# LOCAL DETECTION AND ESTIMATION OF MULTIPLE OBJECTS FROM IMAGES WITH OVERLAPPING OBSERVATION AREAS

*Rene Repp*[⋆]*, Günther Koliander*[⋆]*, Florian Meyer*[†]*, and Franz Hlawatsch*[⋆]

[⋆]Institute of Telecommunications, TU Wien, Vienna, Austria ({rene.repp,guenther.koliander,franz.hlawatsch}@nt.tuwien.ac.at)
[†]Centre for Maritime Research and Experimentation, La Spezia, Italy (florian.meyer@cmre.nato.int)

## ABSTRACT

We propose a method for detecting and estimating multiple objects from multiple noisy images with partly overlapping observation areas. The goal is to detect the objects that are "locally" present in the individual observation areas and to estimate their states. Our method is based on a new closed-form expression of the marginal posterior probability hypothesis density (PHD) and admits a distributed implementation. Simulation results demonstrate performance gains over correlation-based and PHD-based methods that do not take advantage of the overlapping observation areas.

***Index Terms***— Random finite set, FISST, probability hypothesis density, PHD, image processing.

## 1. INTRODUCTION

Detecting and estimating objects from observed images is important in applications such as visual tracking [1], autonomous driving [2], remote sensing [3], biomedical analytics [4, 5], and SLAM [6, 7]. Methods for object detection and estimation from image-like data include [8–16]. Recently, methods based on random finite sets (RFSs) [17, 18] have been proposed for applications in target tracking and data fusion. Some of these methods involve the probability hypothesis density (PHD), which is a first-order moment of an RFS distribution [19–21]. An RFS method that detects and tracks multiple objects based on a sequence of images is proposed in [16]. This method uses a multi-Bernoulli filter and assumes a single sensor or multiple sensors with identical observation areas (OAs).

Here, we propose a PHD-based method for the problem of estimating the number and states of objects from multiple noisy images. We allow for overlapping OAs and extend the data model of [16] to this case. However, differently from [16], we consider estimation from a single set of images, rather than sequential estimation from a temporal sequence of images. Our goal, more specifically, is to estimate the numbers and states of the objects that are present in the individual OAs. This amounts to estimating the "local RFS" of the objects present in a given OA. The proposed estimator is based on a new closed-form expression of the marginal posterior PHD. It takes advantage of the overlapping OAs by using all the relevant images, and it can be easily implemented in a decentralized manner. Simulation results demonstrate the superior performance of our estimator compared to both a PHD-based estimator using only the respective image and a classical correlation-based estimator [22, Ch. 3].

This paper is organized as follows. After a review of some RFS fundamentals in Section 2, the data model and a statistical formulation are described in Section 3. The proposed estimator is developed in Section 4. Simulation results are presented in Section 5.

This work was supported by the FWF under grants P27370-N30 and J3886-N31, the National Sustainability Program of the European Commission under grant LO1401, and the NATO Supreme Allied Command Transformation under project SAC000601. F. Meyer is currently with the Laboratory for Information and Decision Systems, MIT, Cambridge, MA, USA.

## 2. RFS FUNDAMENTALS

An RFS $\mathsf{X} \subseteq \mathbb{R}^d$ is a finite set of a random number of random vectors $\mathsf{x} \in \mathbb{R}^d$, i.e., $\mathsf{X} = \{\mathsf{x}_1, \ldots, \mathsf{x}_\mathsf{k}\}$, where $\mathsf{k}$ is a nonnegative random integer and the $\mathsf{x}_i \in \mathbb{R}^d$ are random vectors. (We use sans-serif type to denote random quantities.) Note that the realization $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ is not changed by a permutation of the elements $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. Using the FISST framework [17, 18], an RFS $\mathsf{X}$ is characterized by its probability density function (PDF) $f(X)$, where $X$ is any finite set of elements from $\mathbb{R}^d$. The probability that $\mathsf{X}$ is contained in a given closed set $A \subseteq \mathbb{R}^d$ can be calculated as $\Pr(\mathsf{X} \subseteq A) = \int_A f(X)\,\mathrm{d}X$, where the right-hand side is a *set integral* [17, 18]. Similarly, an $I$-tuple of RFSs $\mathsf{X}_1, \ldots, \mathsf{X}_I$ is characterized by the joint PDF $f(X_1, \ldots, X_I)$, and the probability that the $\mathsf{X}_i$ are contained in closed sets $A_i \subseteq \mathbb{R}^d$ can be calculated by an iterated set integral, i.e.,

$$\Pr(\mathsf{X}_1 \subseteq A_1, \ldots, \mathsf{X}_I \subseteq A_I) = \int_{A_1} \cdots \int_{A_I} f(X_1, \ldots, X_I)\,\mathrm{d}X_I \ldots \mathrm{d}X_1 . \tag{1}$$

The PHD $D(\boldsymbol{x})$ of an RFS $\mathsf{X}$ is a real-valued function on $\mathbb{R}^d$ with the defining property that its integral over any closed subset $A \subseteq \mathbb{R}^d$ gives the expected number of elements of $\mathsf{X}$ contained in $A$, i.e.,

$$\mathsf{E}\{|\mathsf{X} \cap A|\} = \int_A D(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} . \tag{2}$$

Here, $|X|$ denotes the cardinality of $X$.

