Joint Detection and Estimation of Multiple Objects From Image Observations

Ba-Ngu Vo, Ba-Tuong Vo, Nam-Trung Pham, and David Suter

***Abstract*—The problem of jointly detecting multiple objects and estimating their states from image observations is formulated in a Bayesian framework by modeling the collection of states as a random finite set. Analytic characterizations of the posterior distribution of this random finite set are derived for various prior distributions under the assumption that the regions of the observation influenced by individual objects do not overlap. These results provide tractable means to jointly estimate the number of states and their values from image observations. As an application, we develop a multi-object filter suitable for image observations with low signal-to-noise ratio (SNR). A particle implementation of the multi-object filter is proposed and demonstrated via simulations.**

***Index Terms*—Random sets, Multi-Bernoulli, probability hypothesis density (PHD), filtering, images, tracking, track before detect (TBD).**

I. Introduction

THIS paper investigates the problem of jointly estimating the number of objects and their states from image observations. In radar/sonar applications, the objective is to detect and locate the targets from radar/sonar images [1], [11]. In spatial statistics applications, for example, agriculture and forestry, it is of interest to study the underlying spatial distribution of points (plants) from partial observations such as aerial images [8], [9], [14]. The copious amount of available image data renders this so-called multi-object estimation problem an important part of estimation theory and practice.

In many applications involving image data, the estimation is often performed on data that has been preprocessed into point measurements. For example, in forestry applications, aerial image observations are converted to spatial point patterns [8], and in radar tracking, radar images are converted to detections [1], [2], [11]. Compressing the information on the image into a finite set of points is efficient in terms of memory and/or bandwidth as well as computational requirements, and when combined with point measurement based approaches to multi-target tracking [1], [2], [7], [10], [11], [13], [23]–[25] can be very effective for a wide range of applications. However, this approach may be undesirable for applications with low signal-to-noise ratio (SNR), since the information loss incurred in the compression can be significant, and in such cases it can clearly be advantageous to make use of all information contained in the image(s).

We formulate the multi-object estimation problem in a Bayesian framework by modeling the collection of states as a finite set and computing the posterior distribution of the random finite set of states given the image observation. Computing the posterior of a random finite set is intractable in general, even for the special case of superpositional measurement models, where the image observation is the sum of the observations generated by individual states and noise [12]. Hence, drastic but principled approximations are needed. Under the assumption that the regions of the image influenced by individual states do not overlap, we derive closed-form expressions of the posterior for certain classes of priors. These results are applicable to nonsuperpositional measurement models. As an application, we develop a multi-object filter suitable for applications involving image observations with low SNR such as track-before-detect (TBD) (see, e.g., [17], [18], and [20]). We also present a tracking example as proof-of-concept for the proposed approach.

Preliminary results have been announced in the conference paper [26]. This paper presents a more complete analytical and numerical study. In Section II, we cast the problem in a Bayesian framework and detail the observation model considered in this paper. We present our main results for multi-object Bayesian estimation with image observation in Section III. The multi-object filter and its particle implementation is developed in Section IV, while numerical examples are presented in Section V. Concluding remarks and extensions are discussed in Section III.

II. Multi-Object Estimation From Images

Let $x_1, \ldots, x_n \in \mathcal{X} \subseteq \mathbb{R}^d$ denote the state (or parameter) vectors, and let $y = [y_1, \ldots y_m]$ denote the image observation comprising an array of m pixel (or bin) values. The value y_i of the ith pixel can be a real number or a vector depending on the application. For example, in a grayscale image each pixel value is a real number, whereas in a color image, each pixel value is a 3-dimensional vector representing the intensities of the three color channels. Given an image observation y, we consider the problem of jointly estimating the number of states and their values.

Manuscript received December 21, 2009; accepted April 29, 2010. Date of publication May 18, 2010; date of current version August 11, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jerome Idier. This work was supported by the Australian Research Council's Discovery Projects funding schemes and DP0880553 (B. N. Vo, D. Suter) and DP0989007 (B. T. Vo). The work of B. T. Vo was supported by an Australian Research Council Post Doctoral Fellowship.

B.-N. Vo and B.-T. Vo are with the School of Electrical, Electronic and Computer Engineering, The University of Western Australia, Crawley, WA 6009, Australia (e-mail: ba-ngu.vo@uwa.edu.au; ba-tuong.vo@uwa.edu.au).

N.-T. Pham is with the Institute for Infocomm Research, Connexis South Tower, Singapore 138632 (e-mail: ntpham@i2r.a-star.edu.sg).

D. Suter is with the School of Computer Science, University of Adelaide, North Terrace SA 5005, Australia (e-mail: dsuter@cs.adelaide.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2050482

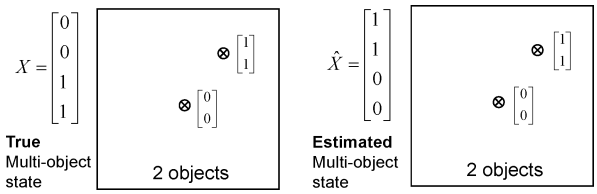

Fig. 1. Hypothetical scenario showing a fundamental inconsistency with vector representations of multi-object states. Individual states are xy positions. The estimate is correct but the error is nonzero.

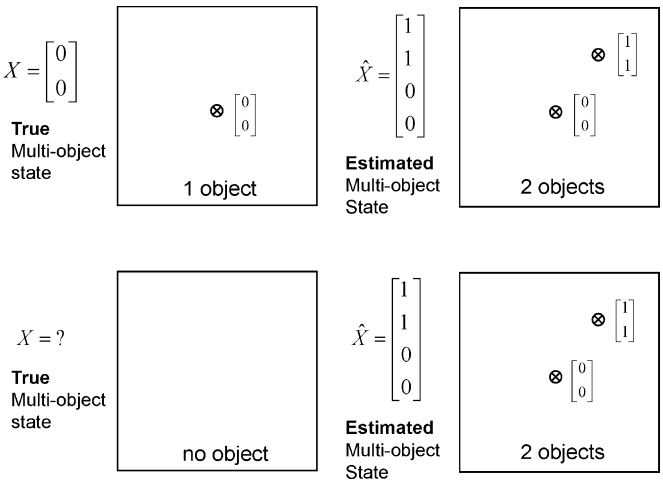

Fig. 2. Hypothetical scenarios showing a fundamental inconsistency with vector representations of multi-object states. Individual states are xy positions. How should the error be assigned when the estimated number of objects is incorrect?

We start by formulating a suitable representation of the multi-object state and cast the estimation problem in a Bayesian framework in Section II-A. The observation model considered in this paper is then described in Section II-B, setting the scene for the main results in Section IV.

A. Multi-Object Bayesian Inferencing

In the context of jointly estimating the number of states and their values, the collection of states, referred to as the *multi-object state*, is naturally represented as a finite set. The rationale behind this representation traces back to a fundamental consideration in estimation theory: estimation error. Without a meaningful notion of estimation error, the output of an estimator has little meaning. Simply stacking individual states into a single vector does not admit a satisfactory notion of error as illustrated in Figs. 1 and 2, with the ground truth represented by the vector X and the estimate represented by the vector $\hat{X}$. Intuitively, for the scenario in Fig. 1, the estimate is correct but the estimation error is $\|X - \hat{X}\| = 2$. While this fundamental inconsistency can be remedied by taking the minimum of the distance over all permutations of the states, i.e., $\min_{\mathrm{perm}(X)} \|X - \hat{X}\| = 0$, there is a more serious problem. What is the error when the estimated and true number of states are different, e.g., the scenarios in Fig. 2? A finite set representation of the multi-object state, $X = \{x_1, \ldots, x_n\}$, admits a mathematically consistent notion of estimation error since distance between sets is a well understood concept.

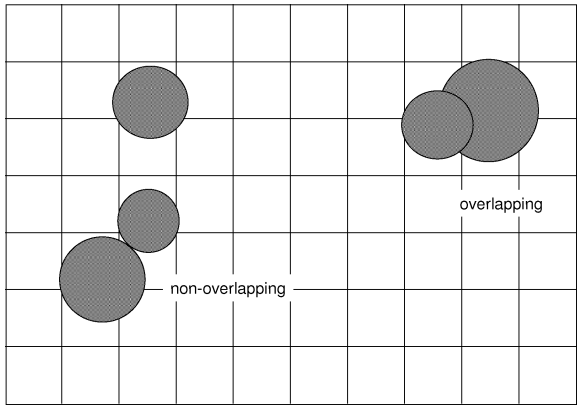

Fig. 3. An illustration of overlapping and nonoverlapping objects.

In the Bayesian estimation paradigm, the state and measurement are treated as realizations of random variables. Since the (multi-object) state X is a finite set, the concept of a random finite set (RFS) is required to cast the multi-object estimation problem in the Bayesian framework. The space of finite subsets of $\mathcal{X}$ does not inherit the usual Euclidean notion of integration and density. Hence, standard tools for random vectors are not appropriate for RFSs. Mahler's finite set statistics (FISST) provides powerful yet practical mathematical tools for dealing with RFSs [10], [11], based on a notion of integration and density that is consistent with point process theory [22]. FISST has attracted substantial interest from academia as well as the commercial sector with the developments of the probability hypothesis density (PHD) and cardinalized PHD filters [10], [22], [23], [13], [24].

Using the FISST notion of integration and density, the posterior probability density $\pi(\cdot \,|\, y)$ of the multi-object state can be computed from the prior π using Bayes rule

$$\pi(X \,|\, y) = \frac{g(y \,|\, X)\pi(X)}{\int g(y \,|\, X)\pi(X)\delta X} \tag{1}$$

where $g(y \,|\, X)$ is the probability density of the observation y given the multi-object state X (the specifics of this density is given in the next subsection), and

$$\int f(X)\delta X = \sum_{i=0}^{\infty} \frac{1}{i!} \int f(\{x_1, \ldots, x_i\}) dx_1 \ldots dx_i$$

is the set integral of a function f taking $\mathcal{F}(\mathcal{X})$, the space of finite subsets of $\mathcal{X}$, to the real line.

