Bayesian Multi-Object Estimation from Image Observations

Ba-Ngu Vo
Dept of EEE
University of Melbourne
Parkville, VIC
bnvo@unimelb.edu.au

Ba-Tuong Vo
School of EECE
University of Western Australia
Crawley, WA
btv@ee.uwa.edu.au

Nam Trung Pham
IRISA/INRIA
Campus de Beaulieu
35042 Rennes Cedex, France
ntpham@irisa.fr

David Suter
School of Computer Science
University of Adelaide
Adelaide, SA
dsuter@cs.adelaide.edu.au

Abstract – *Analytic characterizations of the posterior distribution of a random finite set of states conditioned on image observation are derived, under the assumption that the regions of the observation influenced by individual states do not overlap. These results provide tractable means to jointly estimate the number of states and their values in the Bayesian framework. As an application, we develop a multi-object filter suitable for image observations with low signal to noise ratio. A particle implementation of the multi-object filter is proposed and demonstrated via simulations.*

Keywords: Random sets, Multi-Bernoulli, Filtering, Images, Tracking, Track Before Detect.

1 Introduction

This paper investigates the problem of jointly estimating the number of objects and their states from image observations. 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. Compressing the information on the image into a finite set of points is efficient in terms of memory as well as computational requirements, and is very effective for a wide range of applications [1, 5]. However, this approach may not be adequate for applications with low signal to noise ratio as the information loss incurred in the compression becomes significant, and it is necessary to make use of all information contained in the image(s).

We formulate the multi-object estimation problem in a Bayesian framework by modelling the collection of states to be estimated as a realization of a random finite set. The solution to this problem is intractable in general, even for the special case of superpositional measurement model where the image observation is the sum of the observations generated by individual states and noise [6]. 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 non-superpositional measurement models. As an application we develop a multi-object filter suitable for applications involving image observations with low signal to noise ratio. We also present a tracking example as proof-of-concept for the proposed approach.

2 Multi-object estimation with image

Let $x_1, ..., x_n \in \mathcal{X} \subseteq R^d$ denote the state (or parameter) vectors and let $y = [y_1, ...y_m]$ 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 greyscale 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.

We start by formulating a suitable representation of the multi-object state and cast the estimation problem in a Bayesian framework in subsection 2.1. The observation model considered in this paper is then described in subsection 2.2, setting the scene for the main results in Section 3.

2.1 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 very little meaning. Simply stacking individual states into a single vector does not admit a satisfactory notion of error as illustrated in Figures 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 Figure 1, the estimate is correct but 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. $\min_{perm(X)} \|X - \hat{X}\| = 0$, 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 Figure 2? A finite set representation of the multi-object state, $X = \{x_1, ..., x_n\}$, admits a mathematically consistent notion of estimation error since distance between sets is a well understood concept.

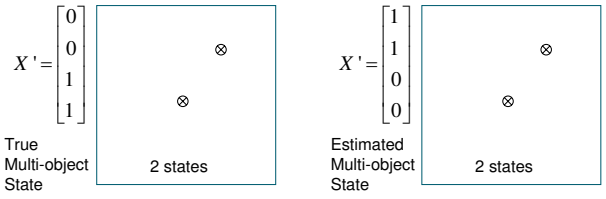

Figure 1: Hypothetical scenario showing a fundamental inconsistency with vector representations of multi-object states

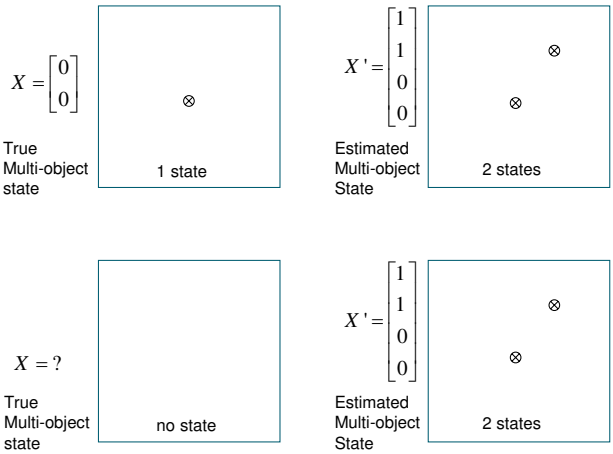

Figure 2: Hypothetical scenario showing a fundamental inconsistency with vector representations of multi-object states. How should the error be assigned when the estimated number of object is incorrect?