An important special RFS is the *Poisson RFS*, whose PDF is [17, 18]

$$f(X) = e^{-\mu}\mu^{|X|} \prod_{\boldsymbol{x} \in X} f(\boldsymbol{x}) , \tag{3}$$

where $\mu > 0$ is the mean of the random variable $|\mathsf{X}|$ and $f(\boldsymbol{x})$ is a vector PDF on $\mathbb{R}^d$ (also called *spatial PDF*). A Poisson RFS is fully characterized by its PHD, which is given by

$$D(\boldsymbol{x}) = \mu f(\boldsymbol{x}) . \tag{4}$$

## 3. DATA MODEL AND STATISTICAL FORMULATION

We assume that an unknown number of objects may be present in one or several observed images. An object is characterized by its *state* $\mathsf{x} \in R \subseteq \mathbb{R}^d$. For example, $\mathsf{x}$ may be the two-dimensional (2-D) position of an object, in which case $R \subseteq \mathbb{R}^2$. We observe a collection $\mathsf{z}_{1:I} = (\mathsf{z}_1, \ldots, \mathsf{z}_I)$ of $I$ gray-scale images, where $\mathsf{z}_i = \big(\mathsf{z}_i^{(1)} \cdots \mathsf{z}_i^{(M)}\big)^{\mathrm{T}} \in \mathbb{R}^M$, $i \in \{1, \ldots, I\}$ comprises the pixel values $\mathsf{z}_i^{(m)}$, $m \in \{1, \ldots, M\}$ of the $i$th image.

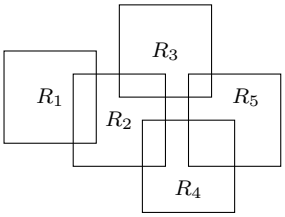



**Fig. 1**. Schematic representation of five pairwise overlapping OAs $R_1, \ldots, R_5$ associated with images $\mathbf{z}_1, \ldots, \mathbf{z}_5$.

To characterize the statistical dependence of the images on the object states, we first assume that only one object with state $\mathbf{x} = \boldsymbol{x}$ is present. Let $T_i(\boldsymbol{x}) \subseteq \{1, \ldots, M\}$ denote the set of those pixel-indices $m$ in image $\mathbf{z}_i$ that are affected by an object with state $\boldsymbol{x}$. An object affects pixel $m \in T_i(\boldsymbol{x})$ by changing the distribution of $\mathbf{z}_i^{(m)}$. More specifically, the conditional PDF of $\mathbf{z}_i^{(m)}$ given $\mathbf{x} = \boldsymbol{x}$ is [16]

$$f\big(z_i^{(m)}\big|\boldsymbol{x}\big) = \begin{cases} \phi_i\big(z_i^{(m)}, \boldsymbol{x}\big) & \text{if } m \in T_i(\boldsymbol{x}), \\ \psi_i\big(z_i^{(m)}\big) & \text{if } m \notin T_i(\boldsymbol{x}), \end{cases}$$

with known functions $\phi_i\big(z_i^{(m)}, \boldsymbol{x}\big)$ and $\psi_i\big(z_i^{(m)}\big)$ whose integral with respect to $z_i^{(m)}$ is 1. We define the $i$th OA $R_i \subseteq R$ as the set of all state vectors $\boldsymbol{x} \in R$ that affect at least one pixel in image $\mathbf{z}_i$, i.e.,

$$R_i \triangleq \{\boldsymbol{x} \in R : T_i(\boldsymbol{x}) \neq \emptyset\}, \quad i \in \{1, \ldots, I\}.$$

Evidently, $\bigcup_{i=1}^I R_i \subseteq R$; for simplicity, we assume $\bigcup_{i=1}^I R_i = R$, i.e., the object affects at least one image $\mathbf{z}_i$. The pairwise intersections $R_{ij} \triangleq R_i \cap R_j = R_{ji}$, $i \neq j$ are allowed to be nonempty, which means that the OAs of two different images $\mathbf{z}_i$ and $\mathbf{z}_j$ may overlap. That is, if $R_{ij} \neq \emptyset$, an object with state $\boldsymbol{x} \in R_{ij}$ affects both images $\mathbf{z}_i$ and $\mathbf{z}_j$. However, we restrict to pairwise overlap, i.e., an object can affect at most two images simultaneously. This means that $R_i \cap R_j \cap R_k = \emptyset$ for any choice of different indices $i$, $j$, and $k$. Fig. 1 shows an example of pairwise overlapping OAs.