B. Multi-Object Likelihood Function

The type of image observation considered in this work is illustrated in Fig. 3. Objects are assumed to be rigid bodies that cannot overlap with each other. In ground target tracking, for example, the objects would be vehicles or stationary objects that must be physically separated.

An object with state x illuminates a set of pixels denoted by $T(x)$, for example, $T(x)$ could be the set of pixels whose centers fall within a certain distance from the position of the object. A pixel $i \in T(x)$, i.e., illuminated by an object with state x, has value distributed according to $\varphi_i(\cdot, x)$, while a pixel $i \notin$

$T(x)$, i.e., not illuminated by any object, has value distributed according to $\phi_i(\cdot)$. More concisely, the probability density of the value y_i of pixel i, given a state x is

$$p(y_i \,|\, x) = \begin{cases} \varphi_i(y_i, x) & i \in T(x), \\ \phi_i(y_i) & i \notin T(x). \end{cases} \tag{2}$$

For example, in TBD (see [17, Ch. 11], [4])

$$\phi_i(y_i) = \mathcal{N}(y_i; 0, \sigma^2)$$
$$\varphi_i(y_i, x) = \mathcal{N}(y_i; h_i(x), \sigma^2)$$

where $\mathcal{N}(\cdot; \mu, \sigma^2)$ denotes a Gaussian density with mean μ and variance σ^2, and $h_i(x)$ is the contribution to pixel i from the state x, which depends on the point spread function, target location and reflected energy. Note that (2) also holds for nonadditive models; see [4], for example.

Under the following assumptions:

- conditioned on the multi-object state, the values of the pixels are independently distributed, and
- the regions of influences of the objects on the image do not overlap, i.e., $x \neq x' \Rightarrow T(x) \cap T(x') = \emptyset$;

the probability density of the observation conditioned on the multi-object state X is given by

$$\begin{aligned} g(y \,|\, X) &= \left(\prod_{x \in X} \prod_{i \in T(x)} \varphi_i(y_i, x) \right) \left(\prod_{i \notin \cup_{x \in X} T(x)} \phi_i(y_i) \right) \\ &= f(y) \prod_{x \in X} g_y(x) \end{aligned} \tag{3}$$

where

$$g_y(x) = \prod_{i \in T(x)} \frac{\varphi_i(y_i, x)}{\phi_i(y_i)}$$
$$f(y) = \prod_{i=1}^{m} \phi_i(y_i).$$

We refer to multi-object likelihood functions of the form (3) as *separable*.

Remark: The terms $f(y)$ and $g_y(x)$ in the representation of the separable likelihood function (3) are unique in the following sense. Suppose that $f(y)$ and $g_y(x)$ are scaled by constants C_1 and C_2 (independent of x, y), respectively, i.e., $f(y)$ and $g_y(x)$ becomes $C_1 f(y)$ and $C_2 g_y(x)$, then

$$C_1 f(y) \prod_{x \in X} C_2 g_y(x) = C_1 C_2^{|X|} g(y \,|\, X)$$

where $|X|$ denotes the number of elements in X. This expression is equal to the multi-object measurement likelihood $g(y \,|\, X)$ for all $X \in \mathcal{F}(\mathcal{X})$ if and only if $C_1 = C_2 = 1$. Consequently, both $f(y)$ and $g_y(x)$ cannot be scaled, even though $f(y)$ cancels out in the calculation of the posterior density $\pi(\cdot \,|\, y)$ via (1), i.e.

$$\pi(X \,|\, y) = \frac{\prod_{x \in X} g_y(x) \pi(X)}{\int \prod_{x \in X} g_y(x) \pi(X) \delta X}.$$

III. Analytic Characterization of the Multi-Object Posterior

This section presents analytic characterizations of the multi-object posterior distribution for the observation model in the previous section and three classes of multi-object priors, namely Poisson, independently and identically distributed (i.i.d.) cluster, and Multi-Bernoulli. These multi-object priors are described next in Section III-A, along with a summary of the mathematical tools used in this work. The main results are presented in Section III-B.

A. Probability Laws and Statistics for RFS

Apart from the probability density, the probability generating functional (PGFl) is another fundamental descriptor of an RFS. Following [3], [10], the *probability generating functional* (PGFl) $G[\cdot]$ of an RFS X on $\mathcal{X}$ is defined by

$$G[h] \equiv \mathbb{E}[h^X] \tag{4}$$

where $\mathbb{E}$ denotes the expectation operator, h is any real-valued function on $\mathcal{X}$ such that $0 \leq h(x) \leq 1$, and

$$h^X \equiv \prod_{x \in X} h(x)$$

with $h^{\emptyset} = 1$ by convention.

The *cardinality* (number of elements) of X, denoted as $|X|$, is a discrete random variable whose *probability generating function* PGF $G(\cdot)$ can be obtained by setting the function h in the PGFl $G[\cdot]$ to a constant z. Note the distinction between the PGF and PGFl by the round and square brackets on the argument. The probability distribution ρ of the cardinality $|X|$ is the Z-transform of the PGF $G(\cdot)$.

The PHD, also known in point process theory as an *intensity function*, is a first-order statistical moment of an RFS, which can be obtained by differentiating the PGFl [3], [10]. For an RFS X on $\mathcal{X}$, its PHD is a nonnegative function v on $\mathcal{X}$ such that for each region $S \subseteq \mathcal{X}$

$$\mathbb{E}[|X \cap S|] = \int_S v(x) dx. \tag{5}$$

In other words, the integral of v over any region S gives the expected number of elements of X that are in S. The local maxima of the PHD are points in $\mathcal{X}$ with the highest local concentration of expected number of elements, and can be used to generate estimates for the elements of X.

There are two simple multi-object estimators which can be obtained from the (posterior) PHD and cardinality distribution. In the first estimator, the estimated number of states, $\hat{N}$, is determined by rounding the PHD mass $\int v(x) dx$, and the estimated states are chosen to be the $\hat{N}$ highest maxima of the PHD v. The second estimator is the same as the first except that the estimated number of states is $\arg\max \rho$, i.e., the maximum *a posteriori* cardinality estimate.

The RFSs pertinent to our key results and their PGFls are summarized in the following, using the standard inner product notation $\langle v, h \rangle = \int v(x) h(x) dx$.

Poisson: A *Poisson* RFS X on $\mathcal{X}$ is one that is completely characterized by its PHD or intensity function v [3]. The cardinality of a Poisson RFS is Poisson with mean $\langle v, 1\rangle$, and for a given cardinality the elements of X are each independent and identically distributed with probability density $v/\langle v, 1\rangle$. The PGFl of a Poisson RFS is

$$G[h] = e^{\langle v, h-1\rangle}. \tag{6}$$

I.I.D. Cluster: An *independent and identically distributed* (i.i.d.) cluster RFS X on $\mathcal{X}$ is completely characterized by a cardinality distribution ρ and a PHD v that satisfy $\sum_{n=0}^{\infty} n\rho(n) = \langle v, 1\rangle$ [3]. For a given cardinality, the elements of an i.i.d. cluster RFS X are each i.i.d. with probability density $v/\langle v, 1\rangle$. The PGFl $G[\,\cdot\,]$ of an i.i.d. cluster RFS is given by

$$G[h] = G\left(\frac{\langle v, h\rangle}{\langle v, 1\rangle}\right) \tag{7}$$

where $G(\,\cdot\,)$ is the probability generating function of the cardinality $|X|$, i.e., the inverse Z-transform of ρ. Note the distinction between the square brackets for functional and round brackets for function.

Bernoulli: A *Bernoulli* RFS X on $\mathcal{X}$ has probability $1 - r$ of being empty, and probability r of being a singleton whose (only) element is distributed according to a probability density p (defined on $\mathcal{X}$). The cardinality distribution of a Bernoulli RFS is a Bernoulli distribution with parameter r. The PGFl of a Bernoulli RFS is (see [11, p. 375])

$$G[h] = 1 - r + r\langle p, h\rangle. \tag{8}$$

Multi-Bernoulli: A *Multi-Bernoulli* RFS X on $\mathcal{X}$ is a union of a fixed number of independent Bernoulli RFSs $X^{(i)}$ with existence probability $r^{(i)} \in (0, 1)$ and probability density $p^{(i)}$ (defined on $\mathcal{X}$), $i = 1, \ldots, M$, i.e.

$$X = \bigcup_{i=1}^{M} X^{(i)}.$$

Using the independence of the $X^{(i)}$, the PGFl of a Multi-Bernoulli RFS is given by (see [11, p. 375])

$$G[h] = \prod_{i=1}^{M} \left(1 - r^{(i)} + r^{(i)} \left\langle p^{(i)}, h\right\rangle\right). \tag{9}$$

A Multi-Bernoulli RFS is thus completely described by the Multi-Bernoulli parameters $\{(r^{(i)}, p^{(i)})\}_{i=1}^{M}$. The parameter $r^{(i)}$ is the existence probability of the ith object while $p^{(i)}$ is the probability density of the state conditioned on its existence. The idea of characterizing target presence/absence by existence probability has also been considered in [15] and [21]. For convenience, the PGFl of the form (9) is abbreviated by $\{(r^{(i)}, p^{(i)})\}_{i=1}^{M}$. The term Multi-Bernoulli is also used to mean a PGFl or a probability density of a Multi-Bernoulli RFS.

Since the PHD of a Multi-Bernoulli is given by

$$v(x) = \sum_{i=1}^{M} r^{(i)} p^{(i)}(x)$$

and the cardinality of a Multi-Bernoulli RFS is a discrete Multi-Bernoulli random variable with parameters $r^{(1)}, \ldots, r^{(M)}$, estimators based on the PHD and cardinality distribution are applicable. Moreover, a more intuitive multi-object estimator can be obtained from the existence probabilities $r^{(i)}$ (and $p^{(i)}$). Given $\{(r^{(i)}, p^{(i)})\}_{i=1}^{M}$, similar to the PHD-based estimator, the estimated number of states $\hat{N}$ is determined by rounding $\arg\max \rho$. However, the estimated states are chosen to be the $\hat{N}$ means (or modes) of the probability densities in $\{(r^{(i)}, p^{(i)})\}_{i=1}^{M}$ with highest existence probabilities.