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 practical mathematical tools for dealing with RFSs [4, 5], based on a notion of integration and density that is consistent with point process theory [12]. In recent years FISST has generated substantial interest due to the developments of the Probability Hypothesis Density (PHD) and Cardinalized PHD filters [4],[12], [13], [8], [14].

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

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

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.

2.2 Multi-object likelihood function

The type of image observation considered in this work is illustrated in Figure 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.

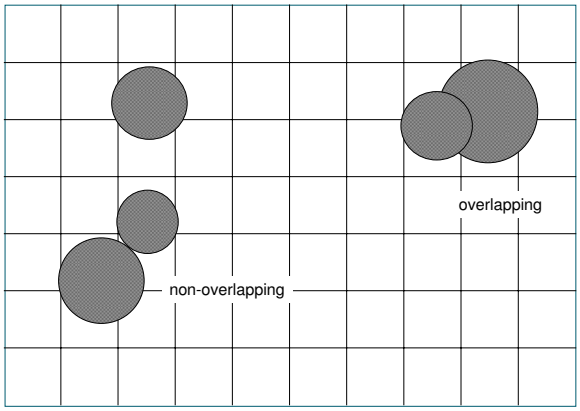

Figure 3: An illustration of overlapping and non-overlapping objects.

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 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 track-before-detect (see [10] Chapter 11),

$$\begin{aligned}\phi_i(y_i) &= \mathcal{N}(y_i;0,\sigma)\\ \varphi_i(y_i,x) &= \mathcal{N}(y_i;h_i(x),\sigma)\end{aligned}$$

where $h_i(x)$ 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 non-additive models, see for example [3].

Under the following assumptions:

- conditional 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 conditional on the multi-object state X is given by

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

where

$$\begin{aligned}g_y(x) &= \prod\nolimits_{i\in T(x)}\frac{\varphi_i(y_i,x)}{\phi_i(y_i)},\\ f(y) &= \prod\nolimits_{i=1}^{m}\phi_i(y_i).\end{aligned}$$

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

3 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 subsection 3.1 along with a summary of the mathematical tools used in this work. The main results are presented in subsection 3.2.

3.1 Probability generating functionals

Apart from the probability density, the probability generating functional (PGFl) is another fundamental descriptor of an RFS. Following [2, 4], the *probability generating functional* (PGFl) $G[\cdot]$ of an RFS X on $\mathcal{X}$ is defined by

$$G[h] \equiv E[h^X], \qquad (4)$$

where 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 *Probability Hypothesis Density* (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 [2, 4]. For an RFS X on $\mathcal{X}$, its PHD is a non-negative function v on $\mathcal{X}$ such that for each region $S \subseteq \mathcal{X}$

$$E\left[|X \cap S|\right] = \int_S v(x)dx, \qquad (5)$$

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.

It has been shown in [7] that two simple multi-object estimators based on the (posterior) PHD and cardinality distribution are Bayes optimal. In the first Bayes optimal 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 Bayes optimal estimator the same as the first except that the estimated number of states is the maximum a posteriori estimate.

The RFS pertinent to our key results and their PGFls are summarized in the following.

Poisson: A *Poisson* RFS X on $\mathcal{X}$ is one that is completely characterized by its PHD or intensity function v [2]. The cardinality (number of elements) of a Poisson RFS is Poisson with mean $\langle v,1\rangle$, where $\langle v,h\rangle$ denotes the standard inner product $\int v(x)h(x)dx$, 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

$$G[h] = e^{\langle v,h-1\rangle}, \qquad (6)$$

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 $\sum_{n=0}^{\infty} n\rho(n) = \langle v,1\rangle$ [2]. 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

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

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.