We now generalize to the case of multiple objects. While the definitions of $T_i(\boldsymbol{x})$ and $R_i$ are still valid, we have to adapt the statistical dependence. We assume that the number and states of the objects are random and that the objects do not possess an inherent order. Accordingly, we model the entirety of all object states as an RFS $\mathsf{X} \subseteq R$ with a given prior PDF $f(X)$. Differently form [16], we do not require $T_i(\boldsymbol{x})$ and $T_i(\boldsymbol{x}')$ for different $\boldsymbol{x}, \boldsymbol{x}' \in X$ to be disjoint; hence a pixel may be affected by more than one object. We also define a "local" object state RFS for each image $\mathbf{z}_i$ as

$$\mathsf{X}_i \triangleq \mathsf{X} \cap R_i \subseteq R_i, \quad i \in \{1, \ldots, I\}. \tag{5}$$

That is, $\mathsf{X}_i$ contains only those elements of $\mathsf{X}$ that lie in $R_i$, i.e., those object states $\boldsymbol{x} \in \mathsf{X}$ affecting image $\mathbf{z}_i$; note that $\bigcup_{i=1}^I \mathsf{X}_i = \mathsf{X}$. Since the OAs $R_i$ are allowed to overlap, certain states $\boldsymbol{x}$ may be contained in two local RFSs simultaneously. More precisely, $\boldsymbol{x} \in \mathsf{X}_i$ and $\boldsymbol{x} \in \mathsf{X}_j$ if and only if $\boldsymbol{x} \in \mathsf{X} \cap R_{ij}$.

The statistical dependence of the $i$th image $\mathbf{z}_i$ on the state RFS $\mathsf{X}$ is described by the *local likelihood function* $f(\boldsymbol{z}_i|X)$. Following [16], we assume that $f(\boldsymbol{z}_i|X)$ factors as

$$f(\boldsymbol{z}_i|X) = \alpha_i(\boldsymbol{z}_i) \prod_{\boldsymbol{x} \in X} \beta_i(\boldsymbol{x}, \boldsymbol{z}_i), \tag{6}$$

with

$$\alpha_i(\boldsymbol{z}_i) = \prod_{m=1}^M \psi_i\big(z_i^{(m)}\big), \tag{7}$$

$$\beta_i(\boldsymbol{x}, \boldsymbol{z}_i) = \prod_{m \in T_i(\boldsymbol{x})} \frac{\phi_i\big(z_i^{(m)}, \boldsymbol{x}\big)}{\psi_i\big(z_i^{(m)}\big)}. \tag{8}$$

By construction, $\mathsf{X}_i$ contains all state vectors $\boldsymbol{x}$ that affect image $\mathbf{z}_i$, i.e.,[1]

$$f(\boldsymbol{z}_i|X) = f(\boldsymbol{z}_i|X_i). \tag{9}$$

Therefore, given $\mathsf{X}_i$, $\mathbf{z}_i$ is conditionally independent of all the other local RFSs $\mathsf{X}_j$, and hence $f(\boldsymbol{z}_i|X_{1:I}) = f(\boldsymbol{z}_i|X_i)$. Furthermore, we assume that given $\mathsf{X}_{1:I}$, $\mathbf{z}_i$ is conditionally independent of all the other $\mathbf{z}_j$. Thus, the *global likelihood function* $f(\boldsymbol{z}_{1:I}|X_{1:I})$ factors as

$$f(\boldsymbol{z}_{1:I}|X_{1:I}) = \prod_{i=1}^I f(\boldsymbol{z}_i|X_{1:I}) = \prod_{i=1}^I f(\boldsymbol{z}_i|X_i).$$

Using Bayes' rule, we then obtain the *joint posterior PDF* (up to a normalization factor) as

$$f(X_{1:I}|\boldsymbol{z}_{1:I}) \propto f(X_{1:I}) f(\boldsymbol{z}_{1:I}|X_{1:I}) = f(X_{1:I}) \prod_{i=1}^I f(\boldsymbol{z}_i|X_i). \tag{10}$$

Here, $f(X_{1:I})$ is the *joint* prior PDF, which has to be determined from the given prior PDF $f(X)$, and $f(\boldsymbol{z}_i|X_i) = f(\boldsymbol{z}_i|X)$ (cf. (9)) is given by (6).

## 4. DETECTION AND ESTIMATION

### 4.1. The RFS Estimator

The problem considered in this paper is to jointly detect the number of objects affecting image $\mathbf{z}_i$ and estimate the states of these objects, for each $i \in \{1, \ldots, I\}$. This problem amounts to *estimating each local RFS* $\mathsf{X}_i$. Because an object can appear in two images, we consider estimation of each $\mathsf{X}_i$ from all the observed images $\mathbf{z}_j = \boldsymbol{z}_j$, $j \in \{1, \ldots, I\}$, or equivalently from $\boldsymbol{z}_{1:I}$, rather than from $\boldsymbol{z}_i$ alone.