B. Closed Form Data Updates

We first present a result concerning the posterior PGFl for the observation likelihood considered in this work (see Section II-B), which allows, with surprising simplicity, the posterior distribution for Poisson, i.i.d. cluster, and Multi-Bernoulli RFS to be characterized analytically.

Proposition 1: Suppose that X is a random finite set on $\mathcal{X}$, with prior PGFl G, and y is a vector observation of X with separable likelihood function, i.e.

$$g(y \,|\, X) = f(y) g_y^X.$$

Then, the posterior PGFl $G[\cdot \,|\, y]$ of X given y is

$$G[h \,|\, y] = \frac{G[hg_y]}{G[g_y]}.$$

Proof: Recall that $\pi(\,\cdot\,)$ and $\pi(\cdot \,|\, y)$ denote the prior and posterior probability densities, respectively. Applying the definition of the PGFl and using (1), i.e., Bayes rule, to obtain the posterior probability density $\pi(\cdot \,|\, y)$ gives

$$\begin{aligned} G[h \,|\, y] &= \int h^X \pi(X \,|\, y)\delta X \\ &= \frac{\int h^X g(y \,|\, X)\pi(X)\delta X}{\int g(y \,|\, X')\pi(X')\delta X'} \\ &= \frac{f(y)\int [hg_y]^X \pi(X)\delta X}{f(y)\int g_y^{X'}\pi(X')\delta X'} \\ &= \frac{G[hg_y]}{G[g_y]}. \end{aligned}$$

□

For a Poisson RFS prior, which is completely characterized by the PHD, the following result shows how the PHD is updated with the observation y, i.e., how the posterior PHD is computed from the prior and the observation.

Corollary 1: Under the premise of Proposition 1, if the prior distribution of X is Poisson with PHD v, then the posterior distribution is also Poisson with PHD $v(\cdot \,|\, y)$ given by

$$v(x \,|\, y) = v(x) g_y(x).$$

Proof: Since X is Poisson with PHD v, its PGFl is given by $G[h] = e^{\langle v, h-1\rangle}$. Using Proposition 1

$$\begin{aligned} G[h \,|\, y] &= \frac{G[hg_y]}{G[g_y]} = \frac{e^{\langle v, hg_y - 1\rangle}}{e^{\langle v, g_y - 1\rangle}} \\ &= e^{\langle v, hg_y - g_y\rangle} = e^{\langle vg_y, h-1\rangle}. \end{aligned}$$

Thus, the posterior is Poisson with PHD vg_y. □

A weaker result has been established in [16] where it was shown that the posterior PHD is proportional to vg_y. Corollary 1 shows that the posterior PHD is equal to vg_y, and that the posterior RFS is Poisson. This result can be generalized to i.i.d. cluster RFSs that are completely characterized by the PHD and cardinality distribution as follows.

Corollary 2: Under the premise of Proposition 1, if the prior distribution of X is i.i.d. cluster with PHD v, and cardinality distribution ρ, then the posterior is also i.i.d. cluster with PHD $v(\cdot\,|\,y)$ and cardinality distribution $\rho(\cdot\,|\,y)$ given by

$$v(x\,|\,y) = v(x)g_y(x)\frac{\sum_{i=0}^{\infty}\frac{(i+1)\rho(i+1)}{\langle v,1\rangle^{i+1}}\langle v,g_y\rangle^i}{\sum_{j=0}^{\infty}\rho(j)\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^j} \tag{10}$$

$$\rho(n\,|\,y) = \frac{\rho(n)\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^n}{\sum_{j=0}^{\infty}\rho(j)\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^j}. \tag{11}$$

Proof: Since X is i.i.d. cluster with intensity v and cardinality distribution ρ, its PGFl is

$$G[h] = G\left(\frac{\langle v,h\rangle}{\langle v,1\rangle}\right)$$

where $G(z) = \sum_{j=0}^{\infty}\rho(j)z^j$ is the PGF of the cardinality of X. Using Proposition 1

$$\begin{aligned} G[h\,|\,y] &= \frac{G[hg_y]}{G[g_y]} = \frac{G\left(\frac{\langle vg_y,h\rangle}{\langle v,1\rangle}\right)}{G\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)} \\ &= \frac{\sum_{n=0}^{\infty}\rho(n)\left(\frac{\langle vg_y,h\rangle}{\langle v,1\rangle}\right)^n}{\sum_{j=0}^{\infty}\rho(j)\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^j} \\ &= \frac{\sum_{n=0}^{\infty}\rho(n)\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^n\left(\frac{\langle vg_y,h\rangle}{\langle vg_y,1\rangle}\right)^n}{\sum_{j=0}^{\infty}\rho(j)\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^j} \end{aligned} \tag{12}$$

where (12) follows from the identity $\langle v,g_y\rangle = \langle vg_y,1\rangle$. To establish that the posterior RFS is indeed an i.i.d. cluster with PHD (10) and cardinality distribution (11), we need to show that the posterior PGFl has the form $G[h\,|\,y] = G((\langle v(\cdot\,|\,y),h\rangle/\langle v(\cdot\,|\,y),1\rangle)|y)$. Note from (10) that $(\langle vg_y,h\rangle/\langle vg_y,1\rangle) = (\langle v(\cdot\,|\,y),h\rangle/\langle v(\cdot\,|\,y),1\rangle)$ (since the quotient of the infinite sums cancel), hence substituting for $(\langle vg_y,h\rangle/\langle vg_y,1\rangle)$ in (12) and using (11) gives

$$\begin{aligned} G[h\,|\,y] &= \sum_{n=0}^{\infty}\frac{\rho(n)\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^n}{\sum_{j=0}^{\infty}\rho(j)\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^j}\left(\frac{\langle v(\cdot\,|\,y),h\rangle}{\langle v(\cdot\,|\,y),1\rangle}\right)^n \\ &= \sum_{n=0}^{\infty}\rho(n\,|\,y)\left(\frac{\langle v(\cdot\,|\,y),h\rangle}{\langle v(\cdot\,|\,y),1\rangle}\right)^n \\ &= G\left(\left.\frac{\langle v(\cdot\,|\,y),h\rangle}{\langle v(\cdot\,|\,y),1\rangle}\right|y\right). \end{aligned}$$

Therefore, the posterior is an i.i.d. cluster process with PHD (10) and cardinality distribution (11). □

Remark: The posterior PHD (10) and cardinality distribution (11) were obtained by differentiating the posterior PGFl $G[\cdot\,|\,y]$ and PGF $G(\cdot\,|\,y)$, respectively. However, for the proof of Corollary 2, it is not necessary to show these steps.

Corollary 1 is a special case of Corollary 2 where the cardinality is Poisson distributed. Whereas Corollaries 1 and 2 characterize the posterior distribution by the PHD and cardinality distribution, the following result characterizes the posterior distribution by a set of existence probabilities and probability densities.

Corollary 3: Under the premise of Proposition 1, if the prior distribution of X is Multi-Bernoulli with parameter set $\{(r^{(i)},p^{(i)}\}_{i=1}^N$, then the posterior is also Multi-Bernoulli, with parameter set

$$\left\{\left(\frac{r^{(i)}\left\langle p^{(i)},g_y\right\rangle}{1-r^{(i)}+r^{(i)}\left\langle p^{(i)},g_y\right\rangle},\frac{p^{(i)}g_y}{\left\langle p^{(i)},g_y\right\rangle}\right)\right\}_{i=1}^N. \tag{13}$$

Proof: Since X is Multi-Bernoulli, with parameter set $\{(r^{(i)},p^{(i)}\}_{i=1}^N$, its PGFl is given by $G[h] = \prod_i^N(1-r^{(i)}+r^{(i)}\langle p^{(i)},h\rangle)$. Using Proposition 1

$$\begin{aligned} G[h\,|\,y] &= \frac{G[hg_y]}{G[g_y]} \\ &= \frac{\prod_{i=1}^N\left(1-r^{(i)}+r^{(i)}\left\langle p^{(i)},hg_y\right\rangle\right)}{\prod_{i=1}^N\left(1-r^{(i)}+r^{(i)}\left\langle p^{(i)},g_y\right\rangle\right)} \\ &= \prod_{i=1}^N\left(\frac{1-r^{(i)}+r^{(i)}\left\langle p^{(i)}g_y,h\right\rangle}{1-r^{(i)}+r^{(i)}\left\langle p^{(i)},g_y\right\rangle}\right) \\ &= \prod_{i=1}^N\left(1-\frac{r^{(i)}\left\langle p^{(i)},g_y\right\rangle}{1-r^{(i)}+r^{(i)}\left\langle p^{(i)},g_y\right\rangle}\right. \\ &\quad \left.+\frac{r^{(i)}\left\langle p^{(i)},g_y\right\rangle}{1-r^{(i)}+r^{(i)}\left\langle p^{(i)},g_y\right\rangle}\left\langle\frac{p^{(i)}g_y}{\left\langle p^{(i)},g_y\right\rangle},h\right\rangle\right). \end{aligned}$$

The ith term in the above product is the PGFl of a Bernoulli RFS. Hence, the posterior is Multi-Bernoulli, with parameter set given by (13). □

C. Multiple Sensor Data Update

Each of the corollaries in the Section III-B can be easily extended to the multiple sensor case, as long as the likelihood functions of the sensors are separable and that the sensors are independent conditional on the multi-object state. Suppose that there are two conditionally independent sensors with observations $y^{(1)}$ and $y^{(2)}$. Then the multi-sensor multi-object posterior density is

$$\pi\left(X\left|y^{(1)},y^{(2)}\right.\right) \propto g\left(y^{(1)}\middle|X\right)g\left(y^{(2)}\middle|X\right)\pi(X).$$

This is the same as updating with $y^{(1)}$ first and then $y^{(2)}$ (or vice-versa).