Multi-Bernoulli: A *multi-Bernoulli* RFS X on $\mathcal{X}$ is a union of a fixed number of independent RFSs $X^{(i)}$ that has probability $1-r^{(i)}$ of being empty, and probability $r^{(i)} \in (0,1)$ of being a singleton whose (only) element is distributed according to a probability density $p^{(i)}$ (defined on $\mathcal{X}$), [5]

$$X = \bigcup_{i=1}^{M} X^{(i)}. \tag{8}$$

Each $X^{(i)}$ is called a *Bernoulli* RFS. Using the independence of the $X^{(i)}$'s, the PGFl of a multi-Bernoulli RFS is given by

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

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 conditional on its existence. For convenience 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

$$v(x) = \sum_{i=1}^{M} r^{(i)} p^{(i)}(x), \tag{10}$$

and the cardinality of a multi-Bernoulli RFS is a discrete multi-Bernoulli random variable with parameters $r^{(1)}, ..., r^{(M)}$, Bayes optimal estimators based on the PHD and cardinality distribution are applicable. Moreover, a more intuitive multi-object estimator can be obtained from the existence probabilities $r^{(i)}$'s (and $p^{(i)}$'s). 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.

3.2 Closed form data-updates

We first present a result concerning the posterior PGFl for the observation likelihood considered in this work (see subsection 2.2), 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: Let p and $p(\cdot|y)$ denote the prior and posterior probability densities. Applying the definition of the PGFl and using Bayes rule to obtain the posterior probability density $p(\cdot|y)$ gives

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

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 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 [9] 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*

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

$$\rho(n|y) = \frac{\rho(n) \left(\frac{\langle v, g_y \rangle}{\langle v, 1 \rangle}\right)^n}{\sum_{j=0}^{\infty} \rho(j) \left(\frac{\langle v, g_y \rangle}{\langle v, 1 \rangle}\right)^j} \tag{12}$$

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,

$$\begin{aligned} G[h|y] &= \frac{G[hg_y]}{G[g_y]} = \frac{G\left(\frac{\langle vg_y,h\rangle}{\langle v,1\rangle}\right)}{G\left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)} \\ &= \frac{\sum_{n=0}^{\infty} \rho(n) \left(\frac{\langle vg_y,h\rangle}{\langle v,1\rangle}\right)^n}{\sum_{j=0}^{\infty} \rho(j) \left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^j} \\ &= \frac{\sum_{n=0}^{\infty} \rho(n) \left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^n \left(\frac{\langle vg_y,h\rangle}{\langle vg_y,1\rangle}\right)^n}{\sum_{j=0}^{\infty} \rho(j) \left(\frac{\langle v,g_y\rangle}{\langle v,1\rangle}\right)^j} \end{aligned} \quad (13)$$

where eq. (13) 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 (11) and cardinality distribution (12), we need to show that the posterior PGFl has the form $G[h|y] = G\left(\frac{\langle v(\cdot|y),h\rangle}{\langle v(\cdot|y),1\rangle}\middle| y\right)$. Note from (11) that $\frac{\langle vg_y,h\rangle}{\langle vg_y,1\rangle} = \frac{\langle v(\cdot|y),h\rangle}{\langle v(\cdot|y),1\rangle}$ (since the quotient of the infinite sums cancel), hence substituting for $\frac{\langle vg_y,h\rangle}{\langle vg_y,1\rangle}$ in (13) and using (12) gives

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

Therefore, the posterior is an i.i.d. cluster process with PHD (11) and cardinality distribution (12) .

Remark: The posterior PHD (11) and cardinality distribution (12) 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*

$$\left\{\left(\frac{r^{(i)}\left\langle p^{(i)}, g_y\right\rangle}{1 - r^{(i)} + r^{(i)}\left\langle p^{(i)}, g_y\right\rangle}, \frac{p^{(i)} g_y}{\left\langle p^{(i)}, g_y\right\rangle}\right)\right\}_{i=1}^{N} \quad (14)$$

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

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

The ith term in the above product is the PGFl of a Bernoulli RFS. Hence, the posterior is a multi-Bernoulli, with parameter set given by (14) .

Remark: Each of the corollaries above can be easily extended to the multiple sensor case, as long as the likelihood functions of the sensors are separable. The 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 the updates are exact, the end result is independent of the order in which the updates are done.

4 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 evolves but the number of states also evolves due to objects appearing or disappearing. Multi-object filtering involves the on-line 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 observation. The filtering formulation and the proposed multi-object filter is described in subsection 4.1 while the particle implementation is described in subsection 4.2. 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 non-linear, 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 [12]. The multi-Bernoulli approach avoids this problem altogether [15].