Our estimator of $\mathsf{X}_i$ is based on the *marginal posterior PHD* $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$, which can be calculated from the *marginal posterior PDF* $f(X_i|\boldsymbol{z}_{1:I})$ (assumed continuous) according to [17, Sec. 16.2]

$$D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I}) = \int_{R_i} \Delta_{X_i}(\boldsymbol{x}) f(X_i|\boldsymbol{z}_{1:I}) \mathrm{d}X_i. \tag{11}$$

Here, $\Delta_{X_i}(\boldsymbol{x}) \triangleq \sum_{\boldsymbol{x}' \in X_i} \delta(\boldsymbol{x} - \boldsymbol{x}')$, where $\delta(\boldsymbol{x})$ is the Dirac delta function. The marginal posterior PDF $f(X_i|\boldsymbol{z}_{1:I})$ is obtained from the joint posterior PDF $f(X_{1:I}|\boldsymbol{z}_{1:I})$ in (10) by integrating out all $X_j$, $j \neq i$, i.e.,

$$f(X_i|\boldsymbol{z}_{1:I}) = \int_{R_{\sim i}} f(X_{1:I}|\boldsymbol{z}_{1:I}) \mathrm{d}X_{\sim i}, \tag{12}$$

where the right-hand side is the iterated set integral over all $R_j$ with $j \neq i$ (cf. (1)). Once the marginal posterior PHD $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$ has been determined, $\mathsf{X}_i$ can be estimated by the following procedure [17, pp. 504–505]. First, the expected number of objects in $R_i$ given the image observations $\mathbf{z}_{1:I} = \boldsymbol{z}_{1:I}$ is calculated as $\mathsf{E}\{|\mathsf{X}_i||\boldsymbol{z}_{1:I}\} = \int_{R_i} D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I}) \mathrm{d}\boldsymbol{x}$ (cf. (2)). An estimate $\hat{K}_i$ of the number of objects in $R_i$ is then obtained by rounding $\mathsf{E}\{|\mathsf{X}_i||\boldsymbol{z}_{1:I}\}$. Next, the positions $\hat{\boldsymbol{x}}_{i,k}$, $k \in \{1, \ldots, \hat{K}_i\}$ of the $\hat{K}_i$ largest local maxima of $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$ are determined. Finally, an estimate of $\mathsf{X}_i$ is given by the set of all $\hat{\boldsymbol{x}}_{i,k}$, i.e., $\hat{X}_i = \{\hat{\boldsymbol{x}}_{i,k}\}_{k=1}^{\hat{K}_i}$.

[1] Indeed, for $\boldsymbol{x} \in X \backslash R_i$, we have $T_i(\boldsymbol{x}) = \emptyset$ and thus (8) yields $\beta_i(\boldsymbol{x}, \boldsymbol{z}_i) = 1$. Hence, from (6), $f(\boldsymbol{z}_i|X) = f(\boldsymbol{z}_i|X \cap R_i)$, which equals $f(\boldsymbol{z}_i|X_i)$ due to (5).

### 4.2. Calculation of the PHD

It remains to calculate the marginal posterior PHD $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$. In principle, this consists of the following steps. *Step 1*: Determine the joint prior PDF $f(X_{1:I})$ from the given prior PDF $f(X)$. *Step 2*: Obtain the marginal posterior PDF $f(X_i|\boldsymbol{z}_{1:I})$ by evaluating (10) and (12). *Step 3*: Obtain $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$ by evaluating (11). In what follows, we will present a general closed-form expression of $f(X_{1:I})$ (Step 1). We will also present closed-form expressions of $f(X_i|\boldsymbol{z}_{1:I})$ (Step 2) and $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$ (Step 3) under the assumption that the prior PDF $f(X)$ is Poisson. Hereafter, we denote by $E$ the set of index pairs $(i,j) \in \{1,\ldots,I\} \times \{1,\ldots,I\}$ such that $R_{ij} \neq \emptyset$ and $i > j$. That is, $E$ consists of the index pairs of all intersecting OAs $R_i$ and $R_j$, where the constraint $i > j$ ensures that the equivalent index pairs $(i,j)$ and $(j,i)$ are counted only once. We also define the "neighbor set" $N_i \triangleq \{j \neq i : R_{ij} \neq \emptyset\}$ as the set of indices $j$ of all $R_j$ intersecting $R_i$. Note that $j \in N_i$ if and only if $i \in N_j$. Finally, for two finite sets $X, Y \subseteq R$, we define $\delta_Y(X)$ via the sifting property $\int_R g(X)\,\delta_Y(X)\,\mathrm{d}X = g(Y)$, for any continuous function $g(X) \in \mathbb{R}$.

