A Tutorial on Bernoulli Filters: Theory, Implementation and Applications

B. Ristic[♭], B.-T Vo[♮], B.-N. Vo[♯] and A. Farina[†]

***Abstract*—Bernoulli filters are a class of exact Bayesian filters for non-linear/non-Gaussian recursive estimation of dynamic systems, recently emerged from the random set theoretical framework. The common feature of Bernoulli filters is that they are designed for stochastic dynamic systems which randomly switch *on* and *off*. The applications are primarily in target tracking, where the switching process models target appearance or disappearance from the surveillance volume. The concept, however, is applicable to a range of dynamic phenomena, such as epidemics, pollution, social trends, etc. Bernoulli filters in general have no analytic solution and are implemented as particle filters or Gaussian sum filters. This tutorial paper reviews the theory of Bernoulli filters as well as their implementation for different measurement models. The theory is backed up by applications in sensor networks, bearings-only tracking, passive radar/sonar surveillance, visual tracking, monitoring/prediction of an epidemic and tracking using natural language statements. More advanced topics of smoothing, multi-target detection/tracking and sensor control are briefly reviewed with pointers for further reading.**

***Index Terms*—Sequential Bayesian estimation, particle filters, random sets, target tracking**

I. Introduction

This paper is devoted to estimation of the state of a dynamic stochastic system (object or phenomenon) which can randomly switch *on* and *off*. Estimation is done sequentially using prior knowledge and a sequence of observations or measurements. Adopting the commonly accepted state-space approach, the state is modelled by a state vector which contains all relevant information required to describe the system. The stochastic models of system switching (on/off), the state evolution with time, and measurements are assumed known. The measurements, which can be related to the state in a nonlinear fashion, are typically noisy, imprecise (possibly fuzzy) and ambiguous, as they may result from imperfect detection.

In the Bayesian approach to dynamic state estimation, the goal is to update on receipt of new measurements the time-varying posterior probability density function (PDF) of the state, using all information available up to that time. Since the posterior PDF embodies all available statistical information about the system, it represents the complete solution to the estimation problem. An optimal (with respect to any criterion) estimate of the state (including a measure of its accuracy), can be obtained from this PDF. The posterior PDF is constructed recursively, in two stages: the prediction and update. The conceptual solution of the recursive propagation of the posterior density forms the basis of the optimal Bayes stochastic filter. The optimal Bayes filter is well known for the so called *standard problem* [1], [2], where the stochastic dynamic system is turned *on* all the time and the measurements are affected only by randomness due to noise (with perfect detection). In the general nonlinear/non-Gaussian case, there is no analytic solution for the standard Bayes filter. The last decade and a half has witnessed great popularity of Monte Carlo based approximate solutions. The resulting *particle filters*, due to their enormous popularity, have been reviewed in several books and tutorials [3]–[8]. The driving forces behind this interest in particle filters have been the ever increasing computational power of computers and a wide range of applications, from navigation and autonomous vehicles to bio-informatics, finance and radar.

In order to deal with a wider scope of nonlinear/non-Gaussian stochastic filtering applications, with possibly multiple on/off switching systems (objects, targets) involved, where detection is imperfect (miss-detections and false alarms) and where measurements or measurement models could be imprecise, the standard Bayes filter has traditionally been used with various additional layers of logic, developed in a fairly ad-hoc manner. This is evident for example in target tracking systems [9], where: (a) the presence or absence of a target is typically established using logic-based track formation [9, Sec.13.3]; (b) the imperfect detection is dealt with using various methods of *data association* [9, Ch.6,7]; (c) imprecise measurements (e.g. attributes) are used in non-Bayesian estimation frameworks (e.g. Dempster-Shafer theory) [9, Ch.9]. While the provided techniques can be described as clever pieces of engineering solutions, their optimality is questionable, hence emphasizing the need to develop a unified theory which will provide the optimal Bayes filter formulation for the aforementioned wider scope of applications.

Sequential Bayesian estimation in conjunction with *random set theory* provides exactly that: an elegant mathematical framework in which one can formulate the optimal Bayes filter for multiple on/off switching systems, with possibly imperfect detection and imprecise measurements or measurement models [10]. While the mathematics is somewhat involved, the results are rewarding, as it will hopefully be demonstrated in this tutorial. Since the implementation of the random set formulation of the optimal Bayes filter for multiple dynamic systems is computationally very demanding (see [10]–[14]),

[♭]Corresponding author: Defence Science and Technology Organisation, ISR Division, Bld. 51, 506 Lorimer Street, Melbourne, VIC 3207, Australia; Tel: +61 3 9626 7688; Fax: +61 3 9626 7088; email: branko.ristic@dsto.defence.gov.au

[♮]B.-T. Vo is with Curtin University, Australia.

[♯]B.-N. Vo is with Curtin University, Australia.

[†]A. Farina is with SELEX Sistemi Integrati, Italy.

four classes of its simplifications or principled approximations have emerged recently: the Bernoulli filter (also known as JoTT or joint target detection and tracking filter) [10], [15]–[18], the probability hypothesis density (PHD) filter [19]–[22], the cardinalised PHD filter [23], [24] and the multi-Bernoulli filter [10], [25].

This tutorial paper covers the Bernoulli filter, its implementations and applications. The Bernoulli filter is the optimal Bayes filter for a *single* dynamic system which can randomly switch *on* or *off*. The applications so far have focused on target tracking, where the interpretation of binary switching is that targets can appear or disappear from the surveillance volume. The concept of binary switching between appearance and disappearance, however, is universal and applies to a number of different dynamic phenomena, such as epidemics, pollution, social trends etc. The key idea in dealing with random on/off switching in the Bayesian filtering framework is the introduction of the *existence* binary random variable. This concept can be traced back to [26], [27], where it was used to derive Kalman-type filters for linear/Gaussian on/off dynamic systems. The main difference between the Bernoulli random finite set formulation and the traditional approach of [26], [27] is that the underlying state is treated as a set (which can be empty or singleton) instead of a vector augmented with the binary existence variable. Bernoulli filters in a general nonlinear/non-Gaussian case have no analytic solution and are therefore implemented approximately, typically as particle filters or Gaussian sum filters.

The main feature of this tutorial is that it considers different measurement models, all of which are important in signal processing. The first model is devoted to measurements as raw intensity signals, such as acoustic or electromagnetic energy, a chemical pollution level, images, and similar. This is important in signal networks, radar/sonar or video surveillance. The jargon typically used to describe object detection and tracking with this type of measurements is *track-before-detect* [28], [29]. The second model is where the measurements are the output of a detector. In this case, one needs to deal with inevitable miss-detections and false alarms. The assumption is that the object of interest is a point and consequently at most one detection (among many) at a particular time is due to the object; the rest are false detections. Naturally which detections are false is unknown. A version of this model is when the object of interest is large with respect to the sensor resolution (non-point or extended object) and consequently can give rise to a number of detections. Again which detections are due to the object and which are false is unknown. Finally, in some situations the measurement function may not be precisely known, or the measurement itself could be imprecise and fuzzy. These type of measurements, referred to as *non-standard* measurements, can be incorporated into the Bayesin estimation framework within the random set theoretical framework. The paper derives the mathematical formulation of each respective Bernoulli filter and for each measurement model.

The tutorial is organised as follows. Sec. II introduces the preliminaries: the recursive equations of the standard optimal Bayes filter and the mathematics of random finite sets. This level of mathematics will be adequate to follow all derivations in the paper. Sec. III describes the stochastic model for binary on/off switching dynamic systems, followed by the prediction equations of the Bernoulli filter. Secs. IV through VII present the update equations of the Bernoulli filter for different measurement models. The model of intensity measurements is considered in Sec. IV. Secs. V and VI present the detector-output measurement model for a point target and an extended target, respectively. Measurement models for imprecise measurements, measurement functions and uncertain implication rules are described in Sec. VII. Two approximate implementations of the Bernoulli filter are presented in Sec.VIII, the Bernoulli particle filter and the Bernoulli Gaussian sum filter. Applications, involving different measurement models, are presented in Sec. IX. They include: sensor networks, bearings-only tracking, visual tracking, monitoring/prediction of an epidemic and tracking using natural language statements. More advanced topics for further research, including multi-sensor distributed Bernoulli filters, smoothing, model parameter estimation, multi-target detection/tracking and sensor control, are briefly reviewed in Sec. X. The tutorial is summarised in Sec.XI.