4.1 Multi-Bernoulli filter for image data

The multi-object filtering problem can be casted as a Bayes filter on the space of finite sets $\mathcal{F}(\mathcal{X})$. Let $\pi_k(\cdot|y_{1:k})$

denote the *multi-object posterior density* at time k. Then, the *multi-target Bayes recursion* propagates $\pi_k(\cdot|y_{1:k})$ in time [4, 5] according to the following prediction and update steps:

$$\pi_{k|k-1}(X_k|y_{1:k-1}) = \int f_{k|k-1}(X_k|X)\pi_{k-1}(X|y_{1:k-1})\delta X \quad (15)$$

$$\pi_k(X_k|y_{1:k}) = \frac{g_k(y_k|X_k)\pi_{k|k-1}(X_k|y_{1:k-1})}{\int g_k(y_k|X)\pi_{k|k-1}(X|y_{1:k-1})\delta X}, \quad (16)$$

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 behaviour 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

$$X_k = \bigcup_{x_{k-1}\in X_{k-1}} S_{k|k-1}(x_{k-1}) \cup \Gamma_k, \quad (17)$$

where Γ_k denotes the multi-Bernoulli RFS of spontaneous births. Assuming that the RFSs constituting the union in (17) are mutually independent, X_k is a multi-Bernoulli RFS conditional on X_{k-1}. Using FISST, the multi-object transition density $f_{k|k-1}(\cdot|\cdot)$ can be derived from the transition equation (17) [4, 5].

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 (15)-(16) 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 a multi-Bernoulli if $\pi_{k-1}(\cdot|y_{1:k-1})$ is a multi-Bernoulli [5]. 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 (15) and update step (16) can be approximated via the following:

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

$$\pi_{k|k-1} = \{(r_{P,k|k-1}^{(i)}, p_{P,k|k-1}^{(i)})\}_{i=1}^{M_{k-1}} \cup \{(r_{\Gamma,k}^{(i)}, p_{\Gamma,k}^{(i)})\}_{i=1}^{M_{\Gamma,k}}, \quad (18)$$

where

$$r_{P,k|k-1}^{(i)} = r_{k-1}^{(i)}\langle p_{k-1}^{(i)}, p_{S,k}\rangle, \quad (19)$$

$$p_{P,k|k-1}^{(i)}(x) = \frac{\langle f_{k|k-1}(x|\cdot), p_{k-1}^{(i)} p_{S,k}\rangle}{\langle p_{k-1}^{(i)}, p_{S,k}\rangle}, \quad (20)$$

$f_{k|k-1}(\cdot|\zeta)$ = *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,*

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

Multi-Bernoulli Update: *Given the predicted multi-Bernoulli parameters* $\{(r_{k|k-1}^{(i)}, p_{k|k-1}^{(i)}\}_{i=1}^{M_{k|k-1}}$*, the updated multi-Bernoulli parameters are*

$$\pi_k = \{(r_k^{(i)}, p_k^{(i)})\}_{i=1}^{M_{k|k-1}} \quad (21)$$

where

$$r_k^{(i)} = \frac{r_{k|k-1}^{(i)}\left\langle p_{k|k-1}^{(i)}, g_y\right\rangle}{1 - r_{k|k-1}^{(i)} + r_{k|k-1}^{(i)}\left\langle p_{k|k-1}^{(i)}, g_y\right\rangle} \quad (22)$$

$$p_k^{(i)} = \frac{p_{k|k-1}^{(i)} g_y}{\left\langle p_{k|k-1}^{(i)}, g_y\right\rangle}. \quad (23)$$

Track merging (heuristic): To account for the non-overlapping assumption, estimates that would overlap on the image observation are merged. A simple way of merging is to combine the existence probabilities $r_k^{(i)}$'s, and densities $p_k^{(i)}$'s of hypothesized objects whose estimates fall within a given distance T_{merge} of each other.

4.2 Sequential Monte Carlo Implementation

In the following, we present a generic sequential Monte Carlo (SMC) implementation of the multi-Bernoulli prediction (18) and update (21) steps.