Unlike the PHD or Multi-Bernoulli updates for point measurements, which are approximations, the posterior parameter updates in Corollaries 1 to 3 exactly capture the necessary and sufficient statistics of the posterior multi-object densities. As such, for multiple sensor update, the exact posterior parameters can be iteratively computed by updating the prior parameters

with sensor 1, then treating the updated parameters as the prior parameters and updating this with sensor 2, and so forth. This procedure is repeated until the list of sensors is exhausted. Since each updates are exact, the end result is exact and independent of the order in which the updates are done.

IV. Multi-Object Filtering With Image Data

This section considers the multi-object filtering problem for image data. Unlike the static setting in the previous section, the multi-object state evolves in time and generates an image observation at each sampling instance. Hence, not only do the values of the states evolve, but the number of states also evolves due to objects appearing or disappearing. Multi-object filtering involves the online estimation of the multi-object state from collected data.

In what follows, we use the Multi-Bernoulli update in the previous section (Corollary 3) to develop a multi-object filtering algorithm for image observations. The filtering formulation and the proposed multi-object filter is described in Section IV-A, while the particle implementation is described in Section IV-B. Similar algorithms can be developed using the PHD update (Corollary 1) or i.i.d. cluster update (Corollary 2). However, since the observation model is highly nonlinear, particle implementations are employed to approximate the PHD, and clustering is needed to extract the estimated states from the particles. The clustering step introduces an additional source of error as well as being computationally expensive [22]. The Multi-Bernoulli approach avoids this problem altogether [25].

A. Multi-Bernoulli Filter for Image Data

The multi-object filtering problem can be cast as a Bayes filter on the space of finite sets $\mathcal{F}(\mathcal{X})$. Let y_k denote the image observation at time k, and $y_{1:k} = (y_1, \ldots, y_k)$ denotes the history of image observations from time 1 to time k. Then, the *multi-object Bayes recursion* propagates the *multi-object posterior density* $\pi_k(\cdot \,|\, y_{1:k})$ in time [10], [11] according to the following prediction and update steps:

$$\pi_{k\,|\,k-1}(X_k \,|\, y_{1:k-1}) = \int f_{k\,|\,k-1}(X_k \,|\, X)\pi_{k-1}(X \,|\, y_{1:k-1})\delta X, \quad (14)$$

$$\pi_k(X_k \,|\, y_{1:k}) = \frac{g_k(y_k \,|\, X_k)\pi_{k\,|\,k-1}(X_k \,|\, y_{1:k-1})}{\int g_k(y_k \,|\, X)\pi_{k\,|\,k-1}(X \,|\, y_{1:k-1})\delta X} \quad (15)$$

where the integrals above are set integrals, $f_{k\,|\,k-1}(\cdot\,|\,\cdot)$ is the *multi-object transition density*, from time $k-1$ to k, and $g_k(\cdot\,|\,\cdot)$ is the *multi-object likelihood* at time k.

The multi-object transition density $f_{k\,|\,k-1}(\cdot\,|\,\cdot)$ encapsulates the underlying models of motions, births and deaths. A popular multi-object transition model is the following. Given a multi-object state X_{k-1} at time $k-1$, each x_{k-1} in X_{k-1} either continues to exist at time k with probability $p_{S,k}(x_{k-1})$ and moves to a new state x_k with probability density[1] $f_{k\,|\,k-1}(x_k \,|\, x_{k-1})$, or dies with probability $1-p_{S,k}(x_{k-1})$. Thus, given a state x_{k-1} at time $k-1$, its behavior at time k is modeled by the Bernoulli RFS

$$S_{k\,|\,k-1}(x_{k-1})$$

with $r = p_{S,k}(x_{k-1})$ and $p(\cdot) = f_{k\,|\,k-1}(\cdot \,|\, x_{k-1})$. The multi-object state X_k at time k is given by the union

$$X_k = \bigcup_{x_{k-1} \in X_{k-1}} S_{k\,|\,k-1}(x_{k-1}) \cup \Gamma_k \quad (16)$$

where Γ_k denotes the Multi-Bernoulli RFS of spontaneous births. Assuming that the RFSs constituting the union in (16) are mutually independent, X_k is a Multi-Bernoulli RFS conditioned on X_{k-1}. Using FISST, the multi-object transition density $f_{k\,|\,k-1}(\cdot\,|\,\cdot)$ can be derived from the transition equation (16), [10], [11].

Since objects do not overlap in the image, it is necessary that the multi-object transition model assigns zero likelihood to multi-object states that contain overlapping objects. More concisely, $f_{k\,|\,k-1}(X \,|\, X') = 0$ if there exist distinct x_1 and $x_2 \in X$ such that $T(x_1) \cap T(x_2) \neq \emptyset$. However, assuming that the objects occupy relatively small regions of the image, the standard multi-object transition model above serves as a reasonable approximation.

The Bayes recursion (14)–(15) is generally intractable. However, under the assumption that the extents of the objects in the image are small, the predicted multi-object density, $\pi_{k\,|\,k-1}(\cdot\,|\,y_{1:k-1})$ is Multi-Bernoulli if $\pi_{k-1}(\cdot\,|\,y_{1:k-1})$ is Multi-Bernoulli [11]. Moreover, by Corollary 3, the updated multi-object density $\pi_k(\cdot\,|\,y_{1:k})$ is also a Multi-Bernoulli if the objects do not overlap. Hence the prediction step (14) and update step (15) can be approximated via the following.

Multi-Bernoulli Prediction: Given the posterior Multi-Bernoulli parameters $\pi_{k-1} = \{(r_{k-1}^{(i)}, p_{k-1}^{(i)})\}_{i=1}^{M_{k-1}}$, at time $k-1$, the predicted Multi-Bernoulli parameters are

$$\pi_{k\,|\,k-1} = \left\{\left(r_{P,k\,|\,k-1}^{(i)}, p_{P,k\,|\,k-1}^{(i)}\right)\right\}_{i=1}^{M_{k-1}} \cup \left\{\left(r_{\Gamma,k}^{(i)}, p_{\Gamma,k}^{(i)}\right)\right\}_{i=1}^{M_{\Gamma,k}} \quad (17)$$

where

$$r_{P,k\,|\,k-1}^{(i)} = r_{k-1}^{(i)} \left\langle p_{k-1}^{(i)}, p_{S,k} \right\rangle \quad (18)$$

$$p_{P,k\,|\,k-1}^{(i)}(x) = \frac{\left\langle f_{k\,|\,k-1}(x\,|\,\cdot), p_{k-1}^{(i)} p_{S,k} \right\rangle}{\left\langle p_{k-1}^{(i)}, p_{S,k} \right\rangle} \quad (19)$$

$f_{k\,|\,k-1}(\cdot\,|\,\zeta)$ = single target transition density at time k, given previous state ζ,

$p_{S,k}(\zeta)$ = probability of target existence at time k, given previous state ζ,

$\left\{\left(r_{\Gamma,k}^{(i)}, p_{\Gamma,k}^{(i)}\right)\right\}_{i=1}^{M_{\Gamma,k}}$ = parameters of the multi−Bernoulli RFS of births at time k.

[1]The same notation is used for multi-object and single-object densities. There is no danger of confusion since for the single-object case the arguments are vectors whereas for the multi-object case the arguments are finite sets.

Multi-Bernoulli Update: Given the predicted Multi-Bernoulli parameters $\pi_{k\,|\,k-1} = \{(r^{(i)}_{k\,|\,k-1}, p^{(i)}_{k\,|\,k-1}\}^{M_{k\,|\,k-1}}_{i=1}$, the updated Multi-Bernoulli parameters are

$$\pi_k = \left\{\left(r^{(i)}_k, p^{(i)}_k\right)\right\}^{M_{k\,|\,k-1}}_{i=1} \tag{20}$$

where

$$r^{(i)}_k = \frac{r^{(i)}_{k\,|\,k-1}\left\langle p^{(i)}_{k\,|\,k-1}, g_y\right\rangle}{1 - r^{(i)}_{k\,|\,k-1} + r^{(i)}_{k\,|\,k-1}\left\langle p^{(i)}_{k\,|\,k-1}, g_y\right\rangle} \tag{21}$$

$$p^{(i)}_k = \frac{p^{(i)}_{k\,|\,k-1} g_y}{\left\langle p^{(i)}_{k\,|\,k-1}, g_y\right\rangle}. \tag{22}$$

Track Merging (Heuristic): To account for the nonoverlapping assumption, estimates that would overlap on the image observation are merged. A simple way of merging is to combine the existence probabilities $r^{(i)}_k$, and densities $p^{(i)}_k$ of hypothesized objects whose estimates fall within a given distance T_{merge} of each other.

State Extraction (Heuristic): Following [25], an intuitive state extraction procedure can be used which is based on the Multi-Bernoulli form of the posterior density: the existence probability $r^{(i)}_k$ indicates how likely it is that the ith hypothesized track is a true track, and the posterior density $p^{(i)}_k$ describes estimated statistics of the current state of the track. Hence, a multi-object state estimate can be obtained by first estimating the number of objects as the expected or maximum *a posteriori* cardinality estimate, and then estimating the individual states by selecting the corresponding number of the means or modes from the state or track densities with the highest existence probabilities. Note also that other state extraction methods are possible, but for simplicity these will not be used in this paper.

B. Sequential Monte Carlo Implementation

In the following, we present a generic sequential Monte Carlo (SMC) implementation for the Multi-Bernoulli prediction step (17) (taken from [25]) and for the new update (20) step.