II. Notation and Preliminaries

A. Standard stochastic Bayes filter

The roots of stochastic filtering theory can be traced back to the early 1960s. Kalman and Bucy [30], [31] formulated the linear filtering theory, while Stratonovich [32] and Kushner [33] pioneered the development of the probabilistic approach to nonlinear filtering.

The discrete-time formulation of the stochastic filtering problem in the Bayesian framework is as follows [1]. Suppose the state vector $\mathbf{x}_k \in \mathcal{X}$ provides the complete specification of the state of a dynamic system (object, phenomenon) at time t_k. Here $\mathcal{X} \subseteq \mathbb{R}^{n_x}$ is the state space, while k is the discrete-time index corresponding to t_k. The stochastic dynamic system is described by two equations:

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{v}_{k-1}, \tag{1}$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{w}_k, \tag{2}$$

referred to as the *dynamics equation* and the *measurement equation*, respectively. The function $\mathbf{f}_{k-1} : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ is a nonlinear transition function defining the evolution of the state vector as a first-order Markov process. The random process $\mathbf{v}_k \in \mathbb{R}^{n_x}$ is independent identically distributed (IID) according to the PDF $p_{\mathbf{v}}$; $\mathbf{v}_k$ is referred to as *process noise*, and its role is to model random disturbances in state evolution. The dimension of the state vector and process noise is $n_x \in \mathbb{N}$. The function $\mathbf{h}_k : \mathbb{R}^{n_x} \to \mathbb{R}^{n_z}$ defines the relationship between the state $\mathbf{x}_k$ and the measurement $\mathbf{z}_k \in \mathcal{Z}$, where $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$ is the measurement space. The random process $\mathbf{w}_k \in \mathbb{R}^{n_z}$, independent of $\mathbf{v}_k$, is also IID with PDF $p_{\mathbf{w}}$, and referred to as *measurement noise*; n_z is the dimensions of the measurement vector.

In the formulation specified by (1)-(2), functions $\mathbf{f}_k$ and $\mathbf{h}_k$, as well as PDFs $p_{\mathbf{v}}$ and $p_{\mathbf{w}}$ are known. Equations (1) and (2) effectively define two probability functions, the *transitional*

density $\pi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) = p_{\mathbf{v}}(\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}_{k-1}))$ and the likelihood function $g_k(\mathbf{z}_k|\mathbf{x}_k) = p_{\mathbf{w}}(\mathbf{z}_k - \mathbf{h}_k(\mathbf{x}_k))$. The goal of stochastic Bayesian filtering is to estimate recursively the posterior PDF of the state, denoted $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, where $\mathbf{z}_{1:k} \equiv \mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_k$.

Assuming the initial density of the state, $p(\mathbf{x}_0)$ is known, the solution is usually presented as a two step procedure. Let $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ denote the posterior PDF at $k-1$. The first step *predicts* the density of the state to time k (when measurement $\mathbf{z}_k$ is available) via the Chapman - Kolmogorov equation [1]:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int \pi_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1}. \tag{3}$$

The second step applies the Bayes rule to *update* the predicted PDF using measurement $\mathbf{z}_k$:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{g_k(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{\int g_k(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) d\mathbf{x}_k}. \tag{4}$$

Knowing the posterior $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, one can compute point estimates of the state, e.g. the expected a posterior (EAP) estimate or the maximum a posterior (MAP) estimate.

B. Random finite sets

A random finite set (RFS) is a random variable that takes values as unordered finite sets. The cardinality of a RFS $\mathbf{X}$ is random and modelled by a discrete distribution $\rho(n) = P\{|\mathbf{X}| = n\}$, where[1] $n \in \mathbb{N}_0$. A RFS $\mathbf{X}$ is completely specified by its cardinality distribution $\rho(n)$ and a family of symmetric joint distributions[2] $p_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $n \in \mathbb{N}_0$, $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{X}$, that characterise the distribution of its elements over the state space $\mathcal{X}$, conditioned on cardinality n.

Since a RFS is nothing but a finite-set valued random variable, the usual probabilistic descriptors of a random variable, such as the PDF and its moments, can be defined for any RFS. Due to its convenience, we adopt Mahler's [10] approach, referred to *finite set statistics (FISST)*. The FISST PDF[3] of a RFS variable $\mathbf{X}$ is denoted by $f(\mathbf{X})$. This PDF is uniquely determined by $\rho(n)$ and $p_n(\mathbf{x}, \ldots, \mathbf{x}_n)$ as follows [10]:

$$f(\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}) = n! \cdot \rho(n) \cdot p_n(\mathbf{x}_1, \ldots, \mathbf{x}_n) \tag{5}$$

for $n \in \mathbb{N}_0$. Using the set integral, defined as

$$\int f(\mathbf{X})\, \delta\mathbf{X} = f(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int f(\{\mathbf{x}_1, \ldots, \mathbf{x}_n\})\, d\mathbf{x}_1 \ldots d\mathbf{x}_n. \tag{6}$$

it is straightforward to show that $f(\mathbf{X})$ of (5) integrates to one (as it should, being a PDF).

The following RFSs are relevant for this paper.

a) Bernoulli RFS: The cardinality distribution $\rho(n)$ of this RFS is Bernoulli. Thus the Bernoulli RFS can either be empty (with probability $1-q$) or have one element (with probability q), distributed over the state space $\mathcal{X}$ according to PDF $p(\mathbf{x})$. The FISST PDF of a Bernoulli RFS $\mathbf{X}$ is given by:

$$f(\mathbf{X}) = \begin{cases} 1-q, & \text{if } \mathbf{X} = \emptyset \\ q \cdot p(\mathbf{x}), & \text{if } \mathbf{X} = \{\mathbf{x}\}. \end{cases} \tag{7}$$

b) IID cluster RFS: Given cardinality $|\mathbf{X}|$, the elements of IID cluster RFS $\mathbf{X}$ are each independent identically distributed (IID) random variables distributed according to PDF $p(\mathbf{x})$ on $\mathcal{X}$. The FISST PDF of $\mathbf{X}$ is:

$$f(\mathbf{X}) = |\mathbf{X}|! \cdot \rho(|\mathbf{X}|) \prod_{\mathbf{x} \in \mathbf{X}} p(\mathbf{x}) \tag{8}$$

Compare (8) with (5). Due to the IID property, the symmetric joint distribution $p_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ in (5) is replaced with the product $\prod_{\mathbf{x} \in \mathbf{X}} p(\mathbf{x})$ in (8).

c) Poisson RFS: A Poisson RFS $\mathbf{X}$ is a special case of the IID cluster RFS, whose cardinality distribution is Poisson, that is:

$$\rho(n) = \frac{e^{-\lambda}\lambda^n}{n!}, \qquad n = 0, 1, 2, \ldots$$

According to (8), its FISST PDF is given by:

$$f(\mathbf{X}) = e^{-\lambda} \prod_{\mathbf{x} \in \mathbf{X}} \lambda\, p(\mathbf{x}). \tag{9}$$

d) Binomial RFS: A Binomial RFS[4] $\mathbf{X}$ is also a special case of the IID cluster RFS, whose cardinality distribution is a binomial distribution with parameters L (number of binary experiments) and q (the probability of success of each of the experiments):

$$\rho(n) = \binom{L}{n} q^n (1-q)^{L-n}, \qquad n = 0, 1, 2, \ldots, L.$$

Its FISST PDF is then:

$$f(\mathbf{X}) = \frac{L!}{(L-|\mathbf{X}|)!}\, q^{|\mathbf{X}|} (1-q)^{L-|\mathbf{X}|} \prod_{\mathbf{x} \in \mathbf{X}} p(\mathbf{x}). \tag{10}$$

Note that if $L = 1$, the Binomial RFS reduces to the Bernoulli RFS.

