further assumed to be *directionally limited* [1] [60, p. 7]. In words, if an object $\mathbf{o}$ does not affect the integral of the PoD of $\mathbf{x}$ itself, it will not do so in the presence of other objects either. The proof is given in Appendix C.

It follows that one can discard $\mu\in M_h\backslash\{m\}$, for which

$$\mathrm{E}_{x,o}[P_D((x,m),\{(o,\mu)\})] \approx \underbrace{\mathrm{E}_x[P_D((x,m),\emptyset)]}_{\triangleq P_D^{\emptyset}(m,h)} \quad (51)$$

where the densities used to compute the expectations are taken from the hypothesis h. Both the left- and right-hand sides of (51) can be implemented using Monte Carlo (MC) integration. From an implementation perspective, the MC samples should be stored for future reference.

Remark: In the absence of any occlusions among all hypotheses, $\bar{P}_D(m,h)=P_D^{\emptyset}(m,h)$ (51). The complexity of the EPoD computation is then $\mathcal{O}(\mathcal{H}M_{\max}^2 N_{\mathrm{MC}})$ in the worst case, where $M_{\max}=\max\{|M_1|,\ldots,|M_{\mathcal{H}}|\}$ and N_{MC} is the number of MC samples.

Finally, we approximate existence probabilities that can be considered high enough by one, so that any combination-set A in (47) excluding the corresponding marks leads to negligible $\omega_h^m(A)$ (48).

Note that this simplification can be computed in a track-oriented manner. That is, for each *local* hypothesis corresponding to the mark m, one can go through local hypotheses of other objects included in the same global hypothesis. The simplification for the given m is likely to be the same among many global hypotheses, eventually leading to the same approximation of (47) for them.

For the next approximation step, it would be generally possible to employ the K-shortest path algorithm used in the early implementations of the GLMB filter [51]. However, only a small number of other Bernoulli components in the set $M_h\backslash\{m\}$ are likely to remain for consideration after the simplification in Section IV-B1. Therefore, we find the following exhaustive computation readily suited.

[1] This geometric assumption is key to the Lebesgue differentiation theorem used in the proof. Spaces like $\mathbb{R}^n$ satisfy this condition automatically.

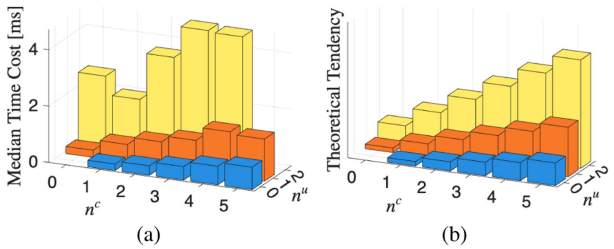


Fig. 1. Complexity analysis of the approximation in Section IV-B2. Not enough data for $n^c=5$ and $n^u=2$ in (a) were available. (a) Experimental cost. (b) Theoretical cost.

2) *Approximation of (47) Directly:* Denote with $M_h^c\subset M_h\backslash\{m\}$ the subset of *certain* marks for which existence probabilities are equal to one, and denote its cardinality with $n^c=|M_h^c|$. Similarly, denote with $M_h^u=M_h\backslash\{m\}\backslash M_h^c$ the set of *uncertain* marks, and denote its cardinality with $n^u=|M_h^u|$. For each $n=n^c, n^c+1,\ldots,n^c+n^u$, compute the terms of the outmost sum in (47) as follows. First, generate directly the subsets (i.e., combinations) $A^u\subseteq M_h^u$, with $|A^u|=\nu$ for each $\nu=0,\ldots,n^u$, such that $A=A^u\cup M_h^c$. For each such A, approximate the corresponding $(n^c+\nu+1)$-way integral in (47) with, e.g., the MC integration.

Under the simplification in Section IV-B1, the total number $2^{|M_h|-1}$ of terms in (47) shrinks to only $2^{(n^u)}$. Neglecting the complexity of evaluating $\omega_h^m(A)$ (48) and assuming the complexity of evaluating an $(n^c+\nu+1)$-way integral in (47) is roughly linear in both n^c and n^u, the computational complexity of the approximation in Section IV-B2 is $\mathcal{O}(N_{MC}(n^c+n^u)2^{(n^u)})$. The theoretical dependence for small n^c and n^u is compared with experimental time costs (median values) in Fig. 1. Details about the experiment are given later in Section V.

C. Summary of the Proposed Strategy

The proposed filter utilizes the following strategy to deal with occlusions. To distinguish it from the ESO strategy (see Section II-C), it is termed as the PRO strategy:

- P : Palm condition the prior distribution of states;
- R: reduce the dependence within the SPO-D model by computing the EPD $P_D(\cdot)$;
- O: obey the SPO model parameterized with the computed $P_D(\cdot)$ to achieve tractable filtering recursion.

The proposed filter will be referred to as MBM-PRO. Analogically, if the ESO strategy (see Section II-C) is applied instead, the algorithm will be referred to as MBM-ESO.

V. APPLICATION TO VISUAL MOT

In [61], an SPO model for pedestrian MOT using a monocular camera was developed with parameters obtained by identification and physical modeling. The MOT-17 dataset [62], [63] was considered, and the Faster R-CNN (FRCNN) BB detections included in the dataset were used for the identification. Pedestrians were modeled in 3-D with

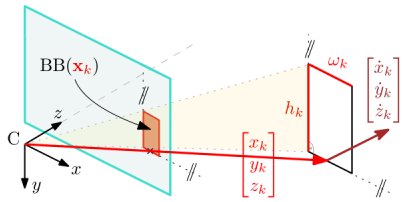


Fig. 2. Illustration of the state $\mathbf{x}_k$ (52) in red. The image plane is in turquoise, and C is the camera center.

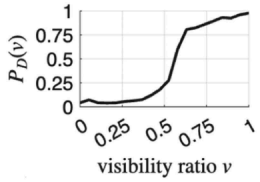


Fig. 3. PoD as a function of visibility ratio.

the state at time step k defined as [64]

$$\mathbf{x}_k = [x_k\ \dot{x}_k\ y_k\ \dot{y}_k\ z_k\ \dot{z}_k\ \omega_k\ h_k]^\top \tag{52}$$

with the positions x_k, y_k, and z_k of the lower-bottom center of the box and the width ω_k and height h_k expressed in meters, while the velocities $\dot{x}_k$, $\dot{y}_k$, and $\dot{z}_k$ in meters per second. Geometric illustration of the measurement and the state quantities is given in Fig. 2. For details, refer to [64]. The fifth state-vector element $[\mathbf{x}_k]_5{=}z_k$ will be called *depth* in the following.

The SPO model [61] involves the constant $P_D = 0.529$, which was estimated from the entire MOT-17 training dataset. Such a modeling was proven to be conceptually incorrect in [61], where the PoD was shown to depend on the *visibility* as follows.

In [61], an estimate $P_D(v)$ of the PoD was computed *conditionally* on the *visibility ratio* $v{\in}[0, 1]$, where v is available in ground-truth (GT) data [62, Table 5]. The resulting $P_D(v)$ curve is far from being constant and is shown in Fig. 3.

Remark: The true PoD likely depends on conditions such as brightness or resolution. $P_D(v)$ (see Fig. 3) is an average over such phenomena contained in the data.

Remark: The ratio v from [62] captures the visibility of pedestrian *BBs* rather than of their *bodies*. Physically, the PoD should depend on the latter visibility. To follow such a modeling, however, a more informative state than $\mathbf{x}_k$ (52) is required. Accordingly, image segmentation measurements could be better suited for such consideration than BB detections used here. Finally, establishing the corresponding PoD model may be challenging. Therefore, the former (i.e., BB) PoD visibility dependence modeling is adopted in this article. While *heuristic* models for such dependence exist [41], [42], [43], the function $P_D(v)$ (see Fig. 3) *estimated from data* is used in this article.

The BB visibility ratio v definition was rather brief in [62]. In this article, the ratio v is modeled without any reference to ground plane; cf. [62, Sect. 2.5]. In particular,

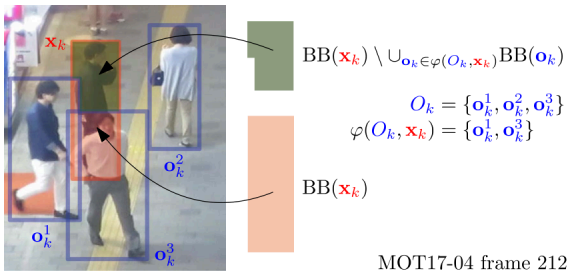


Fig. 4. Illustration of sets whose area is computed in (53).

a definition similar to [42, Sect. 4.2] (cf. [17]) is used as

$$v(\mathbf{x}_k,\ O_k) = \frac{\text{Area}\left(\text{BB}(\mathbf{x}_k)\setminus \cup_{\mathbf{o}_k\in\varphi(O_k,\mathbf{x}_k)}\text{BB}(\mathbf{o}_k)\right)}{\text{Area}\left(\text{BB}(\mathbf{x}_k)\right)} \tag{53}$$

where the function $\text{BB}(\cdot){\subset}\mathbb{R}^2$ yields the BB corresponding to the state, as illustrated in Fig. 4, and

$$\varphi(O_k, \mathbf{x}_k){=}\{\mathbf{o}_k{\in}O_k : [\mathbf{o}_k]_5{<}z_{\max},\ [\mathbf{o}_k]_5{<}[\mathbf{x}_k]_5{-}\kappa\} \tag{54}$$

is the subset of other objects that are considered eligible to occlude $\mathbf{x}_k$. The condition $[\mathbf{o}_k]_5{<}z_{\max}$ means that occlusions by objects that appear further than $z_{\max}{=}15$ m are ignored due to limited depth estimation accuracy for distant objects. The $z_{\max}$ value was selected to mitigate meaningless EPoD values for crowded pedestrians in the distance, and experiments exhibited rather small sensitivity to the value. The condition $[\mathbf{o}_k]_5{<}[\mathbf{x}_k]_5{-}\Delta$ means that occlusions by objects that are in front of $\mathbf{x}_k$ but too close in z-coordinate are ignored to ensure physical validity of occlusion. Here, $\kappa{=}\frac{0.85}{2}$ meters was set to half of a mean BB width [61]. Experiments validated that using either a smaller or larger κ may lead to meaningless results.