SMC Prediction [15]: Suppose that at time $k-1$ the (multi-Bernoulli) posterior multi-object density $\pi_{k-1} = \{(r_{k-1}^{(i)}, p_{k-1}^{(i)})\}_{i=1}^{M_{k-1}}$ is given and each $p_{k-1}^{(i)}$, $i = 1, ..., M_{k-1}$, is comprised of a set of weighted samples $\{w_{k-1}^{(i,j)}, x_{k-1}^{(i,j)}\}_{j=1}^{L_{k-1}^{(i)}}$, i.e.

$$p_{k-1}^{(i)}(x) = \sum_{j=1}^{L_{k-1}^{(i)}} w_{k-1}^{(i,j)} \delta_{x_{k-1}^{(i,j)}}(x).$$

Then, given proposal densities $q_k^{(i)}(\cdot|x_{k-1}, y_k)$ and $b_k^{(i)}(\cdot|y_k)$, the predicted (multi-Bernoulli) multi-object

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

density (18) can be computed as follows

$$\begin{aligned}
r_{P,k|k-1}^{(i)} &= r_{k-1}^{(i)} \sum_{j=1}^{L_{k-1}^{(i)}} w_{k-1}^{(i,j)} p_{S,k}(x_{k-1}^{(i,j)}), \\
p_{P,k|k-1}^{(i)}(x) &= \sum_{j=1}^{L_{k-1}^{(i)}} \tilde{w}_{P,k|k-1}^{(i,j)} \delta_{x_{P,k|k-1}^{(i,j)}}(x), \\
r_{\Gamma,k}^{(i)} &= \textit{parameter given by birth model,} \\
p_{\Gamma,k}^{(i)}(x) &= \sum_{j=1}^{L_{\Gamma,k}^{(i)}} \tilde{w}_{\Gamma,k}^{(i,j)} \delta_{x_{\Gamma,k}^{(i,j)}}(x),
\end{aligned}$$

where

$$\begin{aligned}
x_{P,k|k-1}^{(i,j)} &\sim q_k^{(i)}(\cdot|x_{k-1}^{(i,j)}, y_k),\ j = 1, ..., L_{k-1}^{(i)} \\
w_{P,k|k-1}^{(i,j)} &= \frac{w_{k-1}^{(i,j)} f_{k|k-1}(x_{P,k|k-1}^{(i,j)}|x_{k-1}^{(i,j)}) p_{S,k}(x_{k-1}^{(i,j)})}{q_k^{(i)}(x_{P,k|k-1}^{(i,j)}|x_{k-1}^{(i,j)}, y_k)}, \\
\tilde{w}_{P,k|k-1}^{(i,j)} &= w_{P,k|k-1}^{(i,j)} / \sum_{j=1}^{L_{k-1}^{(i)}} w_{P,k|k-1}^{(i,j)} \\
x_{\Gamma,k}^{(i,j)} &\sim b_k^{(i)}(\cdot|y_k)\ j = 1, ..., L_{\Gamma,k}^{(i)} \\
w_{\Gamma,k}^{(i,j)} &= \frac{p_{\Gamma,k}(x_{\Gamma,k}^{(i,j)})}{b_k^{(i)}(x_{\Gamma,k}^{(i,j)}|y_k)}, \\
\tilde{w}_{\Gamma,k}^{(i,j)} &= w_{\Gamma,k}^{(i,j)} / \sum_{j=1}^{L_{\Gamma,k}^{(i)}} w_{\Gamma,k}^{(i,j)}.
\end{aligned}$$

SMC Update: Suppose that at time k the predicted (multi-Bernoulli) multi-object density $\pi_{k|k-1} = \{(r_{k|k-1}^{(i)}, p_{k|k-1}^{(i)})\}_{i=1}^{M_{k|k-1}}$ is given and each $p_{k|k-1}^{(i)}$, $i = 1, ..., M_{k|k-1}$, is comprised of a set of weighted samples $\{w_{k|k-1}^{(i,j)}, x_{k|k-1}^{(i,j)}\}_{j=1}^{L_{k-1}^{(i)}}$, i.e.

$$p_{k|k-1}^{(i)} = \sum_{j=1}^{L_{k|k-1}^{(i)}} w_{k|k-1}^{(i,j)} \delta_{x_{k|k-1}^{(i,j)}}(x).$$