SMC Prediction: Suppose that at time $k-1$, the (Multi-Bernoulli) posterior multi-object density $\pi_{k-1} = \{(r^{(i)}_{k-1}, p^{(i)}_{k-1})\}^{M_{k-1}}_{i=1}$ is given and each $p^{(i)}_{k-1}, i = 1, \ldots, M_{k-1}$, is comprised of a set of weighted samples $\{w^{(i,j)}_{k-1}, x^{(i,j)}_{k-1}\}^{L^{(i)}_{k-1}}_{j=1}$, i.e.

$$p^{(i)}_{k-1}(x) = \sum_{j=1}^{L^{(i)}_{k-1}} w^{(i,j)}_{k-1}\delta_{x^{(i,j)}_{k-1}}(x).$$

Then, given proposal densities $q^{(i)}_k(\cdot\,|\,x_{k-1}, y_k)$ and $b^{(i)}_k(\cdot\,|\,y_k)$, the predicted (Multi-Bernoulli) multi-object density (17) can be computed as follows:

$$r^{(i)}_{P,k\,|\,k-1} = r^{(i)}_{k-1}\sum_{j=1}^{L^{(i)}_{k-1}} w^{(i,j)}_{k-1} p_{S,k}\left(x^{(i,j)}_{k-1}\right),$$

$$p^{(i)}_{P,k\,|\,k-1}(x) = \sum_{j=1}^{L^{(i)}_{k-1}} \tilde{w}^{(i,j)}_{P,k\,|\,k-1}\delta_{x^{(i,j)}_{P,k\,|\,k-1}}(x)$$

$$r^{(i)}_{\Gamma,k} = \text{parameter given by birth model},$$

$$p^{(i)}_{\Gamma,k}(x) = \sum_{j=1}^{L^{(i)}_{\Gamma,k}} \tilde{w}^{(i,j)}_{\Gamma,k}\delta_{x^{(i,j)}_{\Gamma,k}}(x)$$

where

$$x^{(i,j)}_{P,k\,|\,k-1} \sim q^{(i)}_k\left(\cdot\,\middle|\,x^{(i,j)}_{k-1}, y_k\right), j = 1, \ldots, L^{(i)}_{k-1}$$

$$w^{(i,j)}_{P,k\,|\,k-1} = \frac{w^{(i,j)}_{k-1} f_{k\,|\,k-1}\left(x^{(i,j)}_{P,k\,|\,k-1}\,|\,x^{(i,j)}_{k-1}\right) p_{S,k}\left(x^{(i,j)}_{k-1}\right)}{q^{(i)}_k\left(x^{(i,j)}_{P,k\,|\,k-1}\,|\,x^{(i,j)}_{k-1}, y_k\right)},$$

$$\tilde{w}^{(i,j)}_{P,k\,|\,k-1} = w^{(i,j)}_{P,k\,|\,k-1}\Big/\sum_{j=1}^{L^{(i)}_{k-1}} w^{(i,j)}_{P,k\,|\,k-1}$$

$$x^{(i,j)}_{\Gamma,k} \sim b^{(i)}_k(\cdot\,|\,y_k) j = 1, \ldots, L^{(i)}_{\Gamma,k}$$

$$w^{(i,j)}_{\Gamma,k} = \frac{p_{\Gamma,k}\left(x^{(i,j)}_{\Gamma,k}\right)}{b^{(i)}_k\left(x^{(i,j)}_{\Gamma,k}\,|\,y_k\right)},$$

$$\tilde{w}^{(i,j)}_{\Gamma,k} = w^{(i,j)}_{\Gamma,k}\Big/\sum_{j=1}^{L^{(i)}_{\Gamma,k}} w^{(i,j)}_{\Gamma,k}\ .$$

SMC Update: Suppose that at time k, the predicted (Multi-Bernoulli) multi-object density $\pi_{k\,|\,k-1} = \{(r^{(i)}_{k\,|\,k-1}, p^{(i)}_{k\,|\,k-1})\}^{M_{k\,|\,k-1}}_{i=1}$ is given and each $p^{(i)}_{k\,|\,k-1}$, $i = 1, \ldots, M_{k\,|\,k-1}$, is comprised of a set of weighted samples $\{w^{(i,j)}_{k\,|\,k-1}, x^{(i,j)}_{k\,|\,k-1}\}^{L^{(i)}_{k-1}}_{j=1}$, i.e.

$$p^{(i)}_{k\,|\,k-1} = \sum_{j=1}^{L^{(i)}_{k\,|\,k-1}} w^{(i,j)}_{k\,|\,k-1}\delta_{x^{(i,j)}_{k\,|\,k-1}}(x).$$

Then, the updated (Multi-Bernoulli) multi-object density (20) computed as follows:

$$r^{(i)}_k = \frac{r^{(i)}_{k\,|\,k-1}\varrho^{(i)}_k}{1 - r^{(i)}_{k\,|\,k-1} + r^{(i)}_{k\,|\,k-1}\varrho^{(i)}_k}$$

$$p^{(i)}_k = \frac{1}{\varrho^{(i)}_k}\sum_{j=1}^{L^{(i)}_{k\,|\,k-1}} w^{(i,j)}_{k\,|\,k-1} g_{y_k}\left(x^{(i,j)}_{k\,|\,k-1}\right)\delta_{x^{(i,j)}_{k\,|\,k-1}}(x)$$

where $\varrho^{(i)}_k = \sum_{j=1}^{L^{(i)}_{k\,|\,k-1}} w^{(i,j)}_{k\,|\,k-1} g_{y_k}(x^{(i,j)}_{k\,|\,k-1})$.

Resampling and Implementation Issues: Like the Multi-Bernoulli filter [25], for each hypothesized object, the particles are resampled after the update step, and the number of particles is reallocated in proportion to the probability of existence, as well as restricted to be between a maximum of $L_{\max}$ and minimum of $L_{\min}$. To reduce the growing number of tracks (and particles), objects with existence probabilities below a threshold P are discarded.

The SMC implementation is equivalent to running many particle filters in parallel and is linear in the number of objects, while the track merging and removal is quadratic in the number of objects, since the distances between pairs of estimates need to be computed. The complexity of the merging step can however be reduced to $\mathcal{O}(n\log(n))$ with other computational strategies such as those presented in [5].

Remark: The SMC implementation presented here is reminiscent of sequential importance resampling in the particle filter. On the issue of the proposal function, there are several possible choices. The use of the predicted density is an obvious choice which simplifies computations and performs reasonably in most scenarios. Other choices, such as the track updates from another filter, may improve upon the results, although such considerations are beyond the scope of this paper.

V. Numerical Study

The idea of TBD and its importance in low SNR radar applications was first investigated in [18]. According to the latest survey [4], the Histogram Probabilistic Multi-Hypothesis Tracker (H-PMHT) proposed in [20] is, by far, the best previously developed multi-target TBD technique.

This section demonstrates the performance of our proposed Multi-Bernoulli filter in a TBD application. Comparisons with the H-PMHT implementation as given in [4] will also be shown as a benchmark where appropriate. For these comparisons, both H-PMHT and Multi-Bernoulli-TBD are supplied wiht a priori information on the numbers of targets, and both are given the same initialization, so as to ensure a fair overall comparison. Several scenarios are used to illustrate their relative strengths and weaknesses. In the first, a challenging nonlinear scenario with a fixed and known number of targets illustrates a typical situation where the proposed Multi-Bernoulli-TBD filter significantly outperforms the H-PMHT. This scenario is then extended to having an unknown and time varying number of targets to demonstrate the true capabilities of the Multi-Bernoulli-TBD filter. In the second, a simplified linear scenario with a fixed and known number of targets illustrates the limitation of our proposed approach. This scenario also serves to demonstrate that our implementation of H-PMHT is consistent with that in the survey [4].

A. Multi-Object Miss Distance

We use the optimal subpattern assignment (OSPA) distance between the estimated and true multi-object state as the estimation error, since it jointly captures differences in cardinality and individual elements between two finite sets in a mathematically consistent yet intuitively meaningful way [19].

An intuitive construction of the OSPA distance between two finite sets $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_n\}$ is as follows. The set X with the smaller cardinality is initially chosen as a reference. Determine the assignment between the m points of X and points of Y that minimizes the sum of the distances, subject to the constraint that distances are capped at a preselected maximum or cut-off value c. This minimum sum of distances can be interpreted as the "total localization error", where the pairings allocated by giving the points in X the "benefit of the doubt". All other points (of Y) that remain unassigned are also allocated an error value, whereby each extraneous point is penalized at the maximum or cut-off value c. These errors can interpreted as "cardinality errors," which are "penalized at the maximum rate." The "total error" incurred is then the sum of the "total localization error" and the "total cardinality error." Remarkably, the "per target error" obtained by normalizing "total error" by n (the larger cardinality of the two given sets) is a proper metric [19]. In other words, the "per target error" enjoys all the properties of the usual distance that we normally take for granted on a Euclidean space.

The OSPA metric $\bar{d}_p^{(c)}$ is defined as follows. Let $d^{(c)}(x,y) := \min(c, \|x-y\|)$ for $x, y \in \mathcal{X}$, and Π_k denote the set of permutations on $\{1,2,\ldots,k\}$ for any positive integer k. Then, for $p \geq 1$, $c > 0$, and $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_n\}$

$$\bar{d}_p^{(c)}(X,Y) := \left(\frac{1}{n} \left(\min_{\pi \in \Pi_n} \sum_{i=1}^{m} d^{(c)}(x_i, y_{\pi(i)})^p + c^p(n-m) \right) \right)^{\frac{1}{p}} \tag{23}$$

if $m \leq n$, and $\bar{d}_p^{(c)}(X,Y) := \bar{d}_p^{(c)}(Y,X)$ if $m > n$; and $\bar{d}_p^{(c)}(X,Y) = \bar{d}_p^{(c)}(Y,X) = 0$ if $m = n = 0$.

The OSPA distance is interpreted as a pth-order per-target error, comprised of a pth-order per-target localization error and a pth-order per-target cardinality error. The order parameter p determines the sensitivity of the metric to outliers, and the cut-off parameter c determines the relative weighting of the penalties assigned to cardinality and localization errors. When $p = 1$, the OSPA distance can be interpreted exactly as the sum of the "per-target localization error" and the "per-target cardinality error." For further details, see [19].