Assuming that the PoD depends on pedestrians via the visibility ratio *only*, it follows that

$$P_D(\mathbf{x}_k,\ O_k) = P_D\left(v(\mathbf{x}_k,\ O_k)\right) \tag{55}$$

where $P_D(\cdot)$ on the right-hand side is $P_D(v)$ (see Fig. 3). The SPO model [61] is adopted with the approximate GM implementation [61]. However, the methods developed in this article rely on the assumption that the birth density is MB, which is adopted based on [61, Sect. II.C.1]. A suitable MB is also used as the initial density. Available FRCNN detections were used as the measurements.

For each time step, estimates are drawn for simplicity using the suboptimal Estimator 1 [54, Sect. VI.A] with a *noninformative* threshold of 0.5 for existence probabilities. Since only 2-D GT data are available, estimates are further projected into 2-D for evaluation, which is achieved using the Unscented transform as in [64, Secy. IV]. Ellipsoidal gating was used with a threshold of 6. The number of global hypotheses was capped at 100, with a pruning threshold of -300 for log-weights. For hypothesis h, the Murty's M-best assignments algorithm was used with $M{=}\lceil w_h \cdot 10\rceil$ [51], [54].

A. Performance Evaluation Metric

To evaluate the estimates, CV scores such as *multiple object tracking accuracy* (MOTA), *higher order tracking accuracy* (HOTA), and *identity F1* (IDF1) are often considered in the CV literature [62], [65]. However, the CV scores may behave undesirably [66], [67] and hardly allow for performance evaluation based on user preferences. Therefore, we employ the *trajectory generalized sub-optimal assignment* (TGOSPA) metric [68], which includes several parameters: (i) metric $d(x, y)$ between GT BB x and an estimated BB y; (ii) power parameter $p \geq 1$; (iii) cutoff $c > 0$ that is the maximum possible distance for x and y to be assigned to each other; and (iv) switching penalty $\gamma > 0$. The parameters allow the TGOSPA metric to be tailored for applications such as online surveillance or offline scene understanding [67].

The TGOSPA metric value can be decomposed into errors E resulting from *(i:* E_{TP} *)* properly estimated objects, i.e., *true positives* (TPs); *(ii:* E_{FN} *)* missed objects, i.e., *false negatives* (FNs); *(iii:* E_{FP} *)* false objects, i.e., *false positives* (FPs); and *(iv:* E_{Sw} *)* properly defined *track switches* (Sw). Informally, the TGOSPA metric is given by

$$d_p^{(c,\gamma)}(\mathbf{X}, \mathbf{Y}) = \min_{\text{“over assignments between trajectories”}} \left(\underbrace{\sum_{(x,y)\in \mathrm{TP}} d(x, y)^p}_{\mathsf{E}_{\mathrm{TP}}} + \underbrace{\frac{c^p}{2}|\mathrm{FN}|}_{\mathsf{E}_{\mathrm{FN}}} + \underbrace{\frac{c^p}{2}|\mathrm{FP}|}_{\mathsf{E}_{\mathrm{FP}}} + \underbrace{\gamma^p \mathrm{Sw}}_{\mathsf{E}_{\mathrm{Sw}}} \right)^{\frac{1}{p}} \tag{56}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are sets of GT and estimated trajectories, respectively, the symbol $|\cdot|$ denotes the cardinality, and Sw counts changes within the assignments as

$$\mathrm{Sw} = \sum_{k=1}^{K} \sum_{i=1}^{|X|} \underbrace{s(\mathrm{assignment}_{k-1}^i,\ \mathrm{assignment}_k^i)}_{\triangleq \begin{cases} 0, & \text{if } \mathrm{assignment}_{k-1}^i = \mathrm{assignment}_k^i \\ 1, & \text{if } 0 \neq \mathrm{assignment}_{k-1}^i \neq \mathrm{assignment}_{k-1}^i \neq 0 \\ \frac{1}{2}, & \text{otherwise} \end{cases}} \tag{57}$$

where $\mathrm{assignment}_k^i$ is the index $j \in \{0, 1, \ldots, |Y|\}$ to which the ith GT trajectory is assigned to at time step k with 0, meaning that it is left unassigned. Note that all of TP, FP, and FN depend on the assignments as well. See [68] for a precise definition and approximate computation of (56).

The desired meaning and impact of the diffident terms in (56) vary among applications. In the following, the guidelines presented in [67] are used to select d, p, c, and γ concisely. The terms E_{TP}, E_{FN}, and E_{FP} are further decomposed to better assess performance under occlusions.

1) *TGOSPA Metric for Assessing Occlusions:* The function $d(x, y) = 1 - \mathrm{IoU}(x, y)$ with IoU being the *intersection over union* was proven to be a metric in [67], and it is adopted in this article. The maximum possible cutoff $c = 1$ is used, allowing any two BBs to be assigned as long as they have a nonempty intersection. The power parameter $p = 2.41$ is selected according to [67, eq. (32)] so that estimates with error $0.75 = \frac{c}{\sqrt[p]{2}}$ in d or larger are better to be omitted. That is, tracking algorithms are encouraged to output even very distant predictive estimates with errors up to 0.75 in the metric d. Since $p > 1$, E_{FP} and E_{FN} can be expected to have the largest influence on the final metric value, compared to E_{TP} and Sw. The switching penalty $\gamma = 2.60$ is set according to [67, eq. (32)] so that switches Sw encapsulate long-term track changes that last for at least ten time steps[2] (see [67, Sect. 4.4.2]). Such a choice of γ favors TP pairs that correspond to the same trajectories over pairs with the smallest error for a given time step. Tracking algorithms are thus encouraged to form trajectories without fragments, such as those caused by occlusions.

The visibility ratio is available for each GT BB x in the MOT-17 dataset, further denoted as v_x. The error E_{TP} can thus be decomposed as

$$\begin{aligned} \mathsf{E}_{\mathrm{TP}} &= \sum_{(x,y)\in \mathrm{TP}} (1 - v_x + v_x) d(x, y)^p \\ &= \underbrace{\sum_{(x,y)\in \mathrm{TP}} (1 - v_x) d(x, y)^p}_{\mathsf{E}_{\mathrm{TP}}^{\mathrm{o}}} + \underbrace{\sum_{(x,y)\in \mathrm{TP}} v_x d(x, y)^p}_{\mathsf{E}_{\mathrm{TP}}^{\mathrm{v}}} \end{aligned} \tag{58}$$

where $\mathsf{E}_{\mathrm{TP}}^{\mathrm{o}}$ and $\mathsf{E}_{\mathrm{TP}}^{\mathrm{v}}$ are the portions of E_{TP} corresponding to cases when the GT BBs were occluded and visible, respectively. Furthermore

$$|\mathrm{TP}| = \underbrace{\sum_{(x,y)\in \mathrm{TP}} (1 - v_x)}_{\mathsf{N}_{\mathrm{TP}}^{\mathrm{o}}} + \underbrace{\sum_{(x,y)\in \mathrm{TP}} v_x}_{\mathsf{N}_{\mathrm{TP}}^{\mathrm{v}}} \tag{59}$$

where $\mathsf{N}_{\mathrm{TP}}^{\mathrm{o}}$ and $\mathsf{N}_{\mathrm{TP}}^{\mathrm{v}}$ correspond to occluded and visible BBs, respectively. The error E_{FN} can be decomposed as

$$\mathsf{E}_{\mathrm{FN}} = \overbrace{\frac{c^p}{2} \underbrace{\sum_{x\in \mathrm{FN}} (1 - v_x)}_{\mathsf{N}_{\mathrm{FN}}^{\mathrm{o}}}}^{\mathsf{E}_{\mathrm{FN}}^{\mathrm{o}}} + \overbrace{\frac{c^p}{2} \underbrace{\sum_{x\in \mathrm{FN}} v_x}_{\mathsf{N}_{\mathrm{FN}}^{\mathrm{v}}}}^{\mathsf{E}_{\mathrm{FN}}^{\mathrm{v}}} \tag{60}$$

where $\mathsf{E}_{\mathrm{FN}}^{\mathrm{o}}$ and $\mathsf{E}_{\mathrm{FN}}^{\mathrm{v}}$ are the parts of E_{FN} for the cases when missed GT BBs were occluded or visible, respectively.

B. Results and Discussion

To analyze the performance in challenging real scenarios, the scenarios should obey the assumed SPO-D model. In particular, the camera must be static, and occlusions should be present. We thus selected MO17-02, MOT17-04, and MOT17-09 videos from the MOT-17 dataset.

Performance is evaluated for the MBM filter with: *(i)* the proposed PRO and *(ii)* the existing ESO ([41], [42], [43]; see Section II-C) occlusion-handling strategies that both utilize $P_D(\mathbf{x}_k, O_k)$ (55), and also for *(iii)* the “basic” MBM filter with constant $P_D = 0.529$ as a reference. Additionally, *(iv)* the SORT [69] and *(v)* FasterTracker [9] algorithms were considered as representatives of CV-based algorithms.

[2] The videos considered in this article were captured with 30 frames/s, i.e., 10 time steps correspond to 0.33 s of a video.