Then, the updated (multi-Bernoulli) multi-object density (21) computed as follows

$$\begin{aligned}
r_k^{(i)} &= \frac{r_{k|k-1}^{(i)} \varrho_k^{(i)}}{1 - r_{k|k-1}^{(i)} + r_{k|k-1}^{(i)} \varrho_k^{(i)}} \\
p_k^{(i)} &= \frac{1}{\varrho_k^{(i)}} \sum_{j=1}^{L_{k|k-1}^{(i)}} w_{k|k-1}^{(i,j)} g_{y_k}(x_{k|k-1}^{(i,j)}) \delta_{x_{k|k-1}^{(i,j)}}(x),
\end{aligned}$$

where $\varrho_k^{(i)} = \sum_{j=1}^{L_{k|k-1}^{(i)}} w_{k|k-1}^{(i,j)} g_{y_k}(x_{k|k-1}^{(i,j)})$.

Resampling and Implementation Issues: Like the MemBer filter [15], for each hypothesized object, the particles are resampled after the update step, the number of particles is reallocated in proportion to the probability of existence, as well as restricted 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.

4.3 Simulation study

In this section, we demonstrate the performance of the filter proposed in Section 4.2 on a tracking example with a time varying number of targets observed in noise. A maximum of 10 targets appears on the scene at various instants. The true trajectories are shown in Figure 4 along with the start and stop positions of each track.

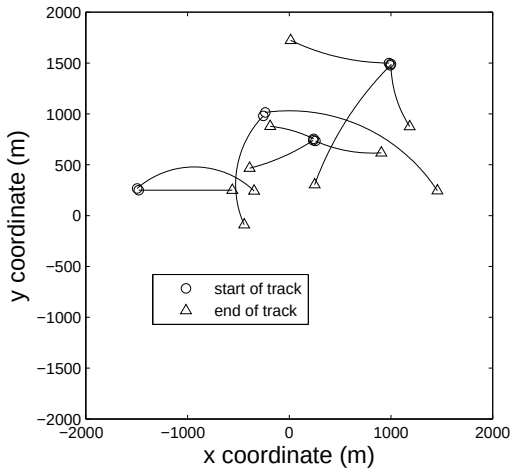

Figure 4: True target trajectories in the surveillance region.

The target state variable $x_k = [\ \tilde{x}_k^T, \omega_k\ ]^T$ comprises the planar position and velocity $\tilde{x}_k^T = [\ p_{x,k}, \dot{p}_{x,k}, p_{y,k}, \dot{p}_{y,k}\ ]^T$ as well as the turn rate ω_k. A nearly constant turn model having varying turn rate is considered:

$$\begin{aligned}
\tilde{x}_k &= F(\omega_{k-1})\tilde{x}_{k-1} + G w_{k-1} \\
\omega_k &= \omega_{k-1} + \Delta u_{k-1}
\end{aligned}$$

where

$$F(\omega) = \begin{bmatrix} 1 & \frac{\sin\omega\Delta}{\omega} & 0 & -\frac{1-\cos\omega\Delta}{\omega} \\ 0 & \cos\omega\Delta & 0 & -\sin\omega\Delta \\ 0 & \frac{1-\cos\omega\Delta}{\omega} & 1 & \frac{\sin\omega\Delta}{\omega} \\ 0 & \sin\omega\Delta & 0 & \cos\omega\Delta \end{bmatrix}, G = \begin{bmatrix} \frac{\Delta^2}{2} & 0 \\ T & 0 \\ 0 & \frac{\Delta^2}{2} \\ 0 & \Delta \end{bmatrix},$$

$w_{k-1} \sim \mathcal{N}(\cdot; 0, \sigma_w^2 I)$, $u_{k-1} \sim \mathcal{N}(\cdot; 0, \sigma_u^2 I)$, $\Delta = 1s$, $\sigma_w = 20m/s^2$, and $\sigma_u = 2\pi/180 rad/s$. The birth process is multi-Bernoulli with density $\pi_\Gamma = \{(r_\Gamma^{(i)}, p_\Gamma^{(i)})\}_{i=1}^4$ where $r_\Gamma^{(1)} = r_\Gamma^{(2)} = 0.02, r_\Gamma^{(3)} = r_\Gamma^{(4)} = 0.03$, $p_\Gamma^{(i)}(x) = \mathcal{N}(x; m_\gamma^{(i)}, P_\gamma)$, $m_\gamma^{(1)} = [\ -1500, 0, 250, 0\ 0\ ]^T$, $m_\gamma^{(2)} = [\ -250, 0, 1000, 0\ 0\ ]^T$, $m_\gamma^{(3)} = [\ 250, 0, 750, 0\ 0\ ]^T$, $m_\gamma^{(4)} = [\ 1000, 0, 1500, 0\ 0\ ]^T$, $P_\gamma = \left(0.25\text{diag}([\ 50,\ 50,\ 50,\ 50,\ 6(\pi/180)\ ]^T)\right)^2$. The probability of target survival is $p_{S,k}(x) = 0.9$.