The calculation of $f(X_{1:I})$ (Step 1) is based on the following result, whose proof is omitted because of space restrictions and will be provided in a future journal publication.

**Theorem 1** *Let $f(X)$ be the prior PDF of $\mathsf{X}$. Then the joint prior PDF $f(X_{1:I}) = f(X_1,\ldots,X_I)$ of the local RFSs $\mathsf{X}_i = \mathsf{X} \cap R_i$ is given for $X_i \subseteq R_i$ by*

$$f(X_{1:I}) = f\left(\bigcup_{i'=1}^{I} X_{i'}\right) \prod_{(i,j)\in E} \delta_{X_i \cap R_j}(X_j \cap R_i)\,.$$

Here, the factor $\prod_{(i,j)\in E} \delta_{X_i\cap R_j}(X_j \cap R_i)$ ensures that all sets $X_i$ and $X_j$ with $(i,j) \in E$ contain the same elements in the overlap area $R_{ij}$. In other words, all events containing sets that do not satisfy this condition are assigned zero probability under integration.

Hereafter, we assume that the prior PDF $f(X)$ is Poisson (see (3)). The next result, whose proof uses Theorem 1, states that the marginal posterior PDFs $f(X_i|\boldsymbol{z}_{1:I})$ (Step 2) are then also Poisson.

**Theorem 2** *Let $f(X)$ be Poisson with mean $\mu > 0$ and spatial PDF $f(\boldsymbol{x})$. Then the marginal posterior PDF $f(X_i|\boldsymbol{z}_{1:I})$ of $\mathsf{X}_i = \mathsf{X} \cap R_i$ given $\mathbf{z}_{1:I} = \boldsymbol{z}_{1:I}$ is again Poisson, i.e., for $X_i \subseteq R_i$*

$$f(X_i|\boldsymbol{z}_{1:I}) = e^{-\mu_i}\mu_i^{|X_i|} \prod_{\boldsymbol{x}\in X_i} f_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})\,,$$

*with mean $\mu_i = \mu\varepsilon_i$ and spatial PDF*

$$f_i(\boldsymbol{x}|\boldsymbol{z}_{1:I}) = \frac{1}{\varepsilon_i} f(\boldsymbol{x})\,\beta_i(\boldsymbol{x},\boldsymbol{z}_i) \prod_{j\in N_i} \gamma_{ij}(\boldsymbol{x},\boldsymbol{z}_j)\,.$$

*Here,*

$$\varepsilon_i = \int_{R_i} f(\boldsymbol{x})\,\beta_i(\boldsymbol{x},\boldsymbol{z}_i) \prod_{j\in N_i} \gamma_{ij}(\boldsymbol{x},\boldsymbol{z}_j)\,\mathrm{d}\boldsymbol{x}\,,$$

*the function $\beta_i(\boldsymbol{x},\boldsymbol{z}_i)$ was defined in* (8)*, and $\gamma_{ij}(\boldsymbol{x},\boldsymbol{z}_j)$ is equal to $\beta_j(\boldsymbol{x},\boldsymbol{z}_j)$ if $\boldsymbol{x} \in R_{ij}$ and* 1 *otherwise.*

Since the marginal posterior PDF $f(X_i|\boldsymbol{z}_{1:I})$ is Poisson with mean $\mu_i$ and spatial PDF $f_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$, the marginal posterior PHD $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$ (Step 3) is given by (4) with obvious modifications, i.e.,

$$D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I}) = \mu_i\, f_i(\boldsymbol{x}|\boldsymbol{z}_{1:I}) = \mu f(\boldsymbol{x})\,\beta_i(\boldsymbol{x},\boldsymbol{z}_i) \prod_{j\in N_i} \gamma_{ij}(\boldsymbol{x},\boldsymbol{z}_j)\,. \tag{13}$$

### 4.3. Distributed Implementation

A distributed implementation of the proposed estimator can be obtained in a straightforward manner. Consider a decentralized sensor network where sensor $i$ observes image $\mathbf{z}_i$. Then, the calculation of $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$ in (13) and the subsequent processing as described in Section 4.1 can be performed locally at sensor $i$ provided that those parts of the "neighboring" images $\boldsymbol{z}_j$, $j \in N_i$ that are affected by the objects in $R_i$ are available at sensor $i$. This means that the tuples $\big(m, z_j^{(m)}\big)_{m\in M_{ji}}$ with $M_{ji} \triangleq \bigcup_{\boldsymbol{x}\in R_{ji}} T_j(\boldsymbol{x})$ have to be transmitted from the "neighbor sensors" $j \in N_i$ to sensor $i$. We conclude that each sensor $i$ has to transmit to its neighbor sensor $j \in N_i$ $|M_{ij}|$ integers $m$ and real numbers $z_i^{(m)}$.