TABLE I
Results for MOT17-02 Video

	CV scores (↑)			TGOSPA metric (↓) and its decomposition									Hz (↑)
	MOTA	HOTA	IDF1	$d_p^{(c,\gamma)}$	E_{TP} |TP|	$\mathsf{E}^{\mathsf{o}}_{\mathsf{TP}}$ $\mathsf{N}^{\mathsf{o}}_{\mathsf{TP}}$	$\mathsf{E}^{\mathsf{v}}_{\mathsf{TP}}$ $\mathsf{N}^{\mathsf{v}}_{\mathsf{TP}}$	E_{FN} |FN|	$\mathsf{E}^{\mathsf{o}}_{\mathsf{FN}}$ $\mathsf{N}^{\mathsf{o}}_{\mathsf{FN}}$	$\mathsf{E}^{\mathsf{v}}_{\mathsf{FN}}$ $\mathsf{N}^{\mathsf{v}}_{\mathsf{FN}}$	E_{FP} |FP|	E_{Sw} Sw	fpps
MBM-PRO	0.287	0.298	0.348	**39.96**	700.71 7320	512.8 2007.3	187.9 5312.7	5630.5 11261	4531.1 9062.2	1099.4 2198.8	457 914	440 44	1.57
MBM-ESO	0.265	0.313	0.362	40.58	466.2 6119	369.6 1437.4	96.6 4681.6	6231 12462	4816.1 9632.2	1414.9 2829.8	517.5 1035	285 28.5	2.41
MBM	0.279	0.316	0.347	40.09	291.6 5751	158.9 1007.1	132.7 4743.9	6415 12830	5031.2 10062	1383.8 2767.5	246.5 493	330 33	2.41
SORT	0.320	0.313	0.364	40.15	315.9 6521	159.6 834.4	156.2 5686.6	6030 12060	5117.6 10235	912.4 1824.8	538 1076	425 42.5	–
FastTracker	0.827	0.581	0.679	24.77	620.2 17175	539.7 9778.7	80.5 7396.3	703.0 1406	645.4 1290.8	57.6 115.2	265.5 531	695.0 69.5	–

TABLE II
Results for MOT17-04 Video

	CV scores (↑)			TGOSPA metric (↓) and its decomposition									Hz (↑)
	MOTA	HOTA	IDF1	$d_p^{(c,\gamma)}$	E_{TP} |TP|	$\mathsf{E}^{\mathsf{o}}_{\mathsf{TP}}$ $\mathsf{N}^{\mathsf{o}}_{\mathsf{TP}}$	$\mathsf{E}^{\mathsf{v}}_{\mathsf{TP}}$ $\mathsf{N}^{\mathsf{v}}_{\mathsf{TP}}$	E_{FN} |FN|	$\mathsf{E}^{\mathsf{o}}_{\mathsf{FN}}$ $\mathsf{N}^{\mathsf{o}}_{\mathsf{FN}}$	$\mathsf{E}^{\mathsf{v}}_{\mathsf{FN}}$ $\mathsf{N}^{\mathsf{v}}_{\mathsf{FN}}$	E_{FP} |FP|	E_{Sw} Sw	fpps
MBM-PRO	0.536	0.486	0.561	**50.31**	836.0 26803	416.9 4137.1	419.1 22665.9	10377.0 20754	7022.0 14044.1	3355.0 6709.9	784.5 1569	590.0 59.0	1.33
MBM-ESO	0.514	0.479	0.556	51.20	446.2 25397	165.4 3541.2	280.8 21855.8	11080.0 22160	7320.0 14640.0	3760.0 7520.0	1007.5 2015	600.0 60.0	1.35
MBM	0.506	0.466	0.536	51.46	863.4 25399	406.6 3801.2	456.8 21597.8	11079.0 22158	7190.0 14380.0	3889.0 7778.0	699.0 1398	655.0 65.5	1.66
SORT	0.540	0.498	0.564	50.39	290.9 26084	102.9 3409.4	187.9 22674.6	10736.5 21473	7385.9 14771.8	3350.6 6701.2	913.5 1827	695 69.5	–
FastTracker	0.988	0.887	0.968	14.19	209.8 47273	108.7 17939.7	101.1 29333.3	142.0 284	120.7 241.5	21.3 42.5	150.0 300	95 9.5	–

TABLE III
Results for MOT17-09 Video

	CV scores (↑)			TGOSPA metric (↓) and its decomposition									Hz (↑)
	MOTA	HOTA	IDF1	$d_p^{(c,\gamma)}$	E_{TP} |TP|	$\mathsf{E}^{\mathsf{o}}_{\mathsf{TP}}$ $\mathsf{N}^{\mathsf{o}}_{\mathsf{TP}}$	$\mathsf{E}^{\mathsf{v}}_{\mathsf{TP}}$ $\mathsf{N}^{\mathsf{v}}_{\mathsf{TP}}$	E_{FN} |FN|	$\mathsf{E}^{\mathsf{o}}_{\mathsf{FN}}$ $\mathsf{N}^{\mathsf{o}}_{\mathsf{FN}}$	$\mathsf{E}^{\mathsf{v}}_{\mathsf{FN}}$ $\mathsf{N}^{\mathsf{v}}_{\mathsf{FN}}$	E_{FP} |FP|	E_{Sw} Sw	fpps
MBM-PRO	0.574	0.465	0.542	**19.35**	138.2 3503	103.2 723.1	35.0 2779.9	911.0 1822	741.7 1483.4	169.3 338.6	34.5 69	175 17.5	2.21
MBM-ESO	0.536	0.505	0.596	19.71	140.6 3318	109.2 646.6	31.3 2671.4	1003.5 2007	779.9 1559.9	223.6 447.1	33.0 66	140 14	4.12
MBM	0.559	0.445	0.515	20.00	47.5 3055	25.1 388.0	22.5 2667.0	1135.0 2270	909.2 1818.5	225.8 451.5	21.5 43	160 16	3.43
SORT	0.530	0.437	0.501	20.69	25.4 2840	7.9 221.1	17.5 2618.9	1242.5 2485	992.7 1985.4	249.8 499.6	22.0 44	190 19	–
FastTracker	0.915	0.614	0.690	12.78	143.4 5173	105.4 2082.6	38.1 3090.4	76.0 152	61.9 123.9	14.1 28.1	49.5 99	195 19.5	–

The SORT algorithm processes FRCNN detections without using video images directly, similarly to the MBM filters, and it is evaluated for a *fair* comparison with CV-based algorithms. The FastTracker algorithm is evaluated as a representative of re-ID-based strategies. It processes both the video images and the FRCNN detections and dominates the MOT-17 Public leader board.[3] The results for MOT17-02, MOT17-04 and MOT17-09 videos are given in Tables I–III, respectively.

As expected, FastTracker performs best in all the considered metrics and videos. The algorithm ranking based on CV scores is difficult to justify clearly and is included for completeness. In the TGOSPA metric, the performance of the filters is comparable to that of SORT, among which the MBM-PRO is the best. Detailed observations and discussions follow.

OBSERVATION 1 (MORE INFORMATIVE POD MODEL ENHANCES MEASUREMENT-TO-TRACK ASSOCIATIONS) Both MBM-PRO and MBM-ESO mostly have a higher number |TP| compared to the basic MBM in all videos. When pedestrians are visible, the occlusion-handling strategies assign considerably higher EPoD to marks compared to $P_D = 0.529$ used otherwise. Measurement-to-track associations are thus more likely for the MBM-PRO and MBM-ESO compared to the basic MBM.

OBSERVATION 2 (PRO AND ESO COMPUTATIONAL COMPLEXITIES ARE HARDLY PREDICTABLE) The algorithms' computational complexity was measured in frames processed per second (fps), which is denoted with *Hz* in the MOT-17 website.[4] Since PoD for visible pedestrians are considerably higher when using either occlusion-handling strategy, the discrepancy among different measurement-to-track associations is larger. This

[3] motchallenge.net/results/MOT17/

[4] The values for SORT and FastTracker of 143.3 and 355.1 Hz, respectively, are reported by the website motchallenge.net/results/MOT17. The values were provided by the algorithms' authors for the entire dataset and not officially evaluated by the MOTChallenge.

TABLE IV
Complexities of ESO and PRO Per Frame

	MOT17-02	MOT17-04	MOT17-09
median time PRO [ms]	225	304	186
median time ESO [ms]	56	182	36

may lead to better computational efficiency for a small number of objects, such as for MBM-ESO in MOT17-09. For a larger number of objects, however, the computational demand is likely to be larger when using either PRO or ESO. It should be noted that implementations of MBM-PRO, MBM-ESO, and the MBM filters was run in R2025b version of MATLAB software on an Apple M4 Air laptop.

Median times spent on a single call of ESO and PRO blocks of code, i.e., median times of the PoD computation for all marks for a single time step, are given in Table IV . As expected, the MBM-PRO computes all $P_D(m)$ values (for a frame) slower than the MBM-ESO in all videos.

OBSERVATION 3 (PRO INCREASES TRACKING PERFORMANCE FOR VISIBLE PEDESTRIANS OVER ESO) The MBM-PRO tracks a larger number $\mathsf{N}^{\mathrm{v}}_{\mathrm{TP}}$ of *visible* pedestrians with a similar error $\mathsf{E}^{\mathrm{v}}_{\mathrm{TP}}$ as MBM-ESO and MBM, and also as SORT except for MOT17-02 video.

OBSERVATION 4 (ESO PRESERVES TRACKING PERFORMANCE FOR VISIBLE PEDESTRIANS) The MBM-ESO tracks a similar or smaller number $\mathsf{N}^{\mathrm{v}}_{\mathrm{TP}}$ of *visible* pedestrians with a similar or smaller error $\mathsf{E}^{\mathrm{v}}_{\mathrm{TP}}$ compared to MBM-PRO and MBM. The latter indicates that MBM-ESO provides less outliers.

OBSERVATION 5 (PRO HANDLES MORE OCCLUSIONS THAN ESO) Among the properly tracked objects, MBM-PRO tracks more occluded pedestrians than MBM-ESO in all videos. This follows since the ratios $\frac{\mathsf{N}^{\mathrm{o}}_{\mathrm{TP}}}{|\mathrm{TP}|}$ are higher for MBM-PRO compared to MBM-ESO: in MOT17-02, we have $0.274>0.235$; in MOT17-04, it is $0.154>0.139$; and in MOT17-09, it is $0.206>0.195$.

OBSERVATION 6 (PRO MISSES LESS PEDESTRIANS) MBM-PRO has less |FN|, $\mathsf{N}^{\mathrm{o}}_{\mathrm{FN}}$ and also $\mathsf{N}^{\mathrm{v}}_{\mathrm{FN}}$ than MBM-ESO and MBM in all the videos. That is, MBM-PRO misses fewer pedestrians under any of the considered circumstances compared to the other filters. Furthermore, MBM-PRO mitigates $\mathsf{N}^{\mathrm{v}}_{\mathrm{FN}}$ over SORT in MOT17-09, has nearly the same $\mathsf{N}^{\mathrm{v}}_{\mathrm{FN}}$ as SORT in MOT17-04; however, it has larger $\mathsf{N}^{\mathrm{v}}_{\mathrm{FN}}$ than SORT in MOT17-02. Nevertheless, MBM-PRO misses less occluded pedestrians then SORT as it has fewer $\mathsf{N}^{\mathrm{o}}_{\mathrm{FN}}$, and also in total |FN|.