Suppose at discrete-time $k = 0, 1, 2 \ldots$, there are $n_k \in \mathbb{N}_0$ objects with states $\mathbf{x}_{k,1}, \ldots, \mathbf{x}_{k,n_k}$, taking values in the state space $\mathcal{X} \subseteq \mathbb{R}^{n_x}$. Both the number of objects n_k and their individual states in $\mathcal{X}$ are random and time-varying. The multi-object state at k, represented by a finite set $\mathbf{X}_k = \{\mathbf{x}_{k,1}, \ldots, \mathbf{x}_{k,n_k}\} \in \mathcal{F}(\mathcal{X})$, can conveniently be modelled as a RFS on $\mathcal{X}$. Here $\mathcal{F}(\mathcal{X})$ is a set of all finite subsets of $\mathcal{X}$. Let us assume that multi-object state is a Markov process with transitional density $\phi_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})$.

Let the measurement of multi-object state $\mathbf{X}_k$ be denoted $\boldsymbol{\Upsilon}_k$. This is a general notation where $\boldsymbol{\Upsilon}_k$ can represent a vector measurement $\mathbf{z}_k$, a random finite set measurement $\mathbf{Z}_k$, or a closed set $\boldsymbol{\zeta}_k$ (a crisp or fuzzy interval), depending on the measurement model we adopt. Measurement models will be

[1] $\mathbb{N}_0$ denotes the set of natural numbers including zero.

[2] A joint distribution function $p_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is said to be symmetric if its value remains unchanged for all of the $n!$ possible permutations of its variables.

[3] While the FISST densities are not probability densities, they have been shown to be equivalent to probability densities relative to some reference measure [13]. Subsequently, we do not distinguish between FISST densities and probability densities of random finite sets.

[4] Note that a binomial point process, introduced in [34], is a completely different concept, with a fixed cardinality.

presented later in the article. Given $\mathbf{X}_k$, measurement $\boldsymbol{\Upsilon}_k$, is assumed to be statistically independent of $\boldsymbol{\Upsilon}_\ell$, where $\ell \neq k$. Let $\varphi_k(\boldsymbol{\Upsilon}_k|\mathbf{X}_k)$ denote the likelihood function of $\boldsymbol{\Upsilon}_k$.

The stochastic filtering problem can now be cast in the random finite set framework. Suppose that at time $k-1$ the posterior FISST PDF of multi-object state $f_{k-1|k-1}(\mathbf{X}_{k-1}|\boldsymbol{\Upsilon}_{1:k-1})$ is known. Here $\boldsymbol{\Upsilon}_{1:k-1} \equiv \boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_{k-1}$ is the sequence of all previous measurements. Then the predicted and updated multi-object posterior densities can be expressed as follows [10]:

$$f_{k|k-1}(\mathbf{X}_k|\boldsymbol{\Upsilon}_{1:k-1}) = \int \phi_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1}) f_{k-1|k-1}(\mathbf{X}_{k-1}|\boldsymbol{\Upsilon}_{1:k-1}) \delta\mathbf{X}_{k-1} \tag{11}$$

$$f_{k|k}(\mathbf{X}_k|\boldsymbol{\Upsilon}_{1:k}) = \frac{\varphi_k(\boldsymbol{\Upsilon}_k|\mathbf{X}_k) f_{k|k-1}(\mathbf{X}_k|\boldsymbol{\Upsilon}_{1:k-1})}{\int \varphi_k(\boldsymbol{\Upsilon}_k|\mathbf{X}) f_{k|k-1}(\mathbf{X}|\boldsymbol{\Upsilon}_{1:k-1}) \delta\mathbf{X}}. \tag{12}$$

The recursion (11)-(12) is a non-trivial generalisation of (3)-(4), because the integrals in (11)-(12) are set integrals and the expressions for $\phi_{k|k-1}(\mathbf{X}_k|\mathbf{X}_{k-1})$ and $\varphi_k(\boldsymbol{\Upsilon}_k|\mathbf{X}_k)$ can be quite involved.

III. The Stochastic Dynamic Model and Prediction

A. The stochastic dynamic model

The common feature of all Bernoulli filters is the model of object (system, target, phenomenon) dynamics. The object, however, may or may not be present in the scene (surveillance region) at a particular time. We therefore model the object state at discrete-time k by the Bernoulli RFS; its state space is $\emptyset \cup \mathcal{S}(\mathcal{X})$, where $\mathcal{S}(\mathcal{X})$ is a set of all singletons[5] $\{\mathbf{x}\}$ such that $\mathbf{x} \in \mathcal{X}$. According to (7), the probabilistic description of a Bernoulli RFS $\mathbf{X}$ is completely specified by the probability q of being a singleton and the PDF $p(\mathbf{x})$, defined on $\mathcal{X}$. In order to verify that this PDF integrates to 1, we need to apply the set integral defined by (6). In the special case where $f(\mathbf{X}) = 0$ for cardinality $|\mathbf{X}| \geq 2$, the set integral simplifies to:

$$\int f(\mathbf{X})\, \delta\mathbf{X} = f(\emptyset) + \int f(\{\mathbf{x}\})\, d\mathbf{x}. \tag{13}$$

Now using (7) we have

$$\int f(\mathbf{X})\, \delta\mathbf{X} = 1 - q + q \int p(\mathbf{x})\, d\mathbf{x} = 1 \tag{14}$$

because $p(\mathbf{x})$ is the (conventional) PDF on $\mathcal{X}$ and hence integrates to 1.

If the object is present (i.e. $\mathbf{X}_k$ is a singleton), then it is assumed that it is a Markov process with a known transitional density $\pi_{k|k-1}(\mathbf{x}|\mathbf{x}')$ during the sampling interval $T_k = t_k - t_{k-1}$. In order to model object appearance and disappearance during the observation period, it is convenient to introduce a binary random variable $\epsilon_k \in \{0, 1\}$ referred to as the *existence*. The convention is that $\epsilon_k = 1$ means that object exists at discrete-time k. Dynamics of ϵ_k is modelled by the first-order two-state Markov chain with a transitional probability matrix (TPM) Π. The elements of the TPM are defined as $[\Pi]_{ij} = P\{\epsilon_k = j-1 | \epsilon_{k-1} = i-1\}$ for $i, j \in \{1, 2\}$. The TPM is adopted as follows:

$$\Pi = \begin{bmatrix} (1-p_b) & p_b \\ (1-p_s) & p_s \end{bmatrix} \tag{15}$$

where $p_b = P\{\epsilon_{k+1} = 1 | \epsilon_k = 0\}$ is the probability of object "birth" during the sampling interval, and $p_s = P\{\epsilon_{k+1} = 1 | \epsilon_k = 1\}$ the probability of target "survival" during the sampling interval. These two probabilities, together with the initial target existence probability $q_0 = P\{\epsilon_0 = 1\}$, are assumed known. If the object appears during the sampling interval T_k, the PDF $b_{k|k-1}(\mathbf{x})$ denotes its birth density.

In summary, the dynamics of the Bernoulli Markov process $\mathbf{X}_k$, for $k = 1, 2, \ldots$, is characterised by the transitional FISST PDF $\phi_{k|k-1}(\mathbf{X}|\mathbf{X}')$, from RFS $\mathbf{X}'$ at $k-1$ to RFS $\mathbf{X}$ at k, specified as follows:

$$\phi_{k|k-1}(\mathbf{X}|\emptyset) = \begin{cases} 1 - p_b & \text{if } \mathbf{X} = \emptyset \\ p_b \cdot b_{k|k-1}(\mathbf{x}) & \text{if } \mathbf{X} = \{\mathbf{x}\} \\ 0 & \text{if } |\mathbf{X}| \geq 2 \end{cases}, \tag{16}$$

$$\phi_{k|k-1}(\mathbf{X}|\{\mathbf{x}'\}) = \begin{cases} 1 - p_s & \text{if } \mathbf{X} = \emptyset \\ p_s \cdot \pi_{k|k-1}(\mathbf{x}|\mathbf{x}') & \text{if } \mathbf{X} = \{\mathbf{x}\} \\ 0 & \text{if } |\mathbf{X}| \geq 2 \end{cases}$$

Now we are in the position to derive the prediction equations of the Bernoulli filter.