## 5. NUMERICAL STUDY

We consider three gray-scale images $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{z}_3$, each of size $100 \times 100$ pixels, where each pixel covers a unit square in $\mathbb{R}^2$. The centers of the images are located at $\boldsymbol{p}_1 = (40\ \ -10)^{\mathrm{T}}$, $\boldsymbol{p}_2 = (50\ \ 50)^{\mathrm{T}}$, and $\boldsymbol{p}_3 = (100\ \ 100)^{\mathrm{T}}$. The object states are 2-D position vectors, i.e., $\mathbf{x} = (\mathsf{x}_1\ \mathsf{x}_2)^{\mathrm{T}} \in R \subseteq \mathbb{R}^2$. The pixel index set $T_i(\boldsymbol{x})$, $i \in \{1,2,3\}$ corresponds to a $5\times 5$ array of pixels whose center is the pixel point closest to $\boldsymbol{x}$. Thus, the OAs $R_i$, $i \in \{1,2,3\}$ are overlapping squares in $\mathbb{R}^2$ (with nonempty intersections $R_{12}$ and $R_{23}$) and $R$ is their union. The global RFS $\mathsf{X}$ is assumed Poisson with $\mu = 4.696$ (corresponding to $\mathsf{X}_2 = \mathsf{X} \cap R_2$ being a Poisson RFS with mean 2) and $f(\boldsymbol{x})$ uniform on $R$. For each simulation run, a different global state set $X$ was randomly drawn from this Poisson distribution.

Given a set of object positions $\mathsf{X} = X$, the pixel values $\mathsf{z}_i^{(m)}$ are randomly generated as

$$\mathsf{z}_i^{(m)} = \sum_{\boldsymbol{x}\in R_i^{(m)}\cap X} h(m;\boldsymbol{x}-\tilde{\boldsymbol{p}}_i) + \mathsf{n}_i^{(m)}. \tag{14}$$

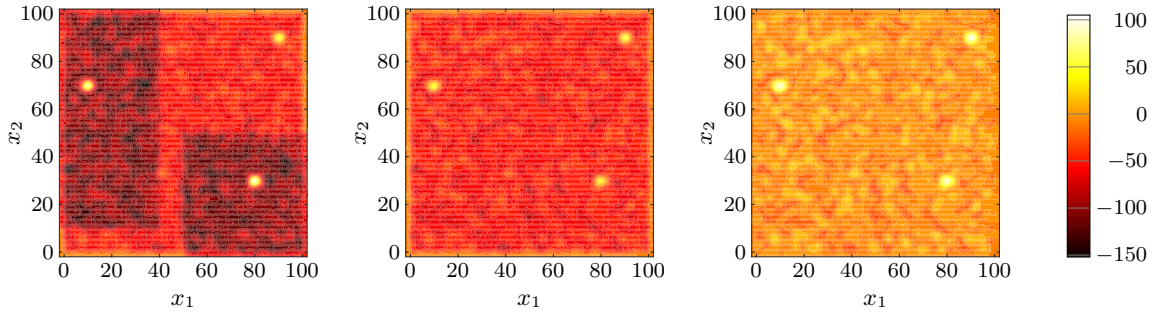
Here, $R_i^{(m)} \triangleq \{\boldsymbol{x} \in R_i : m \in T_i(\boldsymbol{x})\}$ is the set of all object positions $\boldsymbol{x}$ that affect pixel $\mathsf{z}_i^{(m)}$ (note that $\bigcup_m R_i^{(m)} = R_i$). Furthermore, $h(m;\boldsymbol{x})$ is the point-spread function used in [16],

$$h(m;\boldsymbol{x}) = \frac{I_{\mathrm{s}}}{2\pi\sigma_h^2} \exp\left(-\frac{(a-x_1)^2+(b-x_2)^2}{2\sigma_h^2}\right),$$

with source intensity $I_{\mathrm{s}}$, blurring factor $\sigma_h^2 = 2$, and $a$ and $b$ defined by $m = 100\,(a-1) + b$, for $a, b \in \{1,\ldots,100\}$. Finally, $\tilde{\boldsymbol{p}}_i \triangleq \boldsymbol{p}_i - (50\ \ 50)^{\mathrm{T}}$, and $\mathsf{n}_i^{(m)}$ is independent and identically distributed zero-mean Gaussian noise with variance $\sigma^2 = 1$. According to (14), each object affecting pixel $\mathsf{z}_i^{(m)}$, i.e., with position $\boldsymbol{x} \in R_i^{(m)}$, adds a deterministic value[2] $h(m;\boldsymbol{x}-\tilde{\boldsymbol{p}}_i)$ to the random noise value $\mathsf{n}_i^{(m)}$.