B. Observation Model

A common observation model is used for all our experiments. At each time step, the observation is a two-dimensional image consisting of an array of cells with a scalar intensity. In these demonstrations, the observation region is a 45 m $\times$ 45 m square region, and the observation image is a 45 pix $\times$ 45 pix array, giving each cell side lengths of $\Delta_x = \Delta_y = 1$ m. Since the observation array is a two dimensional image, the array index will be treated as an ordered pair of integers $i = (a,b)$, where $1 \leq a, b \leq 45$. The observation model is then given by the probability density of the intensity $y_{i,k}$ of pixel i, at time k, given a state x_k

$$p(y_{i,k} \,|\, x_k) = \begin{cases} \mathcal{N}(y_{i,k}; h_{i,k}(x_k), \sigma^2), & i \in T(x) \\ \mathcal{N}(y_{i,k}; 0, \sigma^2), & i \notin T(x) \end{cases} \tag{24}$$

where $\sigma^2 = 1$ is the noise variance, $h_{i,k}$ is the point spread function given by

$$h_{i,k}(x_k) = \frac{\Delta_x \Delta_y I_k}{2\pi\sigma_h^2} \times \exp\left(-\frac{(\Delta_x a - p_{x,k})^2 + (\Delta_y b - p_{y,k})^2}{2\sigma_h^2} \right)$$

with source intensity $I_k = 30$, blurring factor $\sigma_h^2 = 1$, and $(p_{x,k}, p_{y,k})$ being the position of the state x_k. The effective template $T(x_k)$ is the 4 pix $\times$ 4 pix square region whose center is

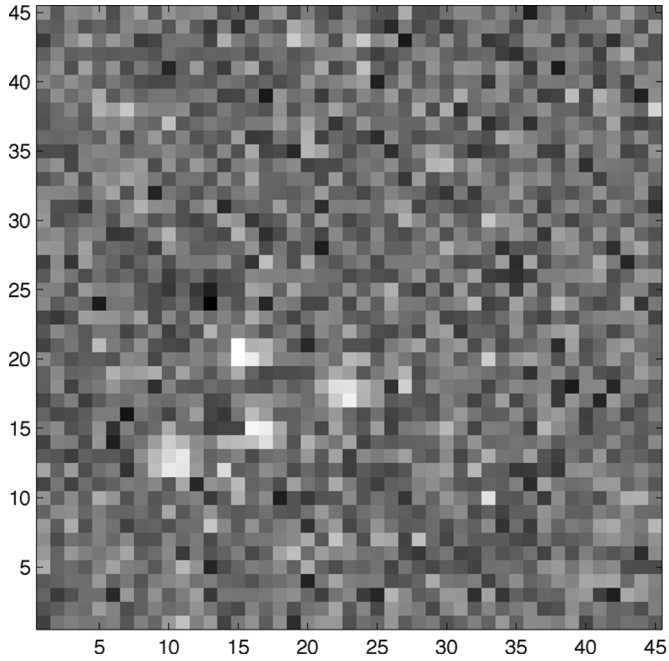

Fig. 4. An image observation of 4 targets at time 10.

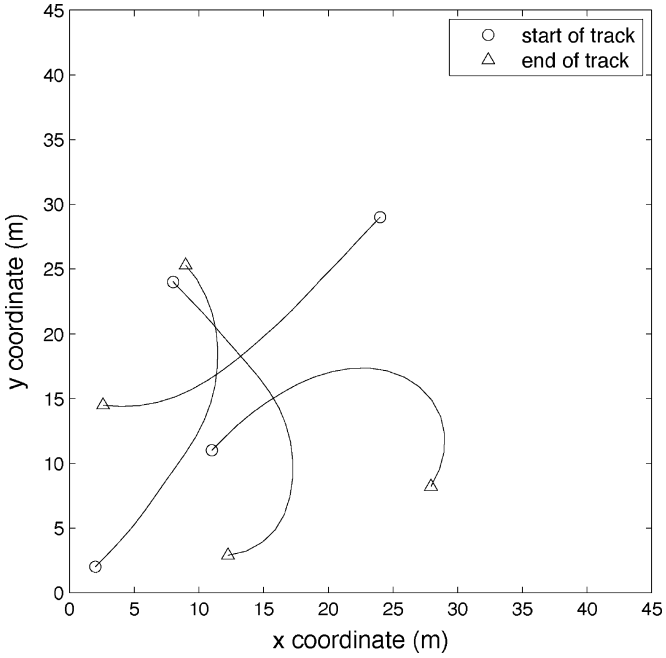

Fig. 5. True tracks on a 2-D plane for 4 targets.

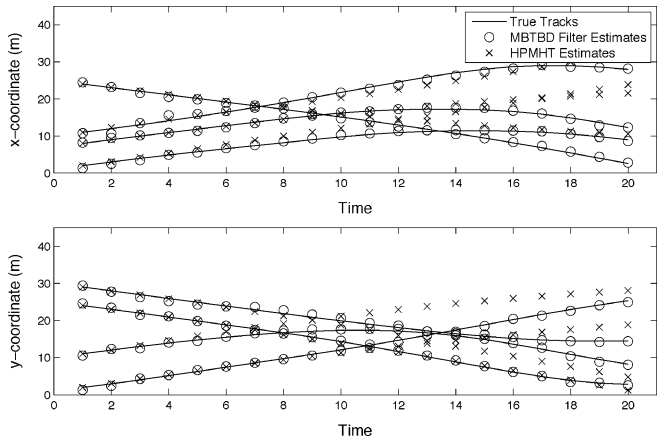

Fig. 6. Outputs of Multi-Bernoulli-TBD filter and H-PMHT for 4 targets with nonlinear dynamical model.

closest to the position of x_k. A typical observation is shown in Fig. 4.

C. Scenario 1

This scenario considers a nonlinear TBD problem in which a fixed and known number of targets follow curved paths. Four targets are present throughout the entire scenario lasting $K = 20$ s. Targets are initialized at separate locations. The single target state $x_k = [\tilde{x}_k^T, \omega_k]^T$ comprises the planar position and velocity $\tilde{x}_k^T = [p_{x,k}, p_{y,k}, \dot{p}_{x,k}, \dot{p}_{y,k}]^T$ and the turn rate ω_k. Target motions are modeled by a nonlinear nearly constant turn model given by

$$x_k = F(\omega_{k-1})x_{k-1} + Gw_{k-1},$$
$$\omega_k = \omega_{k-1} + \Delta u_{k-1}$$

where

$$F(\omega) = \begin{bmatrix} 1 & 0 & \frac{\sin\omega\Delta}{\omega} & -\frac{1-\cos\omega\Delta}{\omega} \\ 0 & 1 & \frac{1-\cos\omega\Delta}{\omega} & \frac{\sin\omega\Delta}{\omega} \\ 0 & 0 & \cos\omega\Delta & -\sin\omega\Delta \\ 0 & 0 & \sin\omega\Delta & \cos\omega\Delta \end{bmatrix},$$

$$G = \begin{bmatrix} \frac{\Delta^2}{2} & 0 \\ 0 & \frac{\Delta^2}{2} \\ \Delta & 0 \\ 0 & \Delta \end{bmatrix}$$

$\Delta = 1$ s, $w_k \sim \mathcal{N}(\cdot; 0, \sigma_w^2 I_2)$, $\sigma_w = 0.1$ m/s^2, $u_k \sim \mathcal{N}(\cdot; 0, \sigma_u^2)$, and $\sigma_u = (\pi/90)$ rad/s.

Both the Multi-Bernoulli-TBD and H-PMHT filters are initialized with the correct target positions. Since H-PMHT assumes prior knowledge of the fixed and known number of targets, to facilitate a fair comparison, the Multi-Bernoulli-TBD filter correspondingly assumes no target births nor deaths. The simulation uses a maximum of $L_{\max} = 5000$ and minimum of $L_{\min} = 1000$ particles per hypothesized track. Tracks with existence probabilities less than $r = 10^{-2}$ are dropped and a maximum of $T_{\max} = 100$ tracks are kept.

Fig. 5 shows the true tracks on a 2-D plane. The performance of the Multi-Bernoulli-TBD and H-PMHT filters are shown for a single sample run in Fig. 6. Note that while target observations appear close to one another, they do not overlap. It can be seen that the Multi-Bernoulli-TBD filter is able to maintain lock on all target tracks and accurately estimate their locations for the entire scenario. On the other hand, it is clear that the H-PMHT filter performs considerably worse. Although H-PMHT is initially able to track the targets, this is mainly due to being given correct initialization. The H-PMHT filter then loses these tracks very quickly due to the relatively high prediction error that generally leads to large deviations between the virtual and true measurements. Consequently, although H-PMHT attempts to

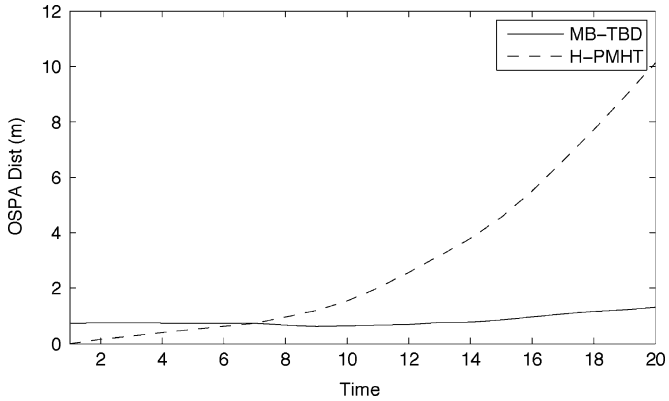

Fig. 7. Monte Carlo averaged OSPA miss distance with $c = 30$ m and $p = 1$ for Multi-Bernoulli-TBD filter and H-PMHT.

find accurate virtual measurements by expectation maximization, the large and difficult search space means that it usually unable to do so.