B. Prediction equations of the Bernoulli filter

The Bernoulli filter, as a sequential Bayesian estimator, estimates recursively the posterior PDF of object state through the prediction and update stages, using the dynamic and measurement models and received measurements. The posterior FISST PDF at time k, denoted $f_{k|k}(\mathbf{X}_k|\boldsymbol{\Upsilon}_{1:k})$, for a Bernoulli RFS has the form (7) and is therefore completely specified by two quantities:

- the posterior probability of object existence $q_{k|k} = P\{|\mathbf{X}_k| = 1 \mid \boldsymbol{\Upsilon}_{1:k}\}$;
- the posterior spatial PDF of $\mathbf{X}_k = \{\mathbf{x}\}$, that is $s_{k|k}(\mathbf{x}) = p(\mathbf{x}_k|\boldsymbol{\Upsilon}_{1:k})$.

The Bernoulli filter therefore needs to propagate only these two quantities over time.

The prediction equations of the Bernoulli filter have been originally derived in [10, App.G.23] and [15, Sec.3.6]. Let the posterior FISST PDF of a Bernoulli RFS at time $k-1$ be $f_{k-1|k-1}(\mathbf{X}_{k-1}|\boldsymbol{\Upsilon}_{1:k-1})$. The prediction equation of the Bayes filter in the RFS framework was given by (11), that is:

$$f_{k|k-1}(\mathbf{X}|\boldsymbol{\Upsilon}_{1:k-1}) = \int \phi_{k|k-1}(\mathbf{X}|\mathbf{X}') f_{k-1|k-1}(\mathbf{X}'|\boldsymbol{\Upsilon}_{1:k-1})\, \delta\mathbf{X}' \tag{17}$$

$$= \phi_{k|k-1}(\mathbf{X}|\emptyset) f_{k-1|k-1}(\emptyset|\boldsymbol{\Upsilon}_{1:k-1}) + \int \phi_{k|k-1}(\mathbf{X}|\{\mathbf{x}'\}) f_{k-1|k-1}(\{\mathbf{x}'\}|\boldsymbol{\Upsilon}_{1:k-1})\, d\mathbf{x}' \tag{18}$$

[5] A singleton is a set whose cardinality is 1.

Equation (18) follows from (17) using (13). Let us first solve (18) for the case $\mathbf{X} = \emptyset$. Using (16) we have:

$$f_{k|k-1}(\emptyset|\boldsymbol{\Upsilon}_{1:k-1}) = \phi_{k|k-1}(\emptyset|\emptyset)\, f_{k-1|k-1}(\emptyset|\boldsymbol{\Upsilon}_{1:k-1}) + \int \phi_{k|k-1}(\emptyset|\{\mathbf{x}'\}) f_{k-1|k-1}(\{\mathbf{x}'\}|\boldsymbol{\Upsilon}_{1:k-1})\, d\mathbf{x}' \quad (19)$$

$$= (1-p_b) f_{k-1|k-1}(\emptyset|\boldsymbol{\Upsilon}_{1:k-1}) + \int (1-p_s) f_{k-1|k-1}(\mathbf{x}'|\boldsymbol{\Upsilon}_{1:k-1})\, d\mathbf{x}' \quad (20)$$

$$= (1-p_b)(1-q_{k-1|k-1}) + (1-p_s) q_{k-1|k-1} \int s_{k-1|k-1}(\mathbf{x}')\, d\mathbf{x}' \quad (21)$$

$$= (1-p_b)(1-q_{k-1|k-1}) + (1-p_s) q_{k-1|k-1} \quad (22)$$

Next we solve (18) for the case $\mathbf{X} = \{\mathbf{x}\}$. Using (16) we have:

$$f_{k|k-1}(\{\mathbf{x}\}|\boldsymbol{\Upsilon}_{1:k-1}) = \phi_{k|k-1}(\{\mathbf{x}\}|\emptyset)\, f_{k-1|k-1}(\emptyset|\boldsymbol{\Upsilon}_{1:k-1}) + \int \phi_{k|k-1}(\{\mathbf{x}\}|\{\mathbf{x}'\}) f_{k-1|k-1}(\mathbf{x}'|\boldsymbol{\Upsilon}_{1:k-1})\, d\mathbf{x}' \quad (23)$$

$$= b_{k|k-1}(\mathbf{x}) p_b (1-q_{k-1|k-1}) + \int p_s \pi_{k|k-1}(\mathbf{x}|\mathbf{x}') q_{k-1|k-1} s_{k-1|k-1}(\mathbf{x}')\, d\mathbf{x}' \quad (24)$$

$$= p_b (1-q_{k-1|k-1}) b_{k|k-1}(\mathbf{x}) + p_s q_{k-1|k-1} \int \pi_{k|k-1}(\mathbf{x}|\mathbf{x}') s_{k-1|k-1}(\mathbf{x}')\, d\mathbf{x}' \quad (25)$$

Additionally we have from (18) and (16) that:

$$f_{k|k-1}(\mathbf{X}|\boldsymbol{\Upsilon}_{1:k-1}) = 0 \quad \text{if } |\mathbf{X}| \geq 2\ . \quad (26)$$

Hence we can easily verify that

$$\int f_{k|k-1}(\mathbf{X}|\boldsymbol{\Upsilon}_{1:k-1})\, \delta\mathbf{X} = 1,$$

and by comparing with (7) we have established that $f_{k|k-1}(\mathbf{X}|\boldsymbol{\Upsilon}_{1:k-1})$ is a FISST probability density for a Bernoulli RFS.

Next we solve for $q_{k|k-1}$ and $s_{k|k-1}(\mathbf{x})$ of $f_{k|k-1}(\mathbf{X}|\boldsymbol{\Upsilon}_{1:k-1})$. Since the predicted FISST PDF is in the form (7), then the left-hand side of (22) equals $1 - q_{k|k-1}$ and we can write (22) as:

$$1 - q_{k|k-1} = (1-p_b)(1-q_{k-1|k-1}) + (1-p_s) q_{k-1|k-1}. \quad (27)$$

This leads to the prediction equation for the probability of existence:

$$q_{k|k-1} = p_b\,(1-q_{k-1|k-1}) + p_s\, q_{k-1|k-1}. \quad (28)$$

Notice that (28) has an intuitive interpretation. Loosely speaking, it states that a predicted target can arise from a new birth or an existing survival. The predicted existence probability consist of two additive terms: the probability of target non-existence and target birth, and the probability of target existence and survival.

Similarly, since the predicted FISST PDF is in the form of (7), the left-hand side of (25) equals $q_{k|k-1} s_{k|k-1}(\mathbf{x})$. This leads to the update equation for the spatial PDF of the Bernoulli filter:

$$s_{k|k-1}(\mathbf{x}) = \frac{p_b\,(1-q_{k-1|k-1})\, b_{k|k-1}(\mathbf{x})}{q_{k|k-1}} + \frac{p_s\, q_{k-1|k-1} \int \pi_{k|k-1}(\mathbf{x}|\mathbf{x}')\, s_{k-1|k-1}(\mathbf{x}')d\mathbf{x}'}{q_{k|k-1}}. \quad (29)$$

Notice also that (29) has an intuitive interpretation. Loosely speaking, it states that the probability density of the predicted state comprises a birth component and a surviving component. The birth component is the birth density weighted by the probability of target non-existence and new birth. The surviving component comprises the standard Chapman-Kolmogorov prediction weighted by the probability of target existence and survival.

Equations (28) and (29) fully specify the prediction step of the Bernoulli filter. Note that if $p_b = 0$, and $q_{k-1|k-1} = 1$, then from (28) $q_{k|k-1} = 1$ while (29) reduces to $s_{k|k-1}(\mathbf{x}) = \int \pi_{k|k-1}(\mathbf{x}|\mathbf{x}')\, s_{k-1|k-1}(\mathbf{x}')d\mathbf{x}'$, which is the standard Chapman-Kolmogorov prediction given by (3).