The surveillance region is a $4000m \times 4000m$ square (see Figure 4) divided into 500×500 pixels or bins (each bin is a square of side $8m$). The observation at time k is a noisy image y_k with the ith pixel value given by

$$y_{i,k} = \begin{cases} h_{i,k}(x) + w_{i,k}, & i \in T(x) \\ w_{i,k}, & i \notin T(x) \end{cases}$$

where $T(x)$ is a 3pixel $\times$ 3pixel square with the central pixel containing the position of x, $w_{i,k} \sim \mathcal{N}(\cdot; 0, 1)$, $h_{i,k}(x) = \sqrt{2}$, giving a 3dB SNR.

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 $P = 10^{-3}$ are dropped. The merge threshold is set to $T_{merge} = 0.5$ times the pixel width.

4.3.1 Multi-object miss-distance

We use the Optimal Sub-Pattern 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 [11].

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 ($100m$ in our example). This minimum sum of distances can be interpreted as the "total localization error", which are assigned by giving the points in X the "benefit of the doubt". All other points which remain unassigned are also charged with an error value, where 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" committed 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 [11]. 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.

A formal statement of the OSPA metric is now shown. 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 $Y = \{y_1, \ldots, y_n\}$,

$$\bar{d}_p^{(c)}(X,Y) := \left(\frac{1}{n} \left(\min_{\pi \in \Pi_n} \sum_{i=1}^{m} d^{(c)}(x_i, y_{\pi(i)})^p + c^p(n-m) \right) \right)^{\frac{1}{p}} \tag{24}$$

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 p-th order per-target error, comprised of a p-th order per-target localization error and a p-th 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 [11].

4.3.2 Numerical results

Figure 5 plots the x and y components of the true trajectories, and filter estimates versus time. The plots indicate that the filter is able to identify all target births and deaths, as well as successfully estimating their states. The estimation error for the 3dB SNR scenario is shown in Figure 6 together with that of a 6dB SNR scenario ($h_{i,k}(x) = 2$). The filter shows good performance with location error in the order of half the side length of the template $T(x)$, i.e. around $12m$ for both scenarios. Although the total error is worst in the 3dB SNR scenario, the filter works surprisingly well, achieving a similar location error to the 6dB case, without dropping any tracks (see also Figure 7).

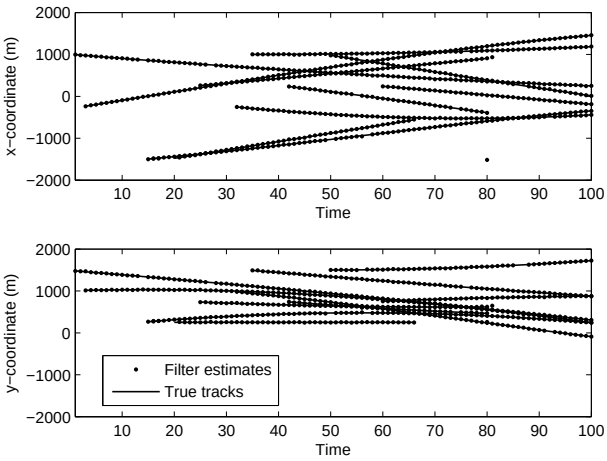

Figure 5: True tracks, and multi-object filter estimates for the scenario with 3dB SNR.