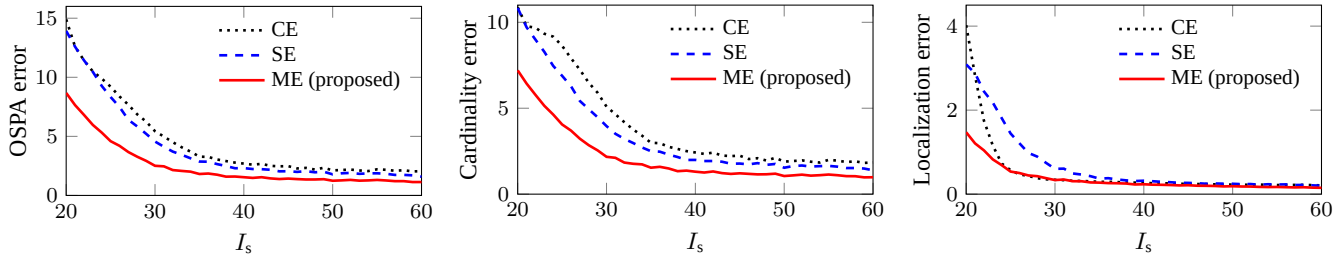
We compare the performance of the proposed estimator, hereafter termed "multiple-image PHD estimator" or briefly ME, with that of two reference methods, namely, a "single-image PHD estimator" (SE) and a single-image correlation-based estimator (CE) [22, Ch. 3]. These estimators apply the same processing—described below—to an estimator-dependent "spatial function." For the ME, the spatial function is the logarithm of the marginal posterior PHD $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$ in (13). For the SE, it is the logarithm of $D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})$ in

[2] Note that the deterministic values due to several (closely spaced) objects may add up in our simulation even though this is not taken into account in our model for the local likelihood function. In fact, we approximate the true local likelihood function $f(\boldsymbol{z}_i|X_i) = f(\boldsymbol{z}_i|X)$ by the expression (6)–(8) with $\phi_i\big(z_i^{(m)},\boldsymbol{x}\big) = \mathcal{N}\big(h(m;\boldsymbol{x}-\tilde{\boldsymbol{p}}_i),\sigma^2\big)$ and $\psi_i\big(z_i^{(m)}\big) = \mathcal{N}(0,\sigma^2)$, where $\mathcal{N}(\mu,\sigma^2)$ denotes a Gaussian PDF with mean $\mu$ and variance $\sigma^2$.




(a) (b) (c)

**Fig. 2**. Example of spatial functions of (a) ME (proposed), (b) SE, and (c) CE, for source intensity $I_\mathrm{s} = 50$.




(a) (b) (c)

**Fig. 3**. Performance of the proposed multi-image PHD estimator (ME) and of the two reference methods (SE and CE) versus the source intensity $I_\mathrm{s}$: (a) mean OSPA error, (b) mean cardinality error, and (c) mean localization error.

(13) with $N_i = \emptyset$, i.e., estimation of $\mathsf{X}_i$ is based solely on the respective image $\boldsymbol{z}_i$. For the CE, it is the correlation function $c(\boldsymbol{x}, \boldsymbol{z}_i) \triangleq \sum_{m \in T_i(\boldsymbol{x})} z_i^{(m)} h(m; \boldsymbol{x} - \tilde{\boldsymbol{p}}_i)$, which is again based solely on $\boldsymbol{z}_i$. The processing applied to the spatial function consists of the following steps. The spatial function is evaluated on the pixel grid. The positions of the 20 largest function values are used to initialize 20 instances of a gradient ascent algorithm, which yield 20 local maxima of the spatial function. We then discard all maxima below a threshold level $t$ to obtain a reduced list of candidate positions. Starting with the first candidate position, we cluster together all candidates inside a circle with radius $r$ and delete them from our candidate list. This procedure is repeated until the list is empty. The position of the largest maximum within each cluster is chosen as an estimate $\hat{\boldsymbol{x}}$ of an object position $\mathbf{x}$. Finally, the set of all the obtained object position estimates $\hat{\boldsymbol{x}}$ is used as the estimate of the local RFS $\mathsf{X}_i$. We note that this procedure differs from that described in Section 4.1 in that it avoids the numerically difficult computation of the integral $\int_{R_i} D_i(\boldsymbol{x}|\boldsymbol{z}_{1:I})\,\mathrm{d}\boldsymbol{x}$. The threshold $t$, gradient ascent step size, and cluster radius $r$ were numerically optimized for each method.

Fig. 2 shows an example of the spatial functions of the three estimators for local RFS $\mathsf{X}_2$ at source intensity $I_\mathrm{s} = 50$. There are three objects present, which are visible as bright spots (corresponding to large local maxima). The SE and CE functions exhibit similar patterns. In the ME function, one can observe an enhancement of the object peaks and a reduction of the noise in the two overlap areas.