In the next three sections we will present the update equations of the Bernoulli filter for different measurement models.

IV. Intensity Measurements

A. Intensity measurement model

Suppose the sensor at our disposal consist of $n \geq 1$ intensity measuring elements at known locations, each reporting instantaneously at time k a measured value $z_k^{(s)}$, $s = 1, \ldots, n$. In this (fairly general) formulation, an element can be one of the following.

a) A pixels of an image: For a monochromatic image, the value in pixel i can be a measure of reflective light or temperature. For a colour image, it can be a measure of the similarity between the colour histogram computed in a box centered at pixel i and a reference histogram [35].

b) A bin in a range-Doppler-azimuth map: This corresponds to the radar context [36], where intensity of a bin refers to the electromagnetic energy gathered in a resolution volume of the corresponding range-Doppler-azimuth cell.

c) A node in a sensor network: The assumption is that all sensor nodes sample the environment synchronously. Intensity in this context depends on the purpose of the sensor network - it can refer to the level of pollution (for a chemical sensor network), radiation (for a network of Geiger-Müller counters), or acoustic energy (for an array of microphones). Note that the placement of nodes in the sensor network can be arbitrary, as long as their locations are known.

We can stack all n measurements into a single measurement vector collected at time t_k:

$$\mathbf{z}_k = \begin{bmatrix} z_k^{(1)} & z_k^{(2)} & \cdots & z_k^{(n)} \end{bmatrix}^{\mathsf{T}}, \quad (30)$$

and express the measurement equation as in (2). The measured intensity in element i can be modelled as a function of the Bernoulli state $\mathbf{X}_k$ as follows:

$$z_k^{(s)} = \begin{cases} h_k^{(s)}(\mathbf{x}_k) + w_k^{(s)} & \text{if } \mathbf{X}_k = \{\mathbf{x}_k\} \\ w_k^{(s)} & \text{if } \mathbf{X}_k = \emptyset \end{cases} \quad (31)$$

where $h_k^{(s)}(\mathbf{x}_k)$ is the contribution to intensity in element $s = 1, \ldots, n$ from an object in the state $\mathbf{x}_k$, and $w_k^{(s)}$ is the background noise in element s. Let us assume that the background noise in element s is independent of noise in other elements, and identically distributed according to PDF $g_0^{(s)}$. The measurement function $h_k^{(s)}(\mathbf{x}_k)$ can take the form of a Gaussian point spread function [6, Ch.11], the inverse distance squared law [37] or an ambiguity function [38]. Let the measurement likelihood in element s in the presence of an object in the state $\mathbf{x}_k$ be denoted $g_1^{(s)}$. The likelihood function of the measurement vector $\mathbf{z}_k$, as a function of the Bernoulli state $\mathbf{X}_k$, can then be expressed as[6]:

$$\varphi_k(\mathbf{z}_k|\mathbf{X}_k) = \begin{cases} \prod_{s=1}^{n} g_1^{(s)}\left(z_k^{(s)}|\mathbf{x}\right) & \text{if } \mathbf{X}_k = \{\mathbf{x}\} \\ \prod_{s=1}^{n} g_0^{(s)}\left(z_k^{(s)}\right), & \text{if } \mathbf{X}_k = \emptyset. \end{cases} \tag{32}$$

The intensity measurement model does not need to distinguish between small and large objects. A large object is typically sensed by many elements (pixels or resolution cells) due to its size. However, even a physically small object can be sensed by many elements of the intensity sensor, due to the spread generated by the measurement function $h_k^{(s)}(\mathbf{x}_k)$. The distinction between small and large objects should be reflected only in the choice of the state vector $\mathbf{x}$; thus for large objects it may be necessary to include the shape/size in the state.

B. Update equations

The update equations of the Bernoulli filter for the intensity measurement model were derived using FISST in [29]. An attempt to derive these equations using the standard non-FISST approach, resulted in very complicated, almost untractable mathematics, see [39].

Recall from Sec.III-B that the output of the prediction step is the FISST PDF $f_{k|k-1}(\mathbf{X}|\boldsymbol{\Upsilon}_{1:k-1})$, which is uniquely specified by the pair $\left(q_{k|k-1}, s_{k|k-1}(\mathbf{x})\right)$ and takes the form (7). Also note that for the intensity measurement model, $\boldsymbol{\Upsilon}_k \equiv \mathbf{z}_k$. The updated FISST PDF given by (12) follows from the Bayes rule:

$$f_{k|k}(\mathbf{X}_k|\mathbf{z}_{1:k}) = \frac{\varphi_k(\mathbf{z}_k|\mathbf{X}_k) \cdot f_{k|k-1}(\mathbf{X}_k|\mathbf{z}_{1:k-1})}{p_k(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{33}$$

where

$$\begin{aligned} p_k(\mathbf{z}_k|\mathbf{z}_{1:k-1}) &= \int \varphi_k(\mathbf{z}_k|\mathbf{X}) \cdot f_{k|k-1}(\mathbf{X}|\mathbf{z}_{1:k-1})\, \delta\mathbf{X} && (34) \\ &= \varphi_k(\mathbf{z}_k|\emptyset)\, f_{k|k-1}(\emptyset|\mathbf{z}_{1:k-1}) + \\ &\quad \int \varphi_k(\mathbf{z}_k|\{\mathbf{x}\})\, f_{k|k-1}(\{\mathbf{x}\}|\mathbf{z}_{1:k-1})\, d\mathbf{x} && (35) \\ &= (1 - q_{k|k-1})\, \varphi_k(\mathbf{z}_k|\emptyset) + \\ &\quad q_{k|k-1} \int \varphi_k(\mathbf{z}_k|\{\mathbf{x}\})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x} && (36) \end{aligned}$$

[6]In applications where intensity measurements are quantised, the expression for $\varphi_k(\mathbf{z}_k|\mathbf{X}_k)$ will involve the integrals (with quantisation thresholds as limits) of likelihoods $g_1^{(s)}$ and $g_0^{(s)}$.

Eq.(35) results from the application of (13). Since $f_{k|k-1}(\mathbf{X}_k|\mathbf{z}_{1:k-1}) = 0$ for $|\mathbf{X}| \geq 2$, we have that

$$f_{k|k}(\mathbf{X}|\mathbf{z}_{1:k}) = 0 \quad \text{if } |\mathbf{X}| \geq 2$$

and obviously from (33)

$$\int f_{k|k}(\mathbf{X}|\mathbf{z}_{1:k})\, \delta\mathbf{X} = 1,$$

from which by comparison with (7) we have established that $f_{k|k}(\mathbf{X}_k|\mathbf{z}_{1:k})$ is a FISST probability density for a Bernoulli RFS.

The update step (33) can be expressed for the case $\mathbf{X}_k = \emptyset$ as:

$$f_{k|k}(\emptyset|\mathbf{z}_{1:k}) = \frac{\varphi_k(\mathbf{z}_k|\emptyset) \cdot f_{k|k-1}(\emptyset|\mathbf{z}_{1:k-1})}{(1 - q_{k|k-1})\, \varphi_k(\mathbf{z}_k|\emptyset) + q_{k|k-1} \int \varphi_k(\mathbf{z}_k|\{\mathbf{x}\})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x}} \tag{37}$$

which is equivalent to

$$1 - q_{k|k} = \frac{1 - q_{k|k-1}}{1 - q_{k|k-1} + q_{k|k-1} \int \ell_k(\mathbf{z}_k|\mathbf{x})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x}} \tag{38}$$

where

$$\ell_k(\mathbf{z}_k|\mathbf{x}) = \frac{\varphi_k(\mathbf{z}_k|\{\mathbf{x}\})}{\varphi_k(\mathbf{z}_k|\emptyset)} = \prod_{s=1}^{n} \frac{g_1^{(s)}\left(z_k^{(s)}|\mathbf{x}\right)}{g_0^{(s)}\left(z_k^{(s)}\right)} \tag{39}$$

is the measurement likelihood ratio. From (38) follows the update equation for the probability of existence:

$$q_{k|k} = \frac{q_{k|k-1} \int \ell_k(\mathbf{z}_k|\mathbf{x})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x}}{1 - q_{k|k-1} + q_{k|k-1} \int \ell_k(\mathbf{z}_k|\mathbf{x})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x}} \tag{40}$$

Eq. (40) is effectively a Bayesian update for the prior existence probability.