OBSERVATION 7 (ESO MAY FAIL HANDLING OCCLUSIONS) In MOT17-04, MBM-ESO provides the least number $\mathsf{N}^{\mathrm{o}}_{\mathrm{TP}}$ of occluded pedestrians. At the same time, it has the largest |FP|. That is, the use of the ESO strategy to handle occlusions may lead to reduced performance.

OBSERVATION 8 (MORE ESTIMATES LEADS TO MORE FALSE POSITIVES) MBM-PRO and MBM-ESO have more overall number |TP|+|FP| of estimates than MBM. While both MBM-PRO and MBM-ESO usually have more TPs (see Observation 1), they also have more FPs relative to MBM. In either PRO or ESO strategy, high existence probabilities of occluded Bernoulli components last longer and yield distant predictive estimates. Whenever such estimates are not assigned to GTs in TGOSPA minimization, they get counted as FPs. This issue is present in FastTracker as well, especially in MOT17-04, where the error $\mathsf{E}_{\mathrm{FP}}=150$ is only slightly lower than $\mathsf{E}_{\mathrm{TP}}=209.8$, and it has higher influence on the metric value compared to E_{FN} and E_{Sw}.

OBSERVATION 9 (HANDLING OCCLUSIONS $\neq$ HANDLING SWITCHES) Although a switch may have several graphical explanations [67], at least one Sw appears whenever a GT pedestrian is successively successfully tracked by two different trajectories (their fragments) for at least ten frames by each trajectory. It turns out that the algorithms get rewarded more for successfully estimating occluded objects than they get penalized for an occasional switch. Note that if γ was set even larger, such track fragments may fail to be counted as Sw [67]. Although FastTracker handles occlusions, it has the largest number of switches in MOT17-02 (and also in MOT17-09). Furthermore, the largest TGOSPA error term for FastTracker in MOT17-09 is the E_{Sw}. Except for MOT17-04, MBM-PRO has more switches than MBM-ESO. Nevertheless, the filters mostly have fewer switches than SORT.

OBSERVATION 10 (MODEL-BASED APPROACH CAN BEAT AD-HOC-BASED METHODS) Utilizing the same input data, the MBM-PRO beats SORT in all the videos by improving the performance of the basic MBM. That is, the use of a more realistic model led to better results. However, algorithms such as FastTracker will be hard to outperform since they process considerably larger amounts of information (entire images). It should be noted that such ad-hoc-based algorithms are usually unable to reliably quantify their estimation errors, e.g., in probabilistic terms, since they do not utilize probabilistic models, which may be problematic for safety-critical applications.

The experiment indicated that the proposed occlusion-handling strategy PRO improves tracking performance not only under occlusion, but also for visible pedestrians, and that it also mitigates false negatives. The strategy, however, provides more false positives and switches. On the other hand, this effect also appears in advanced ad-hoc-based algorithms processing not only BB detections such as the FastTracker. The proposed strategy thus appears promising for dealing with occlusions in visual tracking.

VI. APPLICABILITY AMONG DIFFERENT SENSORS

While the considered framework applies among many sensors [30], the SPO-D model utilized in this work requires the sensed data to be segmented into *clusters*; cf. [22] and [44]. Furthermore, at most one cluster per object must exist in the set of clusters for each time step. This section elaborates on the generality of this approach for dealing

with occlusions using various sensing equipment. Note that obtaining reliable model parameters is essential to enable *fully model-based* tracking.

A. Camera-Based Tracking

In the visual tracking example considered before, the said clusters are BBs given by the RFCNN detector. Many alternative BB detectors exist [62]. To find suitable measurement model parameters *for the chosen detector*, including the PoD as a function of the visibility, the method from [70] can be utilized. Note that objects such as vehicles can be tracked visually as well, for which suitable motion and measurement models must be given. Furthermore, object class likelihoods (if provided by the detector) can be incorporated into the tracking algorithm [71].

With pixelwise segmentation [15], clusters take the form of sets of pixels covering a potential object. Although the object visibility becomes physically straightforward, the PoD function model must be developed.

Stereo cameras further output depth measurements [16]. Such measurements can be easily accounted for in the considered tracking framework by adding a suitable equation alongside the BB measurement model. Direct depth measurements naturally result in considerably higher depth estimation accuracy and thus better visibility computation accuracy; cf. (53).

Employing sensors other than video cameras can apparently be done along the lines of the above text.

B. Laser-Based Tracking

LiDAR point clouds contain range measurements [26]. Employing suitable segmentation algorithm, the resulting clusters can be treated analogously to stereo cameras. Furthermore, Doppler measurements may be available depending on the considered measurement unit. The PoD function model would depend on the chosen segmentation algorithm. Laser range finder could be understood as a special case of LiDAR that lacks the vertical axis [21], [44]. Since Dai et al. [44] utilized the GLMB tracking framework, most parts of the model developed therein can be adopted directly for the proposed MBM-PRO filter.

VII. CONCLUSION

This article focused on MOT problems where occlusions among objects may arise. A principled approximation was proposed, leading to a straightforward and fully model-based strategy for handling occlusions. The essence of the strategy is the calculation of the conditional expectation related to detection probability. As the solution necessitates marked state densities to be proven optimal, the article investigated the implementation of filters using MBMs with distinct marks. The application of visual tracking demonstrated that the proposed strategy outperformed others. Future work could focus on incorporating Poisson birth or addressing implementation for unmarked objects.

APPENDIX A
RIGOROUS PALM CONDITIONING

A point process could be understood as a *collection of points drawn at random*. When the number of points is locally finite and they do not coincide, the point process can be represented as an RFS. However, counting measure representation is most common in the literature discussing Palm conditioning, see [47, Ch. 13, Ch. 15.5], and [48, Sect. 3.2] for details. In this appendix, the counting measure treatment is employed to incorporate Palm conditioning into FISST, i.e., the theory used in this article.

A. Preliminaries

In general [30], the underlying spaces $\mathcal{X}$ and $\mathcal{Z}$ can be assumed *locally compact*, *second countable*, and *Hausdorff* (LCSCH) topological spaces. With such an assumption, one can rigorously tackle applications involving more general spaces than $\mathbb{R}^n$ [30, Chapter 18]. For instance, objects may move on a manifold, such as on a globe, or objects that live in state spaces of different dimensions may coexist. LCSCH spaces are Polish [72, p. 13 and 29], i.e., any LCSCH space can be turned into a *complete separable metric* (CSM) space. A metric on the underlying state space is usually needed to measure the performance of tracking algorithms[5] [68]. Therefore, being CSM seems as the convenient assumption for $\mathcal{X}$ and $\mathcal{Z}$.

The *Dirac measure centered at* $\mathbf{x}\in\mathcal{X}$, denoted with $\delta_{\mathbf{x}}$ is the set function, such that $\delta_{\mathbf{x}}(B)=1$ if $\mathbf{x}\in B\subset\mathcal{X}$ and zero otherwise. The *counting measure* N_Ξ *associated with* $\Xi\subset\mathcal{X}$ is a set function that simply counts the number of points of Ξ within B, i.e.,

$$N_\Xi(B)=\sum_{\mathbf{x}\in\mathcal{X}}\delta_{\mathbf{x}}(B)=|\Xi\cap B|. \tag{61}$$

Let $(\mathfrak{X},\mathfrak{S},P_\Xi)$ be the probability space corresponding to the RFS Ξ. Briefly, the set $\mathfrak{X}=\uplus_{n\geq 0}\mathcal{X}^{(n)}$ is the space such that $\Xi\in\mathfrak{X}$, where $B^{(n)}$ is the set of all subsets of the closed set $B\subseteq\mathcal{X}$ containing n elements and $B^{(0)}=\{\emptyset\}$ is the empty configuration. The set $\mathfrak{S}$ is a convenient sigma algebra on $\mathfrak{X}$ (see [29, Appendix F] and [73, Ch. 4]) and P_Ξ is the probability measure.

To give the RPD, the following definitions are needed. The *first moment measure* D_Ξ is defined as

$$D_\Xi(B)=\mathrm{E}\left[N_\Xi(B)\right] \tag{62}$$

where $\mathrm{E}[\cdot]$ is the expectation operator. The first moment measure of any RFS must be absolutely continuous with respect to (w.r.t.) some reference measure; otherwise, it fails to form a locally finite set [74, p. 138], thus

$$D_\Xi(B)=\int_B D_\Xi(\mathbf{x})\,d\mathbf{x} \tag{63}$$

[5] The metric must be chosen such that the space is CSM in that metric. Being Polish, such a metric is guaranteed to exist for a LCSCH space. If $\mathcal{X}=\mathbb{R}^n$, any metric on $\mathbb{R}^n$, including the cut-off metric, can be used.

where $D_\Xi(\mathbf{x})$ is the PHD (18) of the RFS Ξ. The same holds true for the so-called *Janossy densities* [74, p. 136], which are in the FISST literature defined via *set derivatives* $\frac{\delta}{\delta X}$ [73, Ch. 4] as

$$p_\Xi(X) = \frac{\delta \beta_\Xi(\cdot)}{\delta X}, \qquad \beta_\Xi(G) = P_\Xi\left(\uplus_{n\geq 0} G^{(n)}\right) \tag{64}$$

where $\beta_\Xi(G)$ is the *belief mass function* evaluated at the closed set $G \subseteq \mathcal{X}$ and $p_\Xi(\{\mathbf{x}^1, \ldots, \mathbf{x}^n\})$ is the nth-order Janossy density aka the FISST density function. Probabilities can thus be computed as [29, p. 714]

$$P_\Xi(S) = \int_{\chi^{-1}(S \cap \mathcal{X}^{(|X|)})} p_\Xi(X) \delta X \tag{65}$$

where χ is the mapping from vectors to sets defined as $\chi([\mathbf{x}^1, \ldots, \mathbf{x}^n]^\top) = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$.

The *Campbell measure* C_Ξ corresponding to Ξ is the measure on $\mathcal{X} \times \mathfrak{X}$ satisfying (cf. [47, pp. 270–271])

$$C_\Xi(B \times S) = \mathrm{E}\left[\, N_\Xi(B) \cdot \mathbf{1}_S \,\right] \tag{66a}$$

$$= \int_{\chi^{-1}(S \cap \mathcal{X}^{(|X|)})} |X \cap B| \, p_\Xi(X) \delta X \tag{66b}$$

for each measurable $B \subseteq \mathcal{X}$ and each event $S \in \mathfrak{S}$, where $\mathbf{1}_S$ is the indicator random variable of the event $S \in \mathfrak{S}$.