To assess the performance of the three estimators, we use the first-order mean OSPA distance [23] between the true local RFS $\mathsf{X}_2$ and the corresponding estimate $\hat{\mathsf{X}}_2$. The base metric of the OSPA distance is chosen as the Euclidean distance, and the cutoff parameter is set to 30. In addition, we consider the cardinality and localization error components (the first-order OSPA distance is the sum of these components [23]), which measure the average deviation between the true and estimated object number and object positions, respectively. Fig. 3 shows the mean OSPA, cardinality, and localization errors obtained with the three estimators for various source intensities $I_\mathrm{s}$. These mean errors were estimated by averaging over 10000 simulation runs per intensity value. According to the OSPA curves, the proposed ME outperforms the two reference methods, with the largest performance gains observed for small intensity values, i.e., in the low-SNR regime. The superior performance of ME demonstrates the advantages of leveraging the overlap of the OAs by basing the estimation of each local RFS on all the relevant images. The error floors visible in Fig. 3 are probably due to clustering errors and, in the case of ME and SE, our approximation of the true local likelihood function by (6). Both the overall OSPA error and the cardinality error of ME are lower than the respective errors of SE, which in turn are slightly lower than those of CE. The localization error of ME is lower than that of SE and CE for small $I_\mathrm{s}$ and similar otherwise. It is also seen that for $I_\mathrm{s}$ between about 22 and 37, somewhat surprisingly, CE achieves better localization results than SE.

## 6. CONCLUSION

We proposed a PHD-based method for the problem of locally estimating the number and states of objects from multiple noisy images whose observation areas (OAs) are allowed to overlap. Our estimator is based on a closed-form expression of the marginal posterior PHD and takes advantage of the overlap of the OAs. A distributed implementation with moderate communication cost can be easily obtained. Simulation results demonstrated that our estimator outperforms single-image PHD-based and correlation-based estimators that do not exploit the overlapping OAs.

## 7. REFERENCES


[1] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1442–1468, Jul. 2014.

[2] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun, "Towards fully autonomous driving: Systems and algorithms," in *Proc. IEEE IV-11*, Baden-Baden, Germany, Jun. 2011, pp. 163–168.

[3] T. Lillesand, R. W. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation*, Wiley, New York, NY, USA, 2014.

[4] R. J. Adrian, "Particle-imaging techniques for experimental fluid mechanics," *Annu. Rev. Fluid Mech.*, vol. 23, no. 1, pp. 261–304, 1991.

[5] A. Genovesio, T. Liedl, V. Emiliani, W. J. Parak, M. Coppey-Moisan, and J. C. Olivo-Marin, "Multiple particle tracking in 3-D+t microscopy: Method and application to the tracking of endocytosed quantum dots," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1062–1070, May 2006.

[6] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–110, Jun. 2006.

[7] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part II," *IEEE Robot. Autom. Mag.*, vol. 13, no. 3, pp. 108–117, Sep. 2006.

[8] Y. Barniv, "Dynamic programming solution for detecting dim moving targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 21, no. 1, pp. 144–156, Jan. 1985.

[9] M. G. Rutten, N. J. Gordon, and S. Maskell, "Recursive track-before-detect with target amplitude fluctuations," *IEE Proc. Radar Sonar Navig.*, vol. 152, no. 5, pp. 345–352, Oct. 2005.

[10] S. J. Davey, M. Wieneke, and H. Vu, "Histogram-PMHT unfettered," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 3, pp. 435–447, Jun. 2013.

[11] B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina, "A tutorial on Bernoulli filters: Theory, implementation and applications," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3406–3430, Jul. 2013.

[12] W. Yi, M. R. Morelande, L. Kong, and J. Yang, "An efficient multi-frame track-before-detect algorithm for multi-target tracking," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 3, pp. 421–434, Jun. 2013.

[13] M. J. Walsh, M. L. Graham, R. L. Streit, T. E. Luginbuhl, and L. A. Mathews, "Tracking on intensity-modulated sensor data streams," in *Proc. IEEE AC-01*, Big Sky, MT, USA, Mar. 2001, vol. 4, pp. 1901–1909.

[14] J. Lund and M. Rudemo, "Models for point processes observed with noise," *Biometrika*, vol. 87, no. 2, pp. 235–249, Jun. 2000.

[15] Y. Bar-Shalom and X.-R. Li, *Multitarget-Multisensor Tracking*, Yaakov Bar-Shalom, Storrs, CT, USA, 1995.

[16] B.-N. Vo, B.-T. Vo, N. T. Pham, and D. Suter, "Joint detection and estimation of multiple objects from image observations," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5129–5141, Oct. 2010.

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

[18] R. Mahler, *Advances in Statistical Multisource-Multitarget Information Fusion*, Artech House, Norwood, MA, USA, 2014.

[19] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

[20] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multitarget filtering with random finite sets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1224–1245, Oct. 2005.

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

[22] R. Brunelli, *Template Matching Techniques in Computer Vision*, Wiley, New York, NY, USA, 2009.

[23] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multiobject filters," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.