The update equation for the spatial PDF can be derived from (33) for the case $\mathbf{X}_k = \{\mathbf{x}\}$ using the fact that $f_{k|k}(\{\mathbf{x}\}|\mathbf{z}_{1:k}) = q_{k|k}\, s_{k|k}(\mathbf{x})$. First we have:

$$q_{k|k}\, s_{k|k}(\mathbf{x}) = \frac{q_{k|k-1} \ell_k(\mathbf{z}_k|\mathbf{x}) s_{k|k-1}(\mathbf{x})}{(1 - q_{k|k-1} + q_{k|k-1} \int \ell_k(\mathbf{z}_k|\mathbf{x}) s_{k|k-1}(\mathbf{x})\, d\mathbf{x}} \tag{41}$$

which using (40) and (39) leads to:

$$s_{k|k}(\mathbf{x}) = \frac{\varphi_k(\mathbf{z}_k|\{\mathbf{x}\})\, s_{k|k-1}(\mathbf{x})}{\int \varphi_k(\mathbf{z}_k|\{\mathbf{x}\})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x}} \tag{42}$$

Equations (40) and (42) are the update equations of the Bernoulli filter for the intensity measurement model. Note that (42) is effectively the same as the conventional nonlinear Bayesian filter update equation (4).

V. Detector Output Measurements for a Point Target

A. Measurement model

The intensity measurement model can become computationally intractable if the number of sensor elements n is too large. In order to reduce the data flow, in many cases it is necessary to introduce a detector and report only the values which are

above a certain threshold. This is illustrated in Fig.1, where $n = 32$, but only three detections, $\mathbf{z}_1$, $\mathbf{z}_2$ and $\mathbf{z}_3$, are reported for further processing. Note that the true target detection can be missing due to the low signal-to-noise ratio or excessively high detection threshold. A low detection threshold increases the probability of detection but creates more false detections.

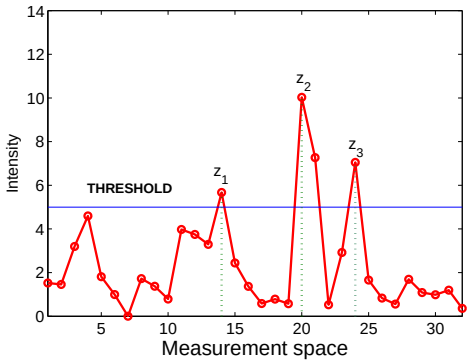


Figure 1. Detection threshold reduces the data flow: of 32 measurements reported by the sensor, only a set $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3\}$ passed the detection threshold (the candidates for detections are only the local maxima)

The measurements produced by a detector can mathematically be represented by a random finite set $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\}$. The main feature of this representation is that both the cardinality $m = |\mathbf{Z}|$ and the position of elements $\mathbf{z} \in \mathbf{Z}$ in the measurement space $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$, are random. Also note that the order of detections in the RFS $\mathbf{Z}$ is irrelevant.

The RFS $\mathbf{Z}$ can be seen as a union of two independent random finite sets:

$$\mathbf{Z} = \mathbf{C} \cup \mathbf{W} \tag{43}$$

where $\mathbf{C}$ is the RFS of false detections (also known as clutter) and $\mathbf{W}$ is the RFS due to the object of interest. We assume here a "point object" (target) which, if detected, produces only one detection (the measurement model for non-point or extended objects will be considered in Sec.VI). For a point object, the RFS $\mathbf{W}$ can be either an empty set or a singleton, depending on whether the object has been detected or not. We are now after the mathematical model of the likelihood function of the measurement set $\mathbf{Z}$, denoted $\varphi(\mathbf{Z}|\mathbf{X})$, where $\mathbf{X}$ is (as always) the Bernoulli RFS. The likelihood function $\varphi(\mathbf{Z}|\mathbf{X})$ will therefore have two forms, one for $\mathbf{X} = \emptyset$ and the other for $\mathbf{X} = \{\mathbf{x}\}$.

For the case $\mathbf{X} = \emptyset$, the measurement set will consist of false detections only, i.e. $\mathbf{Z} = \mathbf{C} \cup \emptyset = \mathbf{C}$. Typically the number of false detections in the measurement set is modelled by the Poisson distribution:

$$P\{|\mathbf{C}| = s\} = \frac{e^{-\lambda}\lambda^s}{s!}, \qquad s = 0, 1, 2, \ldots \tag{44}$$

where λ is the excepted value. Furthermore, conditioned on $|\mathbf{C}|$, false detections are modelled as independent, identically distributed (IID) random vectors taking values from $\mathcal{Z}$ with PDF $c(\mathbf{z})$. A RFS with such characteristics was introduced in Sec.II-B as the Poisson RFS [10]. According to (9), the FISST PDF of clutter only detections has the form:

$$\varphi(\mathbf{Z}|\emptyset) = \kappa(\mathbf{Z}) = e^{-\lambda} \prod_{\mathbf{z}\in\mathbf{Z}} \lambda\, c(\mathbf{z}). \tag{45}$$

In order to derive the likelihood function $\varphi(\mathbf{Z}|\{\mathbf{x}\})$ we first need to specify the FISST PDF of the RFS $\mathbf{W}|\{\mathbf{x}\}$, denoted as $\eta(\mathbf{W}|\{\mathbf{x}\})$. This is a Bernoulli RFS which can be either empty (if the object is undetected, i.e. $\mathbf{W} = \emptyset$) or a singleton (if object is detected and caused a measurement $\mathbf{z}$, i.e. $\mathbf{W} = \{\mathbf{z}\}$). Hence we can write:

$$\eta(\mathbf{W}|\{\mathbf{x}\}) = \begin{cases} 1 - p_d(\mathbf{x}), & \text{if } \mathbf{W} = \emptyset \\ p_d(\mathbf{x})\; g(\mathbf{z}|\mathbf{x}), & \text{if } \mathbf{W} = \{\mathbf{z}\} \end{cases} \tag{46}$$

where $g(\mathbf{z}|\mathbf{x})$ is the (conventional) likelihood function of measurement $\mathbf{z}$ due to the object in state $\mathbf{x}$ and $p_d(\mathbf{x})$ is the probability of detecting the object in state $\mathbf{x}$.

Now we are in the position to find the expression for $\varphi(\mathbf{Z}|\{\mathbf{x}\})$ using one of the fundamental results of the FISST calculus: the convolution formula [10, p.385]. For a union of independent RFS, as in (43), the FISST PDF is given by:

$$\varphi(\mathbf{Z}|\{\mathbf{x}\}) = \sum_{\mathbf{W}\subseteq\mathbf{Z}} \eta(\mathbf{W}|\{\mathbf{x}\})\, \kappa(\mathbf{Z} \setminus \mathbf{W}) \tag{47}$$

where sign $\setminus$ denotes the set-difference operation and κ was defined by (45). Since the RFS $\mathbf{W}$ can be either empty or a singleton, the summation in (47) greatly simplifies and we have:

$$\begin{aligned} &\varphi(\mathbf{Z}|\{\mathbf{x}\}) \\ &\quad = \eta(\emptyset|\{\mathbf{x}\}) \cdot \kappa(\mathbf{Z}) + \sum_{\mathbf{z}\in\mathbf{Z}} \eta(\{\mathbf{z}\}|\{\mathbf{x}\}) \cdot \kappa(\mathbf{Z} \setminus \{\mathbf{z}\}) \qquad (48) \\ &\quad = \kappa(\mathbf{Z}) \left[1 - p_d(\mathbf{x}) + p_d(\mathbf{x}) \sum_{\mathbf{z}\in\mathbf{Z}} g(\mathbf{z}|\mathbf{x}) \frac{\kappa(\mathbf{Z} \setminus \{\mathbf{z}\})}{\kappa(\mathbf{Z})}\right] \qquad (49) \end{aligned}$$

Next we derive the update equations of the Bernoulli filter for the described measurement model.