B. Palm Conditioning Within FISST

The *Palm distribution* $P_{\Xi|\mathbf{x}\in\Xi}$ is defined as the Radon–Nikodým derivative of the Campbell measure in the first argument w.r.t the first moment measure D_Ξ (62). Since D_Ξ (62) is absolutely continuous, it follows that for any fixed event $S \in \mathfrak{S}$,

$$C_\Xi(B \times S) = \int_B P_{\Xi|\mathbf{x}\in\Xi}(S|\mathbf{x}) D_\Xi(\mathbf{x}) \, d\mathbf{x} \tag{67}$$

which can be understood as the *disintegration* of C_Ξ w.r.t. the first component. Note that $P_{\Xi|\mathbf{x}\in\Xi}(S|\mathbf{x})$ is a probability measure in S and a function in $\mathbf{x}$. Its density can be defined using several methods. In the point process literature, it is common to take the Radon–Nikodým derivative of $P_{\Xi|\mathbf{x}\in\Xi}(\cdot\,|\mathbf{x})$ w.r.t. the *unit-rate Poisson point process* [47, Ch. 10.4]. In FISST, however, belief mass functions are used instead.

For a fixed measurable $B \subseteq \mathcal{X}$, define

$$\beta_{C,\Xi}(G; B) \triangleq C_\Xi\left(B \times (\cup_{n\geq 0} G^{(n)})\right) \tag{68a}$$

$$= \int_B \underbrace{P_{\Xi|\mathbf{x}\in\Xi}\left((\cup_{n\geq 0} G^{(n)})|\mathbf{x}\right)}_{\triangleq \beta_{\Xi|\mathbf{x}\in\Xi}(G|\mathbf{x})} D_\Xi(\mathbf{x}) \, d\mathbf{x} \tag{68b}$$

where $\beta_{C,\Xi}$ is the belief mass function corresponding to C_Ξ in the second argument and $\beta_{\Xi|\mathbf{x}\in\Xi}$ is the belief mass function corresponding to $P_{\Xi|\mathbf{x}\in\Xi}$ with $\mathbf{x}$ fixed. Note that $\beta_{C,\Xi}(G; B)$ (68a) can be used instead of $C_\Xi(B \times S)$ thanks to the Choquet theorem; see [29, p. 713] and [73, Ch.3]. Assuming that it exists, the set derivative of $\beta_{C,\Xi}(G; B)$ (68a) where B is still fixed is denoted with

$$p_{C,\Xi}(X; B) \triangleq \frac{\delta \beta_{C,\Xi}(\cdot\,; B)}{\delta X} = \int_B \underbrace{\frac{\delta \beta_{\Xi|\mathbf{x}\in\Xi}(\cdot\,|\mathbf{x})}{\delta X}}_{\triangleq p_{\Xi|\mathbf{x}\in\Xi}(X|\mathbf{x})} D_\Xi(\mathbf{x}) \, d\mathbf{x} \tag{69}$$

where $X \subset \mathcal{X}$ is a finite set such that $\mathbf{x} \in X$ and $\mathbf{x} \in B$, while $p_{\Xi|\mathbf{x}\in\Xi}$ is the FISST density corresponding to $\beta_{\Xi|\mathbf{x}\in\Xi}$, i.e., the Palm density. Substituting $X = O \cup \{\mathbf{x}\}$ into (69) yields

$$p_{C,\Xi}(O \cup \{\mathbf{x}\}; B) = \int_B \underbrace{p_{\Xi|\mathbf{x}\in\Xi}(O \cup \{\mathbf{x}\}|\mathbf{x})}_{\triangleq p_{\Theta|\mathbf{x}\in\Xi}(O|\mathbf{x})} D_\Xi(\mathbf{x}) d\mathbf{x} \tag{70}$$

where the function $O \mapsto p_{\Xi|\mathbf{x}\in\Xi}(O \cup \{\mathbf{x}\}|\mathbf{x})$ for a given $\mathbf{x}$ is defined as the desired RPD, i.e.,

$$p_{\Theta|\mathbf{x}\in\Xi}(O|\mathbf{x}) \triangleq p_{\Xi|\mathbf{x}\in\Xi}(O \cup \{\mathbf{x}\}|\mathbf{x}) \tag{71}$$

with $\Theta = \Xi \setminus \{\mathbf{x}\}$ being the RFS of points in Ξ besides $\mathbf{x}$.

The relation (17) suitable for practical computation of RPD remains to be established. Employing (66b) directly, an alternative form of (68a) involves the FISST density

$$\beta_{C,\Xi}(G; B) = \int_G \left(\sum_{\mathbf{x}\in X} \delta_{\mathbf{x}}(B)\right) p_\Xi(X) \, \delta X \tag{72a}$$

$$= \int_B \int_G p_\Xi(X \cup \{\mathbf{x}\}) \, \delta X \, d\mathbf{x} \tag{72b}$$

where the representation (61) was used and the last two equations follow as a simple generalization of the proof of [75, Thm. 2], i.e., a straightforward extension of [29, eq. (4.74)] from Dirac "functions" to measures.

Using the fundamental theorem of multiobject calculus [73, pp. 159–161], the set derivative (69) of $\beta_{C,\Xi}(G; B)$ (72b) thus also equals to

$$p_{C,\Xi}(X; B) = \frac{\delta \beta_{C,\Xi}(\cdot\,; B)}{\delta X} = \int_B p_\Xi(X \cup \{\mathbf{x}\}) d\mathbf{x} \tag{73}$$

provided that $p_\Xi(X) < M$ for some large $M \in \mathbb{R}^+$ for all $X \in \mathfrak{X}$ using the Lebesgue dominated convergence theorem. Substituting $X = O \cup \{\mathbf{x}\}$ into (73) and using (70) yields

$$\int_B p_\Xi(O \cup \{\mathbf{x}\}) \, d\mathbf{x} = \int_B p_{\Theta|\mathbf{x}\in\Xi}(O|\mathbf{x}) \, D_\Xi(\mathbf{x}) \, d\mathbf{x} \tag{74}$$

which must hold for any measurable $B \subseteq \mathcal{X}$, and thus, the integrands must be equal almost everywhere (provided that $\mathcal{X}$ is *directionally limited* [60, p. 7] using the Lebesgue differentiation theorem). Assuming that the densities are continuous, the integrands are equal everywhere in $\mathcal{X}$; thus

$$p_\Xi(O \cup \{\mathbf{x}\}) = p_{\Theta|\mathbf{x}\in\Xi}(O|\mathbf{x}) \, D_\Xi(\mathbf{x}) \tag{75}$$

for any $\mathbf{x} \in \mathcal{X}$. Dividing both sides by $D_\Xi(\mathbf{x})$, while assuming it is nonzero, yields the desired relation (17).

APPENDIX B
PROOF OF PROPOSITION 1 AND ITS SPECIAL CASE

The proof of Proposition 1 uses the following Lemma; cf. [47, p. 512] and [48, p. 48].

LEMMA 1 (CAMPBELL–MECKE FORMULA FOR RFSS) Let a function $f : \mathcal{X}\times\mathfrak{X}\to\mathbb{R}^+$ be measurable w.r.t. the Campbell measure. Let Ξ be an RFS on $\mathcal{X}$ with a density $p_\Xi(X)$ and PHD $D_\Xi(\mathbf{x})$. Then

$$\mathrm{E}_\Xi\left[\sum_{\mathbf{x}\in\Xi} f(\mathbf{x}, \Xi\backslash\{\mathbf{x}\})\right] = \int \mathrm{E}_{\Theta|\mathbf{x}}\left[f(\mathbf{x}, \Theta)\right] D_\Xi(\mathbf{x})d\mathbf{x}. \quad (76)$$

PROOF: Directly

$$\begin{aligned}\mathrm{E}_\Xi\left[\sum_{\mathbf{x}\in\Xi} f(\mathbf{x}, \Xi\backslash\{\mathbf{x}\})\right] &= \int \sum_{\mathbf{x}\in X} f(\mathbf{x}, X\backslash\{\mathbf{x}\}) p_\Xi(X)\delta X \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{i=1}^{n}\int\int\cdots\int\int\cdots\int \\ &\quad f(\mathbf{x}^i, \{\mathbf{x}^1, \ldots, \mathbf{x}^{i-1}, \mathbf{x}^{i+1}, \ldots, \mathbf{x}^n\}) \\ &\quad \times p_\Xi(\{\mathbf{x}^1, \ldots, \mathbf{x}^{i-1}, \mathbf{x}^{i+1}, \ldots, \mathbf{x}^n\}\cup\{\mathbf{x}^i\}) \\ &\quad \times d\mathbf{x}^1\cdots d\mathbf{x}^{i-1}d\mathbf{x}^{i+1}\cdots d\mathbf{x}^n d\mathbf{x}^i. \quad (77a)\end{aligned}$$