To confirm these single sample run observations, 1000 Monte Carlo trials are performed and averaged. Fig. 7 shows the estimation errors in terms of the Monte Carlo averaged OSPA distance (for $c = 30$ m, $p = 1$) for both the Multi-Bernoulli-TBD and H-PMHT approaches. These results confirm that the Multi-Bernoulli-TBD approach performs accurately and consistently in the sense that it maintains lock on all tracks and correctly estimates target positions. The estimation error of the H-PMHT, while initially smaller than that of the Multi-Bernoulli-TBD, increases rapidly and the filter eventually diverges. This occurs because the H-PMHT filter is given the correct initialization, and for the first few times steps, the predictions for the target states given by the dynamical model happen to fall close to the true states. After this time however, the predictions for the target states degrade, which eventually leads to complete track loss and consequently very high estimation errors.

An indicative assessment of the computational demands of the two filters is now shown by benchmarking their processing times on a standard laptop computer. With the same experimental setup, the computational times and estimation errors of the two filters are compared for varying observation image resolutions, that is, for varying values of the number of pixels per metre (and hence total number of pixels) in the observed image. Figs. 11 and 12 respectively show a comparison of the processing times and scenario averaged estimation errors. It can be seen that the processing times for the proposed Multi-Bernoulli TBD approach are fairly constant across the range of resolutions surveyed, whereas the processing times for the reference H-PMHT algorithm appear to increase exponentially with image resolution. Moreover, the estimation errors incurred by the Multi-Bernoulli TBD approach are fairly consistent, whereas the H-PMHT algorithm is still seen to diverge as indicated by the large miss distance values. This latest observation provides further confirmation of the comparative performance of the two filters, even when the image resolution is high, meaning that there is a higher information content in the observations. This can be expected since the Multi-Bernoulli TBD approach is better able to utilize the information contained in the observations in comparison to H-PMHT.

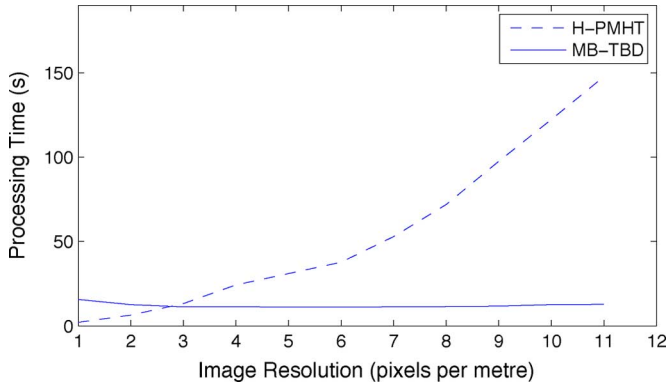

Fig. 8. Comparison of the processing times for varying image resolution.

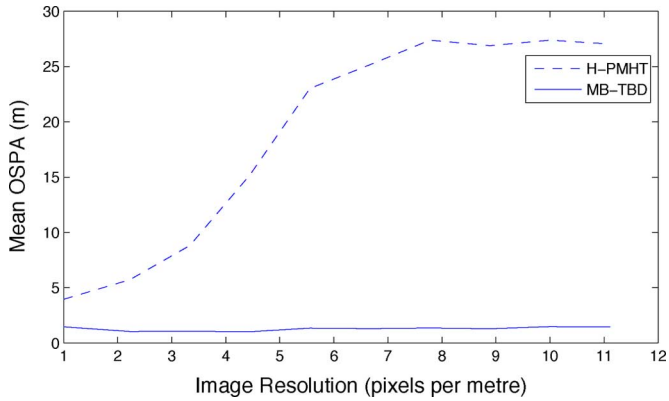

Fig. 9. Comparison of the estimation errors given by the time averaged OSPA miss distance with $c = 30$ m and $p = 1$ for varying image resolution.

Unknown number of targets: To further demonstrate the capabilities of the Multi-Bernoulli-TBD approach, this experiment is extended to having a time varying and unknown number of targets. A maximum of 4 targets are present at any time, and there are various target births and deaths throughout the scenario duration of $K = 20$ s. The same dynamical model is used except that births and deaths are now introduced into the model to accommodate the time varying number of targets. The birth process is a Multi-Bernoulli density $\pi_\Gamma = \{(r_\Gamma, p_\Gamma^{(i)})\}_{i=1}^{4}$ where $r_\Gamma = 0.01$, $p_\Gamma^{(i)}(x) = \mathcal{N}(x; m_\gamma^{(i)}, P_\gamma)$, $m_\gamma^{(1)} = [7, 25, 1, -1, \pi/360]^T$, $m_\gamma^{(2)} = [10, 10, 1, 1, \pi/360]^T$, $m_\gamma^{(3)} = [1, 1, 1, 1, \pi/360]^T$, $m_\gamma^{(4)} = [25, 30, -1, -1, \pi/360]^T$, and $P_\gamma = \mathrm{diag}([0.5, 0.5, 0.5, 0.5, 6(\pi/180)]^T)^2$. The death process is specified by a uniform single target probability of target survival $p_{S,k}(x) = 0.99$. The true tracks on a 2-D plane are shown Fig. 10. To cope with an increased number of hypotheses, in addition to track pruning, track merging is performed with threshold $T_{\mathrm{merge}} = 0.75$ times the pixel width.

The tracking results for a single sample are shown in Fig. 11. These results suggest that the Multi-Bernoulli-TBD approach is able to initiate, maintain, terminate and consequently estimate all target tracks satisfactorily, although there are occasional delays in the initiation and termination of tracks. This is confirmed by the results of 1000 Monte Carlo trials, for which the average estimation errors given by the OSPA distance (for $c = 30$ m, $p = 1$) are shown in Fig. 12, along with the breakdown into its localization and cardinality components. The localization error

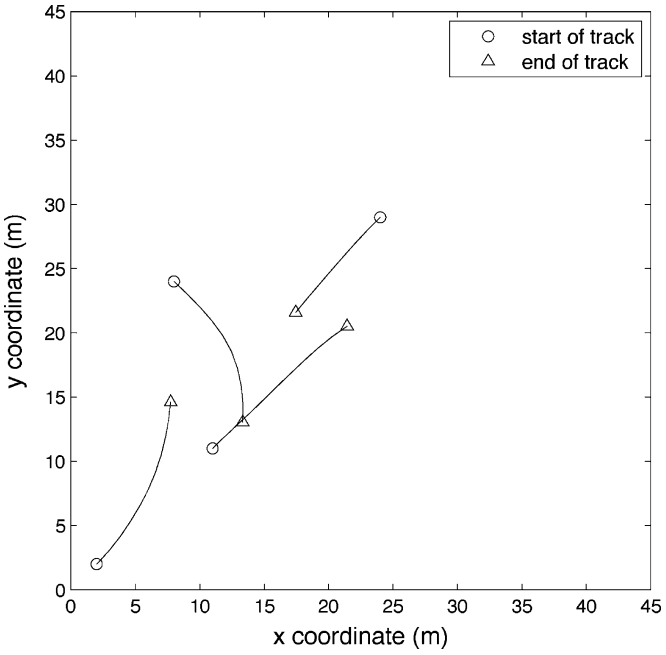

Fig. 10. True tracks on a 2-D plane for 4 targets appearing and disappearing at different times.

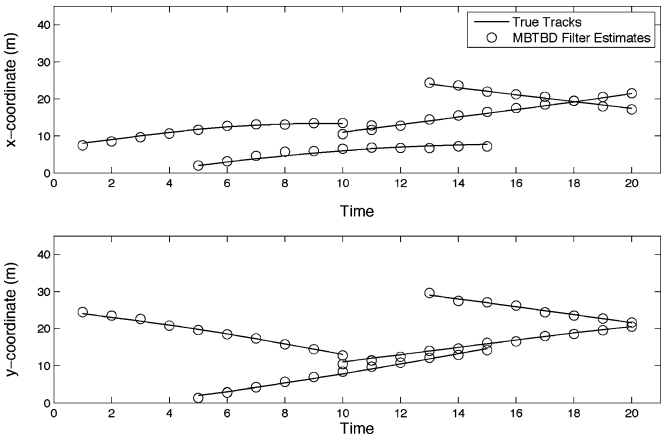

Fig. 11. Output of Multi-Bernoulli-TBD filter with an unknown and time varying number of targets following a nonlinear dynamical model.

does not grow with time, indicating that the position estimates are accurate and stable. The cardinality errors are also relatively stable, exhibiting jumps at instants involving target births or deaths.

Remark: Although this example addresses a TBD problem, the proposed Multi-Bernoulli filter can actually be applied to more general settings. For example, the proposed filter has been successfully applied in computer vision applications such as tracking sports players from a video stream [6].

D. Scenario 2

To show the limitation of the proposed filter when the nonoverlapping assumption is violated, consider the scenario adopted in the survey [4], with four targets initially colocated near the origin and moving along straight lines. The target state $x_k = [p_{x,k}, \dot{p}_{x,k}, p_{y,k}, \dot{p}_{y,k}]$ comprises only the position and

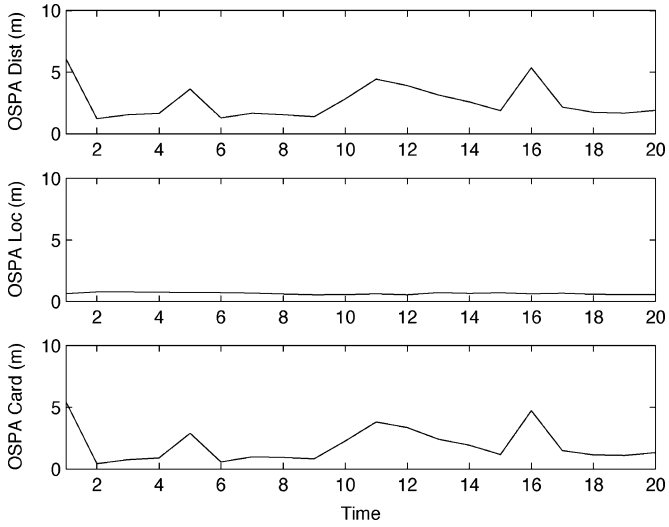

Fig. 12. Monte Carlo averaged OSPA miss distance with $c = 30$ m and $p = 1$ along with localization and cardinality components.

velocity of the target. The dynamical model is linear Gaussian given by

$$x_k = F x_{k-1} + w_k \tag{25}$$

where $w_k \sim \mathcal{N}(\cdot; 0, Q)$ is the process noise

$$F = \begin{bmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$Q = \sigma_\nu^2 \begin{bmatrix} \frac{\Delta^3}{3} & \frac{\Delta^2}{2} & 0 & 0 \\ \frac{\Delta^2}{2} & \Delta & 0 & 0 \\ 0 & 0 & \frac{\Delta^3}{3} & \frac{\Delta^2}{2} \\ 0 & 0 & \frac{\Delta^2}{2} & \Delta \end{bmatrix}$$

$\Delta = 1$ s is the sampling period, and $\sigma_\nu^2 = 0.001$ is the standard deviation of the process noise. Again, both the Multi-Bernoulli-TBD and H-PMHT filters are initialized with the correct target positions, and since H-PMHT assumes prior knowledge of the fixed and known number of targets, the Multi-Bernoulli-TBD filter correspondingly assumes no target births, deaths or spawns. A maximum of $L_{\max} = 5000$ and minimum of $L_{\min} = 1000$ particles per hypothesized track is imposed. Tracks with existence probabilities less than $r = 10^{-2}$ are dropped and a maximum of $T_{\max} = 100$ tracks are kept.