B. Update equations

Update equations for the Bernoulli filter using the detector-output measurement model of point targets was originally derived in [10, App.G.24] and [15, Sec.3.6]. Recall from Sec.III-B, the output of the prediction step is the FISST PDF $f_{k|k-1}(\mathbf{X}|\boldsymbol{\Upsilon}_{1:k-1})$, which is uniquely determined by the pair $\left(q_{k|k-1}, s_{k|k-1}(\mathbf{x})\right)$ and takes the form (7). For the measurement model we consider here, $\boldsymbol{\Upsilon}_k \equiv \mathbf{Z}_k$. The updated FISST PDF follows from (12):

$$f_{k|k}(\mathbf{X}_k|\mathbf{Z}_{1:k}) = \frac{\varphi_k(\mathbf{Z}_k|\mathbf{X}_k) \cdot f_{k|k-1}(\mathbf{X}_k|\mathbf{Z}_{1:k-1})}{f_k(\mathbf{Z}_k|\mathbf{Z}_{1:k-1})} \tag{50}$$

where $\varphi_k(\mathbf{Z}_k|\mathbf{X}_k)$ was specified by (45) and (49), and

$$\begin{aligned} & f_k(\mathbf{Z}_k|\mathbf{Z}_{1:k-1}) \\ &= \int \varphi_k(\mathbf{Z}_k|\mathbf{X}) \cdot f_{k|k-1}(\mathbf{X}|\mathbf{Z}_{1:k-1})\, \delta\mathbf{X} \quad (51) \\ &= \varphi_k(\mathbf{Z}_k|\emptyset)\, f_{k|k-1}(\emptyset|\mathbf{Z}_{1:k-1}) + \\ &\quad \int \varphi_k(\mathbf{Z}_k|\{\mathbf{x}\})\, f_{k|k-1}(\{\mathbf{x}\}|\mathbf{Z}_{1:k-1})\, d\mathbf{x} \quad (52) \\ &= (1-q_{k|k-1})\, \varphi_k(\mathbf{Z}_k|\emptyset) + \\ &\quad q_{k|k-1} \int \varphi_k(\mathbf{Z}_k|\{\mathbf{x}\})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x} \quad (53) \end{aligned}$$

Using (45) and (49), the expression in (53) can be written (after a few steps) as:

$$\begin{aligned} f_k(\mathbf{Z}_k|\mathbf{Z}_{1:k-1}) = \kappa(\mathbf{Z}_k) \Bigg\{ 1 - q_{k|k-1} \int p_d(\mathbf{x}) s_{k|k-1}(\mathbf{x}) d\mathbf{x} \\ + q_{k|k-1} \sum_{\mathbf{z}\in\mathbf{Z}_k} \frac{\kappa(\mathbf{Z}_k \setminus \{\mathbf{z}\})}{\kappa(\mathbf{Z}_k)} \int p_d(\mathbf{x}) g_k(\mathbf{z}|\mathbf{x}) s_{k|k-1}(\mathbf{x}) d\mathbf{x} \Bigg\} \end{aligned} \quad (54)$$

where notation $g_k(\mathbf{z}|\mathbf{x})$ emphasizes the (possibly) time-varying aspect of this likelihood function. Again, since $f_{k|k-1}(\mathbf{X}_k|\mathbf{Z}_{1:k-1}) = 0$ for $|\mathbf{X}| \geq 2$, we have that

$$f_{k|k}(\mathbf{X}|\mathbf{Z}_{1:k}) = 0 \quad \text{if } |\mathbf{X}| \geq 2$$

and obviously from (50)

$$\int f_{k|k}(\mathbf{X}|\mathbf{Z}_{1:k})\, \delta\mathbf{X} = 1,$$

from which by comparison with (7) we have established that $f_{k|k}(\mathbf{X}_k|\mathbf{Z}_{1:k})$ is a FISST probability density for a Bernoulli RFS.

Let us now work out the posterior $f_{k|k}(\mathbf{X}_k|\mathbf{Z}_{1:k})$ for the case $\mathbf{X}_k = \emptyset$. First note that using (45), we can write:

$$\frac{\kappa(\mathbf{Z}_k \setminus \{\mathbf{z}\})}{\kappa(\mathbf{Z}_k)} = \frac{1}{\lambda\, c(\mathbf{z})} \quad (55)$$

Using (45), (54), (55) and the fact that $f_{k|k}(\mathbf{X}_k|\mathbf{Z}_{1:k}) = 1 - q_{k|k}$, equation (50) leads to the update equation for the probability of existence:

$$q_{k|k} = \frac{1-\Delta_k}{1 - q_{k|k-1}\Delta_k}\, q_{k|k-1} \quad (56)$$

where

$$\begin{aligned} \Delta_k = & \int p_d(\mathbf{x}) s_{k|k-1}(\mathbf{x}) d\mathbf{x} \\ & - \sum_{\mathbf{z}\in\mathbf{Z}_k} \frac{\int p_d(\mathbf{x}) g_k(\mathbf{z}|\mathbf{x})\, s_{k|k-1}(\mathbf{x}) d\mathbf{x}}{\lambda c(\mathbf{z})} \end{aligned} \quad (57)$$

If $p_d = const$ (independent of the target state), then (57) simplifies to:

$$\Delta_k = p_d \left(1 - \sum_{\mathbf{z}\in\mathbf{Z}_k} \frac{\int g_k(\mathbf{z}|\mathbf{x})\, s_{k|k-1}(\mathbf{x}) d\mathbf{x}}{\lambda c(\mathbf{z})} \right) \quad (58)$$

since $s_{k|k-1}(\mathbf{x})$ being the conventional PDF, integrates to 1.

By evaluation of the posterior $f_{k|k}(\mathbf{X}_k|\mathbf{Z}_{1:k})$ for the case $\mathbf{X}_k = \{\mathbf{x}\}$ from (50), and using the fact that $f_{k|k}(\{\mathbf{x}\}|\mathbf{Z}_{1:k}) = q_{k|k}\, s_{k|k}(\mathbf{x})$, one can derive the update equation for the spatial PDF $s_{k|k}(\mathbf{x})$. The final expression is:

$$s_{k|k}(\mathbf{x}) = \frac{1 - p_d(\mathbf{x}) + p_d(\mathbf{x}) \sum_{\mathbf{z}\in\mathbf{Z}_k} \frac{g_k(\mathbf{z}|\mathbf{x})}{\lambda\, c(\mathbf{z})}}{1-\Delta_k}\, s_{k|k-1}(\mathbf{x}) \quad (59)$$

Note that if $q_{k|k-1} = 1$ and $p_d = 1$, then from (56) and (58) it follows that $q_{k|k} = 1$. Furthermore, under the additional assumption that there are no false detections, the measurement set $\mathbf{Z}_k$ has to be a singleton whose only element $\mathbf{z}$ is due to the object of interest. Then (59) simplifies to:

$$s_{k|k}(\mathbf{x}) = \frac{g_k(\mathbf{z}_k|\mathbf{x})\, s_{k|k-1}(\mathbf{x})}{\int g_k(\mathbf{z}_k|\mathbf{x})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x}} \quad (60)$$

which is the standard Bayes filter update equation (4).

Notice also that both (56) and (59) can be interpreted as comprising $|\mathbf{Z}_k|+1$ independent contributions or components. The first component is derived from the hypothesis that the state is not detected. Each of the remaining $|\mathbf{Z}_k|$ components is derived from the hypothesis that the state is detected and generated the measurement $\mathbf{z} \in \mathbf{Z}_k$. All quantities are of course normalised.

VI. Detector Output Measurements for an Extended Target

We have earlier described a point object as an object that can cause at most one detection. In situations where the object is far away and smaller than sensor resolution (e.g. radar surveillance of aircraft), this is a reasonable assumption. However, there are also many situations where this assumption is not appropriate. For example, if a high resolution radar is used for maritime surveillance, a ship (object of interest) can appear in many resolution cells. In computer vision this is particularly widespread, because an object typically occupies the whole region of an image, consisting of many pixels.