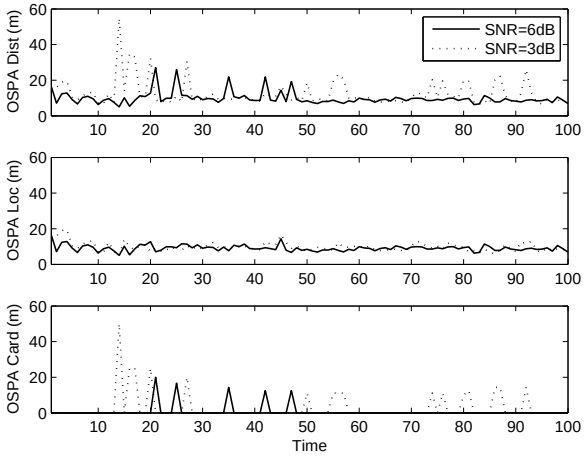

Figure 6: Multi-object estimation error for the 6dB and 3dB SNR scenarios.

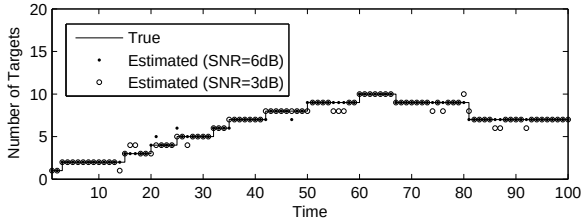

Figure 7: True, and estimated number of targets for the 6dB and 3dB SNR scenarios.

We stress that due to the assumption of non-overlapping objects, the proposed filter does not handle target crossing, and that our examples require accurate prior knowledge of birth locations. The proposed approach can theoretically accommodate more diffused birth densities. However, in implementation, many more particles are required to avoid particle collapse because g_y has a very narrow support. An auxiliary particle implementation might be more efficient.

5 Conclusion

While more comprehensive evaluations are needed, preliminary simulations have demonstrated that the proposed random finite set approach enabled a tractable solution to the multi-object estimation problem for image data. There are, however, a number of limitations of the proposed approach. The particle implementation suffers from particle degeneracy due to the reweighting of the particles by a function with very small support. More efficient particle implementation and analytic implementation are two venues for further work. 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*. Academic Press, San Diego, 1988.

[2] D. Daley and D. Vere-Jones, *An introduction to the theory of point processes*. Springer-Verlag, 1988.

[3] S. Davey, M. Rutten, and B., Cheung, "A comparison of detection performance for several Track-Before-Detect algorithms," *EURASIP Jrn. Advances in Signal Processing*, Vol. 2008 , Issue 1, Article 41, 2008.

[4] R. Mahler, "Multi-target Bayes filtering via first-order multi-target moments," *IEEE Trans. AES*, Vol. 39, No. 4, pp. 1152–1178, 2003.

[5] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House, 2007.

[6] R. Mahler, "Multitarget-moment filters for nonstandard measurement models," *Proc. SPIE Defense & Security Symposium*, Vol. 6968, 2008.

[7] R. Mahler, "Bayes-Optimality of PHD and CPHD filters, " submitted to *IEEE Trans. Signal Processing*, 2009.

[8] R. Mahler, "PHD filters of higher order in target number," *IEEE Trans. Aerospace & Electronic Systems*, vol. 43, no. 3, July 2007.

[9] N. T. Pham, H. Weimin, and S. Ong, "Tracking multiple objects using probability hypothesis density filter and color measurements," *Proc. IEEE Int. Conf. Multimedia and Expo*, pp. 1511–1514, Beijing, China, 2007.

[10] B. Ristic, S. Arulampalam, and N. J. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.

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

[12] B.-N. Vo, S. Singh and A. Doucet, "Sequential Monte Carlo methods for multi-target filtering with random finite sets", *IEEE Trans. Aerospace & Electronic Systems*, vol. 41, no. 4, pp. 1224–1245, 2005.

[13] B.-N. Vo and W.-K. Ma, "The Gaussian mixture Probability Hypothesis Density filter," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[14] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Analytic implementations of the Cardinalized Probability Hypothesis Density filter," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3553–3567, Jul. 2007.

[15] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The cardinality balanced multi-target multi-Bernoulli filter and its implementations," *IEEE Trans. Signal Processing*, vol. 57, no. 2, pp. 409–423, Feb. 2009.

Acknowledgement

This research was supported under Australian Research Council's Discovery Projects funding schemes and DP0880553 and DP0989007.