For each n, the integrals for $i=1, \ldots, n$ are the same. Noting that the summand for $n = 0$ can be neglected

$$\begin{aligned}&\mathrm{E}_\Xi\left[\sum_{\mathbf{x}\in\Xi} f(\mathbf{x}, \Xi\backslash\{\mathbf{x}\})\right] \\ &= \int \sum_{n=1}^{+\infty}\frac{n}{n!}\int\cdots\int f(\mathbf{x}, \{\mathbf{o}^1, \ldots, \mathbf{o}^{n-1}\}) \\ &\quad \times p_\Xi(\{\mathbf{o}^1, \ldots, \mathbf{o}^{n-1}\}\cup\{\mathbf{x}\})d\mathbf{o}^1\cdots d\mathbf{o}^{n-1}\cdot d\mathbf{x} \quad (78a) \\ &= \int \sum_{m=0}^{+\infty}\frac{1}{m!}\int\cdots\int f(\mathbf{x}, \{\mathbf{o}^1, \ldots, \mathbf{o}^{m}\}) \\ &\quad \times \underbrace{p_\Xi(\{\mathbf{o}^1, \ldots, \mathbf{o}^{m}\}\cup\{\mathbf{x}\})}_{p_{\Theta|\mathbf{x}}(\{\mathbf{o}^1,\ldots,\mathbf{o}^m\}|\mathbf{x})D_\Xi(\mathbf{x})} d\mathbf{o}^1\cdots d\mathbf{o}^m\cdot d\mathbf{x} \quad (78b) \\ &= \int\int f(\mathbf{x}, O)p_{\Theta|\mathbf{x}}(O|\mathbf{x})\delta O\, D_\Xi(\mathbf{x})d\mathbf{x} \quad (78c)\end{aligned}$$

using substitution $m{=}n{-}1$ and the RPD (17). □

COROLLARY 2 If the space $\mathcal{X}$ is a joint space with marks $\mathcal{X}{=}\mathcal{X}^{\mathcal{S}}\times\mathcal{M}$ and the RFS Ξ is a marked RFS on $\mathcal{X}^{\mathcal{S}}$ with marks in $\mathcal{M}$, then (76) from Lemma 1 becomes

$$\begin{aligned}&\mathrm{E}_\Xi\left[\sum_{\mathbf{x}\in\Xi} f(\mathbf{x}, \Xi\backslash\{\mathbf{x}\})\right] \\ &\quad = \sum_{m\in\mathcal{M}} \mathrm{E}_{\Theta,x|(\cdot,m)}\left[f\left((x, m), \Theta\right)\right]\cdot D(m) \quad (79)\end{aligned}$$

where the inner expectation is a function of m, computed over the density $p_{\Theta|\mathbf{x}}(O|(x, m))\cdot D(x|m)$ for each m fixed, and the PHD of Ξ is $D_\Xi(x, m) = D(x|m)\cdot D(m)$, see (37).

PROOF OF PROPOSITION 1: First, from the structure of Q (32), the optimal parameter can be established as

$$\mathcal{P}^* = \underset{\mathcal{P}:\mathcal{X}\to[0,1]}{\arg\min}\ D_{KL}\left(p \,\|\, q_{\mathcal{P}}\right). \quad (80)$$

Given the definitions of $p(\cdot)$ (30) and $q(\cdot, \mathcal{P})$ (31), notice that the KLD in (80) can be written as

$$D_{KL}\left(p \,\|\, q_{\mathcal{P}}\right) = \mathrm{E}_{X_k}\left[\mathrm{E}_{Z_k|X_k}\left[\log\frac{p_{\text{A-SPO-D}_k}(Z_k|X_k)}{p_{\text{A-SPO}_k}(Z_k|X_k,\ \mathcal{P})}\right]\right]. \quad (81)$$

The following adjustments of the inner expectation follow the classic proof [76, p. 277]. We have that

$$\begin{aligned}&\mathrm{E}_{Z_k|X_k}\left[\log\frac{p_{\text{A-SPO-D}_k}(Z_k|X_k)}{p_{\text{A-SPO}_k}(Z_k|X_k,\ \mathcal{P})}\right] \\ &\quad = \mathrm{E}_{Z_k|X_k}\left[\log\frac{p_{\text{A-SPO-D}_k}(Z_k|X_k)}{p_{\text{A-SPO}_k}(Z_k|X_k,\ \mathcal{P}^{\mathrm{g}})}\right] \\ &\qquad + \mathrm{E}_{Z_k|X_k}\left[\log\frac{p_{\text{A-SPO}_k}(Z_k|X_k,\ \mathcal{P}^{\mathrm{g}})}{p_{\text{A-SPO}_k}(Z_k|X_k,\ \mathcal{P})}\right] \quad (82)\end{aligned}$$

where the *guessed* $\mathcal{P}^g(\mathbf{x}_k){=}\mathrm{E}_{O_k|\mathbf{x}_k}[\,P_D(\mathbf{x}_k, O_k)\,]$ is to be proven to be the optimal $\mathcal{P}^*$. The first term on the right-hand side of (82) does not depend on $\mathcal{P}$, and since KLD is nonnegative, it can be neglected. Using the product form of the auxiliary SPO measurement likelihood (28), we have

$$\begin{aligned}\mathrm{E}_{Z_k|X_k}\left[\log\frac{p_{\text{A-SPO}_k}(Z_k|X_k,\ \mathcal{P}^{\mathrm{g}})}{p_{\text{A-SPO}_k}(Z_k|X_k,\ \mathcal{P})}\right] &= \cancel{\mathrm{E}_{Z_k|X_k}\left[\log\frac{p_{C_k}(Z^0)}{p_{C_k}(Z^0)}\right]} \\ &\quad + \mathrm{E}_{Z_k|X_k}\left[\sum_{\mathbf{x}_k\in X_k}\log\frac{p_{D(\mathbf{x}_k)}(Z^{\mathcal{L}(\mathbf{x}_k)}|\mathbf{x}_k, \mathcal{P}^g)}{p_{D(\mathbf{x}_k)}(Z^{\mathcal{L}(\mathbf{x}_k)}|\mathbf{x}_k, \mathcal{P})}\right]. \quad (83)\end{aligned}$$

$D_{\mathrm{KL}}(p \,\|\, q_{\mathcal{P}})$ in (80) can thus be substituted with

$$\Delta = \mathrm{E}_{X_k}\left[\sum_{\mathbf{x}_k\in X_k}\mathrm{E}_{Z_k|X_k}\left[\log\frac{p_{D(\mathbf{x}_k)}(Z^{\mathcal{L}(\mathbf{x}_k)}|\mathbf{x}_k, \mathcal{P}^g)}{p_{D(\mathbf{x}_k)}(Z^{\mathcal{L}(\mathbf{x}_k)}|\mathbf{x}_k, \mathcal{P})}\right]\right]. \quad (84)$$

Note that the measurements $Z^{\mathcal{L}(\mathbf{x}_k)}$ given X_k are Bernoulli distributed. Using [30, eq. (3.53)], the inner expectation

$$\begin{aligned}&\mathrm{E}_{Z_k|X_k}\left[\log\frac{p_{D(\mathbf{x}_k)}(Z^{\mathcal{L}(\mathbf{x}_k)}|\mathbf{x}_k, \mathcal{P}^g)}{p_{D(\mathbf{x}_k)}(Z^{\mathcal{L}(\mathbf{x}_k)}|\mathbf{x}_k, \mathcal{P})}\right] \\ &\quad = (1 - P_D(\mathbf{x}_k, X_k\backslash\{\mathbf{x}_k\}))\log\left(\frac{1-\mathcal{P}^g(\mathbf{x}_k)}{1-\mathcal{P}(\mathbf{x}_k)}\right) \\ &\qquad + P_D(\mathbf{x}_k, X_k\backslash\{\mathbf{x}_k\})\int\log\left(\frac{\mathcal{P}^g(\mathbf{x}_k)\cancel{L(z|\mathbf{x}_k)}}{\mathcal{P}(\mathbf{x}_k)\cancel{L(z|\mathbf{x}_k)}}\right)L(z|\mathbf{x}_k)dz. \quad (85)\end{aligned}$$

Using Lemma 1, the shorthand variable Δ (84) becomes

$$\begin{aligned}\Delta &= \int \mathrm{E}_{O_k|\mathbf{x}_k}[P_D(\mathbf{x}_k, O_k)]\log\left(\frac{\mathcal{P}^g(\mathbf{x}_k)}{\mathcal{P}(\mathbf{x}_k)}\right) \\ &\quad \times \cancel{\int L(z|\mathbf{x}_k)dz}\cdot D_\Xi(\mathbf{x}_k)d\mathbf{x}_k \\ &\quad + \int \mathrm{E}_{O_k|\mathbf{x}_k}[1{-}P_D(\mathbf{x}_k, O_k)]\log\left(\frac{1-\mathcal{P}^g(\mathbf{x}_k)}{1-\mathcal{P}(\mathbf{x}_k)}\right)D_\Xi(\mathbf{x}_k)d\mathbf{x}_k. \quad (86)\end{aligned}$$

Since $\mathcal{P}^g(\mathbf{x}_k) = \mathrm{E}_{O_k|\mathbf{x}_k}[\,P_D(\mathbf{x}_k, O_k)\,]$, it follows that (86) is zero if and only if $\mathcal{P} = \mathcal{P}^g$ and thus $\mathcal{P}^* = \mathcal{P}^g$. □

COROLLARY 3 (SPECIAL CASE OF PROPOSITION 1) Assume that $\mathcal{P}$ is a function of only the mark in (80), i.e., $\mathcal{P}: \mathcal{M} \to [0, 1]$. Take $\mathcal{P}^g(m)$ to be $\mathrm{E}_{O_k, x_k|(\cdot,m)}[\, P_D((x_k, m)), O_k)\,]$. Using Corollary 2, the variable Δ (84) can be adjusted further as

$$\begin{aligned}\Delta &= \sum_{m\in\mathcal{M}} \mathrm{E}_{O_k,x_k|(\cdot,m)}[1-P_D((x_k,m),O_k)]\log\left(\frac{1-\mathcal{P}^g(m)}{1-\mathcal{P}(m)}\right)D(m)\\ &+ \sum_{m\in\mathcal{M}} \mathrm{E}_{O_k,x_k|(\cdot,m)}[P_D((x_k,m),O_k)]\log\left(\frac{\mathcal{P}^g(m)}{\mathcal{P}(m)}\right)D(m) \quad (87)\end{aligned}$$

which is zero if and only if $\mathcal{P}(m)$ is equal to $\mathrm{E}_{O_k,x_k|(\cdot,m)}[\, P_D((x_k, m)), O_k)\,]$, i.e., $\mathcal{P}^* = \mathcal{P}^g$. □

APPENDIX C
PROOF OF COROLLARY 1

Corollary 1 is a special case of the following lemma.