A comparison between the performance of the Multi-Bernoulli-TBD and H-PMHT approaches is shown in Fig. 13. In contrast with the previous scenario, H-PMHT performs better than the Multi-Bernoulli-TBD filter. The Multi-Bernoulli-TBD filter performs poorly since the targets are initially colocated, thereby violating the original rigid objects assumption in Section II-B. Consequently, the Multi-Bernoulli-TBD filter cannot verify the target tracks from the observed image due to the inconsistency with the assumed observational model. However, H-PMHT performs well in this scenario since the targets travel along straight lines, thereby allowing easy and accurate predictions. The virtual measurements computed by expectation maximization then converge very quickly to the

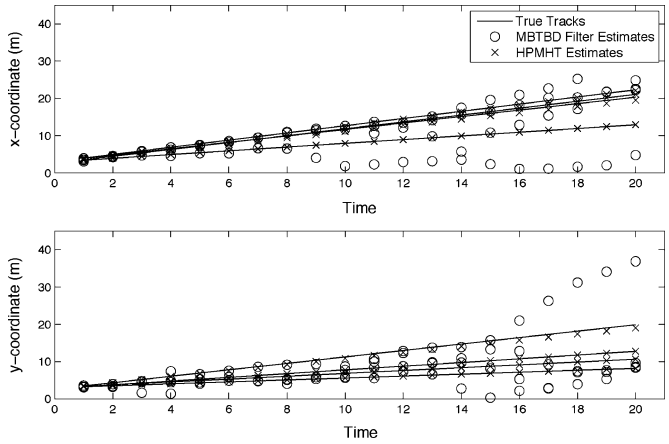

Fig. 13. Outputs of Multi-Bernoulli-TBD and H-PMHT for 4 linear tracks, initially colocated near the origin.

true measurements. Fig. 13 also shows that we have a valid implementation of H-PMHT, since our numerical result replicates that in the survey [4].

VI. Conclusion

We have demonstrated that the random finite set approach enabled a tractable solution to the multi-object estimation problem for image data and proposed a multi-object filter for image observations. Numerical studies show that the proposed multi-object filter is more effective than existing methods, when the regions of the image influenced by individual states do not overlap. There are, however, a number of limitations of the proposed multi-object filter. For applications with large images, the particle implementation suffers from particle degeneracy due to the reweighting of the particles by a function with extremely small support. More efficient particle implementations and analytic implementations are two venues for further work. Due to the assumption of nonoverlapping objects, the proposed filter does not handle target crossings. Extending the proposed approach to accommodate overlapping objects is an important problem that has wider applicability.

References

[1] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association*. San Diego, CA: Academic, 1988.
[2] S. Blackman, *Multiple Target Tracking With Radar Applications*. Norwood, MA: Artech House, 1986.
[3] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*. New York: Springer-Verlag, 1988.
[4] S. Davey, M. Rutten, and B. Cheung, "A comparison of detection performance for several track-before-detect algorithms," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, 2008, Article 41.
[5] A. Gray and A. Moore, "N-body problems in statistical learning," *Adv. Neural Inf. Process. Syst.*, no. 4, 2000.
[6] R. Hoseinnezhad, B.-N. Vo, D. Suter, and B.-T. Vo, "Multiobject filtering from image sequence without detection," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Dallas, TX, Mar. 2010.
[7] C. Hue, J.-P. Le Cadre, and P. Perez, "Sequential Monte Carlo methods for multiple target tracking and data fusion," *IEEE Trans. Signal Process.*, vol. 50, no. 2, Feb. 2002.
[8] J. Lund and M. Rudemo, "Models for point processes observed with noise," *Biometrika*, vol. 87, pp. 235–249, 2000.
[9] J. Lund and E. Thonnes, "Perfect simulation and inference for point processes given noisy observations," *Computat. Statist.*, vol. 19, pp. 317–336, 2004.
[10] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, 2003.
[11] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA: Artech House, 2007.
[12] R. Mahler, "Multitarget-moment filters for nonstandard measurement models," in *Proc. SPIE Defense Security Symp.*, 2008, vol. 6968.
[13] R. Mahler, "PHD filters of higher order in target number," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, Jul. 2007.
[14] J. Moller and R. Waagepeterson, "Modern statistics for spatial point processes," *Scand. J. Statist.*, vol. 34, pp. 643–684, 2006.
[15] D. Musicki and R. Evans, "Linear joint integrated probabilistic data association —LJIPDA," in *Proc. 41st IEEE Conf. Decision and Control*, Las Vegas, NV, 2002.
[16] N. T. Pham, H. Weimin, and S. Ong, "Tracking multiple objects using probability hypothesis density filter and color measurements," in *Proc. IEEE Int. Conf. Multimed. Expo*, Beijing, China, 2007, pp. 1511–1514.
[17] B. Ristic, S. Arulampalam, and N. J. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Norwood, MA: Artech House, 2004.
[18] D. J. Salmond and H. Birch, "A particle filter for track-before-detect," in *Proc. Amer. Control Conf.*, Arlington, VA, Jun. 2001, vol. 5, pp. 3755–3760.
[19] 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.
[20] R. L. Streit, M. L. Graham, and M. J. Walsh, "Multitarget tracking of distributed targets using histogram-PMHT," *Digital Signal Process.*, vol. 12, no. 2–3, pp. 394–404, 2002.
[21] J. Vermaak, S. Maskell, and M. Briers, "A unifying framework for multitarget tracking and existence," in *Proc. 8th Int. Conf. Inf. Fusion*, Philadelphia, PA, 2005.
[22] 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, 2005.
[23] 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.
[24] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3553–3567, Jul. 2007.
[25] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The cardinality balanced multitarget multi-Bernoulli filter and its implementations," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 409–423, Feb. 2009.
[26] B.-N. Vo, B.-T. Vo, N.-T. Pham, and D. Suter, "Bayesian multiobject estimation from image observations," in *Proc. 12th Int. Conf. Inf. Fusion*, Seattle, Wash. DC, 2009.

Ba-Ngu Vo received the Bachelor degrees jointly in science and electrical engineering with first class honors from the University of Western Australia, Crawley, in 1994, and the Ph.D. degree from the Curtin University of Technology, in 1997.

He has held various research positions before joining the Department of Electrical and Electronic Engineering, University of Melbourne, Australia, in 2000. Currently he is Winthrop Professor and Chair of Signal Processing in the School of Electrical Electronic and Computer Engineering, University of Western Australia. His research interests are signal processing, systems theory, and stochastic geometry with emphasis on target tracking, robotics, computer vision.

Dr. Vo is a recipient of the Australian Research Council's inaugural Future Fellowship.

Ba Tuong Vo was born in Perth, Australia, in 1982. He received the B.Sc. degree in applied mathematics and the B.E. degree in electrical and electronic engineering (with first-class honors) in 2004, and the Ph.D. degree in engineering (with Distinction) in 2008, all from the University of Western Australia, Crawley.

He is currently an Assistant Professor and Australian Postdoctoral Fellow with the School of Electrical, Electronic and Computer Engineering, University of Western Australia. His primary research interests are in point process theory, filtering and estimation, and multiple object filtering.

Nam-Trung Pham received the Ph.D. degree in computer engineering from the National University of Singapore in 2008.

After a year and a half working as a Research Engineer with INRIA Rennes, he joined the Department of Computer Vision and Image Understanding, Institute for Infocomm Research, Singapore, in 2009 as a Research Fellow. His current research focuses on multiple object tracking, object detection, random sets, and pattern recognition.

David Suter received the B.Sc. degree in applied mathematics and physics from The Flinders University of South Australia, in 1977, the Grad. Dip. Comp. from the Royal Melbourne Institute of Technology, in 1984, and the Ph.D. degree in computer science from La Trobe University in 1991.

He was a Lecturer with La Trobe University from 1988 to 1991. He then became a Senior Lecturer in 1992, Associate Professor in 2001, and Professor from 2006–2008 with Monash University, Melbourne, Australia. Since 2008, he has been a professor with the School of Computer Science, The University of Adelaide, Australia, where he is currently head of the School of Computer Science. He also serves on the Australian Research Council (ARC) College of Experts. His research activities focus on topics such as robust statistical methods, motion estimation from images (including optic flow), structure from motion, image segmentation, biomedical image analysis, human motion capture and animation, visual tracking, activity detection and classification, face recognition, and the construction of building models from laser scan, and image data.

Dr. Suter is an Associate Editor for the *International Journal of Computer Vision*, the *Journal of Mathematical Imaging and Vision*, and IPSJ *Transactions on Computer Vision and Applications* (having previously served as an Associate Editor for the *International Journal of Image and Graphics*, and for *Machine Vision and Applications*).