LEMMA 2 Let $p_{\mathbf{x}}$ and $p_{\mathbf{o}}$ be given continuous probability densities on a *directionally limited* [60, p. 7] (complete and) separable metric space $\mathcal{X}$, then the implication

$$\begin{aligned}&\int_{A_1}\int_{A_2} P_D(\mathbf{x},\{\mathbf{o}\})\, p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{o}}(\mathbf{o})\, d\mathbf{x}d\mathbf{o}\\ &\qquad = \int_{A_1}\int_{A_2} P_D(\mathbf{x},\emptyset)\, p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{o}}(\mathbf{o})\, d\mathbf{x}d\mathbf{o} \quad (88a)\\ \Rightarrow\ &\int_{B_1}\int_{B_2} P_D(\mathbf{x},O\uplus\{\mathbf{o}\})\, p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{o}}(\mathbf{o})\, d\mathbf{x}d\mathbf{o}\\ &\qquad = \int_{B_1}\int_{B_2} P_D(\mathbf{x},O)\, p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{o}}(\mathbf{o})\, d\mathbf{x}d\mathbf{o} \quad (88b)\end{aligned}$$

holds for all measurable $A_1, A_2,\ B_1, B_2 \subseteq \mathcal{X}$.

PROOF: The integrand in (88a) is a product of integrable functions by Axiom 1 and thus locally integrable. Since A_1 and A_2 are arbitrary measurable sets on $\mathcal{X}$, open balls can be taken in the Lebesgue differentiation theorem[6] [60, pp. 3–7] to claim that the integrands in (88a) are equal almost everywhere in $\mathcal{X}\times\mathcal{X}$, i.e., for all Lebesgue points

$$P_D(\mathbf{x},\{\mathbf{o}\})\, p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{o}}(\mathbf{o}) = P_D(\mathbf{x},\emptyset)\, p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{o}}(\mathbf{o}). \quad (89)$$

Therefore, wherever $p_{\mathbf{x}}(\mathbf{x})\neq 0$ and $p_{\mathbf{o}}(\mathbf{o})\neq 0$, i.e., for almost all $\mathbf{x}\in\mathrm{supp}(p_{\mathbf{x}})$ and $\mathbf{o}\in\mathrm{supp}(p_{\mathbf{o}})$, we have also that $P_D(\mathbf{x},\{\mathbf{o}\})=P_D(\mathbf{x},\emptyset)$, with supp being the support. Thus by Axiom 1, we also have that $P_D(\mathbf{x}, O\uplus\{\mathbf{o}\})=P_D(\mathbf{x},O)$ for almost all $\mathbf{x}\in\mathrm{supp}(p_{\mathbf{x}})$ and $\mathbf{o}\in\mathrm{supp}(p_{\mathbf{o}})$. Since $p_{\mathbf{x}}(\mathbf{x})$ and $p_{\mathbf{o}}(\mathbf{o})$ are zero almost everywhere in the complements of $\mathrm{supp}(p_{\mathbf{x}})$ and $\mathrm{supp}(p_{\mathbf{o}})$, respectively, we have that

$$P_D(\mathbf{x},O\uplus\{\mathbf{o}\})\, p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{o}} = P_D(\mathbf{x},O)\, p_{\mathbf{x}}(\mathbf{x})p_{\mathbf{o}}(\mathbf{o}) \quad (90)$$

almost everywhere in $\mathcal{X}$. Integrating (90) on the prescribed B_1 and B_2 yields (88b), which concludes the proof. □

[6] We use [60, Thm. 1.8] under [60, Remark 1.13] and [60, Example 1.15(f)]. For $\mathcal{X} = \mathbb{R}^n$, no further assumptions would be needed assuming that the reference measure is Radon [60, Remark 1.13].

Corollary 1 results directly from Lemma 2 by setting $A_1 = A_2 = B_1 = B_2 = \mathcal{X}$ and noting that $p_{\mathbf{o}}(\mathbf{o})$ integrates to one on $\mathcal{X}$. Note that if the densities are, e.g., Gaussian, the assumption (88a) can be satisfied only approximately.

REFERENCES

[1] B.-N. Vo et al., *Multitarget Tracking*. Hoboken, NJ, USA: Wiley, 2015, pp. 1–15.

[2] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA, USA: Artech House, 1999.

[3] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comp. Vis. Pattern Rec.*, 2012, pp. 3354–3361.

[4] M. Abbaspour and M. A. Masnadi-Shirazi, "Online multi-object tracking with δ-GLMB filter based on occlusion and identity switch handling," *Image Vis. Comput.*, vol. 127, 2022, Art. no. 104553.

[5] T. Rathnayake, A. K. Gostar, R. Hoseinnezhad, R. Tennakoon, and A. Bab-Hadiashar, "On-line visual tracking with occlusion handling," *Sensors*, vol. 20, no. 3, 2020, Art. no. 929.

[6] N. L. Baisa, "Occlusion-robust online multi-object visual tracking using a GM-PHD filter with CNN-based re-identification," *J. Vis. Commun. Image Rep.*, vol. 80, 2021, Art. no. 103279.

[7] A. Specker, D. Stadler, L. Florin, and J. Beyerer, "An occlusion-aware multi-target multi-camera tracking system," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2021, pp. 4168–4177.

[8] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "BoT-SORT: Robust associations multi-pedestrian tracking," 2022, *arXiv:2206.14651*.

[9] H. Hashempoor and Y. D. Hwang, "FastTracker: Real-time and accurate visual tracking," 2025, *arXiv:2508.14370*.

[10] Y. Zhang et al., "ByteTrack: Multi-object tracking by associating every detection box," *Avidan*, vol. 13682, S., Brostow, G., Cissé, M., Farinella, and G.M., Hassner, T. (eds) Computer Vision ECCV, Springer, Cham, 2022. [Online]. Available: https://doi.org/10.1007/978-3-031-20047-2_1

[11] Z. Liu, X. Wang, C. Wang, W. Liu, and X. Bai, "SparseTrack: Multi-object tracking by performing scene decomposition based on pseudo-depth," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 5, pp. 4870–4882, May 2025, doi: 10.1109/TCSVT.2024.3524670.

[12] K. G. Quach, H. Le, P. Nguyen, C. N. Duong, T. D. Bui, and K. Luu, "Depth perspective-aware multiple object tracking," in *Engineering Applications of AI and Swarm Intelligence*. Berlin, Germany: Springer, 2023.

[13] S. A. Memon, H. Son, W.-G. Kim, A. M. Khan, M. Shahzad, and U. Khan, "Tracking multiple unmanned aerial vehicles through occlusion in low-altitude airspace," *Drones*, vol. 7, no. 4, 2023, Art. no. 24.

[14] D. Musicki, "Limits of linear multitarget tracking," in *Proc. 7th Int. Conf. Inf. Fusion*, 2005, Art. no. 6.

[15] B. Wu and R. Nevatia, "Tracking of multiple, partially occluded humans based on static body part detection," in *Proc. IEEE Comput. Soci. Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 951–958.

[16] A. Ess, B. Leibe, K. Schindler, and L. van Gool, "Robust multiperson tracking from a mobile platform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1831–1846, Oct. 2009.

[17] C. Wojek, S. Walk, S. Roth, K. Schindler, and B. Schiele, "Monocular visual scene understanding: Understanding multi-object traffic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 882–897, Apr. 2013.

[18] Y. Ma and Q. Chen, "Depth assisted occlusion handling in video object tracking," in *Advances in Visual Computing*. Berlin, Germany: Springer, 2010, pp. 449–460.

[19] A. O. Ercan, A. E. Gamal, and L. J. Guibas, "Object tracking in the presence of occlusions via a camera network," in *Proc. 6th Int. Symp. Inf. Proc. Sens. Net.*, 2007, pp. 509–518.

[20] M. Yang, Y. Liu, L. Wen, Z. You, and S. Z. Li, "A probabilistic framework for multitarget tracking with mutual occlusions," in *Proc. IEEE Conf. Comp. Vis. Pattern Recognit.*, 2014, pp. 1298–1305.

[21] K. Wyffels and M. Campbell, "Negative observations for multiple hypothesis tracking of dynamic extended objects," in *Proc. Amer. Control Conf.*, 2014, pp. 642–647.
[22] K. Wyffels and M. Campbell, "Negative information for occlusion reasoning in dynamic extended multiobject tracking," *IEEE Trans. Robot.*, vol. 31, no. 2, pp. 425–442, Apr. 2015.
[23] B. Yang and R. Yang, "Interactive particle filter with occlusion handling for multi-target tracking," in *Proc. 12th Int. Conf. Fuzzy Syst. Knowl. Discov.*, 2015, pp. 1945–1949.
[24] H.-N. Hu et al., "Joint monocular 3D vehicle detection and tracking," in *Proc. IEEE/CVF Intern. Conf. Comp. Vis.*, 2019, pp. 5389–5398.
[25] S. Jovanoska, F. Govaers, R. Thomä, and W. Koch, "Dynamic occlusion handling in the PHD filter for range-only tracking: Proof of concept," in *Proc. Sens. Data Fusion: Trends, Solutions Appl.*, 2014, pp. 1–6.
[26] J.-F. Chen, C.-C. Wang, and C.-F. Chou, "Multiple target tracking in occlusion area with interacting object models in urban environments," *Robot. Auton. Syst.*, vol. 103, pp. 68–82, 2018.
[27] A. Andriyenko, S. Roth, and K. Schindler, "An analytical formulation of global occlusion reasoning for multi-target tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2011, pp. 1839–1846.
[28] Y.-M. Song, K. Yoon, Y.-C. Yoon, K. C. Yow, and M. Jeon, "Online multi-object tracking with GMPHD filter and occlusion group management," *IEEE Access*, vol. 7, pp. 165103–165121, 2019.
[29] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, 2007.
[30] R. P. S. Mahler, *Advances in Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, 2014.
[31] R. Streit, R. B. Angle, and M. Efe, *Analytic Combinatorics for Multiple Object Tracking*. Berlin, German: Springer, 2021.
[32] D. Y. Kim, B.-N. Vo, B.-T. Vo, and M. Jeon, "A labeled random finite set online multi-object tracker for video data," *Pattern Recognit.*, vol. 90, pp. 377–389, 2019.
[33] M. Beard, B.-T. Vo, and B.-N. Vo, "Bayesian multi-target tracking with merged measurements using labelled random finite sets," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1433–1447, Mar. 2015.
[34] D. E. Clark and J. Houssineau, "FAA di Bruno's formula for Gateaux differentials and interacting stochastic population processes," 2012, *arXiv:1202.0264*.
[35] T. Yang, Q. Pan, J. Li, and S. Li, "Real-time multiple objects tracking with occlusion handling in dynamic scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 970–975.
[36] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle, "Appearance models for occlusion handling," *Image Vis. Comput.*, vol. 24, no. 11, pp. 1233–1243, 2006.
[37] K. Granstrom, C. Lundquist, and O. Orguner, "Extended target tracking using a Gaussian-mixture PHD filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 3268–3286, Oct. 2012.
[38] M. Motro and J. Ghosh, "Measurement-wise occlusion in multi-object tracking," in *Proc. 21st Int. Conf. Inf. Fusion*, 2018, pp. 2384–2391.
[39] S. Jovanoska, F. Govaers, R. Thomä, and W. Koch, "Dynamic-occlusion likelihood incorporation in a PHD filter based range-only tracking system," in *Proc. 18th Int. Conf. Inf. Fusion*, 2015, pp. 1078–1084.
[40] L. Strand, J. Honer, and A. Knoll, "Modeling inter-vehicle occlusion scenarios in multi-camera traffic surveillance systems," in *Proc. 26th Int. Conf. Inf. Fusion*, 2023, pp. 1–8.
[41] J. Ong, B.-T. Vo, B.-N. Vo, D. Y. Kim, and S. Nordholm, "A Bayesian filter for multi-view 3D multi-object tracking with occlusion handling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2246–2263, May 2022.
[42] L. V. Ma, T. T. D. Nguyen, B.-N. Vo, H. Jang, and M. Jeon, "Track initialization and re-identification for 3D multi-view multi-object tracking," *Inf. Fusion*, vol. 111, 2024, Art. no. 102496.
[43] L. Van Ma, T. T. D. Nguyen, C. Shim, D. Y. Kim, N. Ha, and M. Jeon, "Visual multi-object tracking with re-identification and occlusion handling using labeled random finite sets," *Pattern Recognit.*, vol. 156, 2024, Art. no. 110785.
[44] K. Dai, Y. Wang, J.-S. Hu, K. Nam, and C. Yin, "Intertarget occlusion handling in multiextended target tracking based on labeled multi-Bernoulli filter using laser range finder," *IEEE/ASME Trans. Mechatron.*, vol. 25, no. 4, pp. 1719–1728, Aug. 2020.
[45] L. Lamard, R. Chapuis, and J.-P. Boyer, "Dealing with occlusions with multi targets tracking algorithms for the real road context," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 371–376.
[46] L. Lamard, R. Chapuis, and J.-P. Boyer, "CPHD filter addressing occlusions with pedestrians and vehicles tracking," in *Proc. IEEE Intell. Veh. Symp.*, 2013, pp. 1125–1130.
[47] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, 2nd ed. Berlin, Germany: Springer, 2008.
[48] A. Baddeley, I. Bárány, R. Schneider, and W. Weil, *Stochastic Geometry: Lectures Given at the C. I. M. E. Summer School held in Martina Franca, Italy, September 13–18, 2004*. Berlin, Germany: Springer, 2007.
[49] B.-T. Vo and B.-N. Vo, "A random finite set conjugate prior and application to multi-target tracking," in *Proc. 7th Int. Conf. Intell. Sens. Sens. Netw. Inf. Process.*, 2011, pp. 431–436.
[50] B.-N. Vo and B.-T. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3460–3475, Jul. 2013.
[51] B.-N. Vo, B.-T. Vo, and D. Phung, "Labeled random finite sets and the Bayes multi-target tracking filter," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6554–6567, Dec. 2014.
[52] R. Mahler, "Integral-transform derivations of exact closed-form multitarget trackers," in *Proc. 19th Int. Conf. Inf. Fusion*, 2016, pp. 950–957.
[53] Á. F. García-Fernández, Y. Xia, K. Granström, L. Svensson, and J. L. Williams, "Gaussian implementation of the multi-Bernoulli mixture filter," in *Proc. 22th Int. Conf. Inf. Fusion*, 2019, pp. 1–8.
[54] Á. F. García-Fernández, J. L. Williams, K. Granström, and L. Svensson, "Poisson multi-Bernoulli mixture filter: Direct derivation and implementation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1883–1901, Aug. 2018.
[55] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, "The labeled multi-Bernoulli filter," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3246–3260, Jun. 2014.
[56] J. L. Williams, "Marginal multi-Bernoulli filters: RFS derivation of MHT, JIPDA, and association-based MeMBer," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 3, pp. 1664–1687, Jul. 2015.
[57] T. L. Song, D. Mušicki, and Y. Kim, "Multi-target tracking with target state dependent detection," in *Proc. 15th Int. Conf. Inf. Fusion*, 2012, pp. 324–329.
[58] G. Jones, Á. F. García-Fernández, and C. Blackman, "Nonmyopic GOSPA-driven Gaussian Bernoulli sensor management," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 6, pp. 7628–7642, Dec. 2024.
[59] Á. F. García-Fernández, L. Svensson, J. L. Williams, Y. Xia, and K. Granström, "Trajectory Poisson multi-Bernoulli filters," *IEEE Trans. Signal Process.*, vol. 68, pp. 4933–4945, 2020.
[60] J. Heinonen, *Lectures on Analysis on Metric Spaces*. New York, NY, USA: Springer, 2001.
[61] J. Krejčí, O. Kost, Y. Xia, L. Svensson, and O. Straka, "Model-based multi-object visual tracking: Identification and standard model limitations," in *Proc. 28th Int. Conf. Inf. Fusion*, 2025, pp. 1–8.
[62] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler, "MOT16: A benchmark for multi-object tracking," 2016, *arXiv:1603.00831*.
[63] P. Dendorfer et al., "MOTChallenge: A benchmark for single-camera multiple target tracking," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 845–881, 2021.

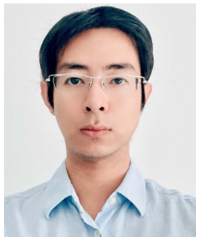
[64] J. Krejčí, O. Kost, O. Straka, and J. Duník, "Pedestrian tracking with monocular camera using unconstrained 3D motion model," in *Proc. 27th Int. Conf. Inf. Fusion*, 2024, pp. 1–8.
[65] J. Luiten et al., "HOTA: A higher order metric for evaluating multi-object tracking," *Int. J. Comput. Vis.*, vol. 129, no. 2, pp. 548–578, 2021.
[66] T. T. D. Nguyen, H. Rezatofighi, B.-N. Vo, B.-T. Vo, S. Savarese, and I. Reid, "How trustworthy are performance evaluations for basic vision tasks?," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 8538–8552, Jul. 2023.
[67] J. Krejčí, O. Kost, O. Straka, Y. Xia, L. Svensson, and Á. F. García-Fernández, "TGOSPA metric parameters selection and evaluation for visual multi-object tracking," 2024, *arXiv:2412.08321*.
[68] Á. F. García-Fernández, A. S. Rahmathullah, and L. Svensson, "A metric on the space of finite sets of trajectories for evaluation of multi-target tracking algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 3917–3928, 2020.
[69] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 3464–3468.
[70] J. Krejčí, O. Kost, and O. Straka, "Bounding box detection in visual tracking: Measurement model parameter estimation," in *Proc. 26th Int. Conf. Inf. Fusion*, 2023, pp. 1–8.
[71] J. Krejčí, O. Straka, J. Vyskočil, M. Jiřík, and U. Dahmen, "Feature-based multi-object tracking with maximally one object per class," in *Proc. 25th Int. Conf. Inf. Fusion*, 2022, pp. 1–8.
[72] A. S. Kechris, *Classical Descriptive Set Theory*. New York, NY, USA: Springer, 1995.
[73] I. Goodman, H. T. Nguyen, and R. Mahler, *Mathematics of Data Fusion*. Norwell, MA, USA: Kluwer, 1997.
[74] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*, 1st ed. Berlin, Germany: Springer, 2003.
[75] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.
[76] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

Jan Krejčí received the M.Sc. degree in cybernetics and control engineering in 2020 from the University of West Bohemia, Pilsen, Czech Republic, where he is currently working toward the Ph.D. degree in cybernetics; advancing nonlinear filtering methods.

His research interests include uncertainty propagation in continuous nonlinear systems, grid-based filtering, and especially multiple object tracking.

Oliver Kost was born in Žatec, Czechoslovakia, in 1991. He received the Ph.D. degree in cybernetics from the Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Pilsen, Czech Republic, in 2022.

He is currently with the NTIS Research Center, University of West Bohemia. His research interests include system identification, navigation systems, and global navigation satellite system spoofing detection.



Yuxuan Xia received the M.Sc. degree in communication engineering and the Ph.D. degree in signals and systems from the Chalmers University of Technology, Gothenburg, Sweden, in 2017 and 2022, respectively.

He was a Postdoctoral Researcher with the Signal Processing Group, Chalmers University of Technology, Gothenburg, Sweden, for one year, and then an Industrial Postdoctoral Researcher with Zenseact and the Division of Automatic Control, Linköping University, Linköping, Sweden. He is currently a Research Assistant Professor with the School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, China. He has organized tutorials on multiobject tracking at IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (FUSION) (2020–2025) and MFI conferences (2024–2025). His research interests include sensor fusion, multiobject tracking, and simultaneous localization and mapping, with a focus on autonomous systems.

Dr. Xia was the recipient of the paper awards at FUSION 2021 and 2024 IEEE International Conference on Multisensor Fusion and Integration.

Lennart Svensson (Senior Member, IEEE) is a Professor of Signal Processing with the Chalmers University of Technology, Gothenburg, Sweden. He has organized a massive open online course on multiple object tracking, available on edX and YouTube. His main research interests include machine learning and Bayesian inference in general, and nonlinear filtering, deep learning, and tracking in particular.

Dr. Svensson received paper awards at the International Conference on Information Fusion in 2009, 2010, 2017, and 2019.

Ondřej Straka (Member, IEEE) received the master's degree in cybernetics and control engineering and the Ph.D. degree in cybernetics from the University of West Bohemia, Pilsen, Czech Republic, in 1999 and 2004, respectively.

He is currently an Associate Professor with the Department of Cybernetics, University of West Bohemia. He has participated in many projects of fundamental research and several projects of applied research. He has authored or coauthored more than 140 journal and conference papers. His research interests include local and global nonlinear state estimation methods, system identification, performance evaluation, and fault detection.

Dr. Straka received the Werner von Siemens Excellence Award in 2014 for the most important result in basic research.