A. Measurement model

The main feature of an extended or non-point object is that it consists of several scattering (feature or measurement generating) points. The number of these scattering points is unknown and time-varying (as the object moves, turns, etc). Since the probability of detecting each scattering point is typically less than one, an extended object can cause zero, one or more detections. In computer vision, for example, the scattering points could be the corners [40, p.320] or invariant features [41, Ch.10].

The measurement set is modelled by a RFS $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m\}$, as in Sec.V-A. Let the number of scattering points be L. We can again represent $\mathbf{Z}$ as a union (43), where $\mathbf{C}$ is the RFS of false detections. The RFS $\mathbf{W}$ is a problem - one can be tempted to adopt the model $\mathbf{W} = \cup_{j=1}^{L} \mathbf{W}_j$, where $\mathbf{W}_j$ are mutually independent Bernoulli RFSs, each corresponding to one scattering point $j = 1, \ldots, L$. This type of RFS is known as the multi-Bernoulli RFS [10]. If we were to continue the derivation in this framework, we would practically develop a method that tracks L individual scattering points, see [42]. This framework, however, is exceptionally

complex and the resulting algorithm computationally difficult to implement in a general non-linear/non-Gaussian case.

We take a different approach, and model the RFS $\mathbf{W}$ by a binomial RFS, which treats object originated detections as having a single source. In this way, instead of tracking individual scattering points, we will track the object centroid and its shape/size[7]. According to Sec.II-B, the cardinality distribution of the binomial RFS is the binomial distribution, while the elements of $\mathbf{W}$ are IID with the spatial distribution $g(\mathbf{w}|\mathbf{x})$. Following (10), the FISST PDF of this RFS is given by:

$$\begin{aligned}&\eta(\mathbf{W}|\{\mathbf{x}\})\\ &= \frac{L!}{(L-|\mathbf{W}|)!}\, p_d^{|\mathbf{W}|}\, [1-p_d]^{L-|\mathbf{W}|} \prod_{\mathbf{w}\in\mathbf{W}} g(\mathbf{w}|\mathbf{x}) \qquad (61)\\ &= \begin{cases} (1-p_d)^L, & \text{if } \mathbf{W}=\emptyset \\ L\, p_d\, (1-p_d)^{L-1}\, g(\mathbf{w}|\mathbf{x}), & \text{if } \mathbf{W}=\{\mathbf{w}\} \\ \cdots & \\ L!\, p_d^L\, g(\mathbf{w}_1|\mathbf{x})\cdots g(\mathbf{w}_L|\mathbf{x}), & \text{if } \mathbf{W}=\{\mathbf{w}_1,\ldots,\mathbf{w}_L\} \end{cases} \qquad (62)\end{aligned}$$

for cardinalities $|\mathbf{W}| = 0, 1, 2, \ldots, L$. Here p_d is the probability of detection of a scattering point; it can be made dependent on the object state $\mathbf{x}$, but this is omitted for simplicity.

Recall that an extended object occupies an area in 2D or a volume in 3D. The state space for an extended object should therefore include, in addition to the kinematic properties of its centroid, the object size in a parametric from. For example, we can assume that the object is an ellipse with the state vector is specified by: $\mathbf{x} = \begin{bmatrix} x & \dot{x} & y & \dot{y} & a & b & c \end{bmatrix}^\intercal$, where (x, y) and $(\dot{x}, \dot{y})$ are the centroid position and velocity in a 2D Cartesian coordinate system, respectively. Components of the state a, b and c determine the size and orientation of its elliptic shape. The likelihood $g(\mathbf{w}|\mathbf{x})$ can be for example the Gaussian distribution whose mean is the object centroid, and whose covariance reflects the elliptic shape of the object.

Note, however, that we can also adopt a different form of the likelihood $g(\mathbf{w}|\mathbf{x})$, based on the interpretation of $\mathbf{x}$ as a closed set. For example, if the object is in the shape of an ellipse, $\mathbf{x}$ can be interpreted as a set of points which belong to the ellipse centered at (x, y) with parameters a, b, c. In this interpretation, the state is modelled by a special type of a random variable which takes values as closed sets, referred to as a random closed set (RCS) [10]. Let $h(\mathbf{x})$ define the mapping from the state space to the measurement space. Since $\mathbf{x}$ is a RCS, then $h(\mathbf{x})$ too is a RCS. For a measurement $\mathbf{w} \in \mathbf{W}$, which is due to a RCS $\mathbf{x}$, the likelihood $g(\mathbf{w}|\mathbf{x})$ is referred to as the *generalised* likelihood function (GLF) [10], [44], and is defined as: $\tilde{g}(\mathbf{w}|\mathbf{x}) = P\{\mathbf{w} \in h(\mathbf{x})\}$. A convenient choice is for example the indicator function, i.e. $\tilde{g}(\mathbf{w}|\mathbf{x}) = \mathbb{I}_{h(\mathbf{x})}(\mathbf{w})$, which equals 1 if $\mathbf{w} \in h(\mathbf{x})$ and zero otherwise. More will be said about generalised likelihoods in Sec.VII.

[7]There is also another similar approach that models the object originated detections by a Poisson RFS, see [43]. Due to the space limitation, this approach will not be considered in the tutorial.

Following the procedure in Sec.V-A, we next apply the convolution formula (47) in order to derive $\varphi_k(\mathbf{Z}|\{\mathbf{x}\})$. For the case $\mathbf{X} = \emptyset$, as usual we have $\varphi_k(\mathbf{Z}|\emptyset) = \kappa(\mathbf{Z})$. The convolution leads to:

$$\begin{aligned}&\varphi(\mathbf{Z}|\{\mathbf{x}\})\\ &\quad= \sum_{\mathbf{W}\subseteq\mathbf{Z}} \eta(\mathbf{W}|\{\mathbf{x}\})\, \kappa(\mathbf{Z}\setminus\mathbf{W}) \qquad (63)\\ &\quad= \eta(\emptyset|\{\mathbf{x}\})\, \kappa(\mathbf{Z}) + \sum_{\Omega\in\mathcal{P}_{1:L}(\mathbf{Z})} \eta(\Omega|\{\mathbf{x}\})\, \kappa(\mathbf{Z}\setminus\Omega) \qquad (64)\end{aligned}$$

where $\mathcal{P}_{1:L}(\mathbf{Z})$ is the set of all subsets of $\mathbf{Z}$ with cardinalities equal to $1, 2, \ldots, L$. If $L \geq |\mathbf{Z}|$, then $\mathcal{P}_{1:L}(\mathbf{Z})$ is the power set of $\mathbf{Z}$ minus the empty set. Using (61) with (55), the expression in (64) simplifies to:

$$\begin{aligned}\varphi(\mathbf{Z}|\{\mathbf{x}\}) = \kappa(\mathbf{Z}) \Big\{ &(1-p_d)^L + \\ &\sum_{\Omega\in\mathcal{P}_{1:L}(\mathbf{Z})} \frac{L!}{(L-|\Omega|)!}\, p_d^{|\Omega|}\, (1-p_d)^{L-|\Omega|} \prod_{\mathbf{z}\in\Omega} \frac{g(\mathbf{z}|\mathbf{x})}{\lambda\, c(\mathbf{z})} \Big\} \qquad (65)\end{aligned}$$

B. Update equations

The updated FISST PDF follows from the Bayes rule, see (12) or (50), where $\varphi_k(\mathbf{Z}|\emptyset) = \kappa(\mathbf{Z})$ and $\varphi_k(\mathbf{Z}_k|\{\mathbf{x}\})$ is given by (65). As usual, the first step is to find $f_k(\mathbf{Z}_k|\mathbf{Z}_{1:k-1})$. Assuming the reader is already familiar with the steps in derivation, we state only the final expression:

$$\begin{aligned}f(\mathbf{Z}_k|\mathbf{Z}_{1:k-1}) = \kappa(\mathbf{Z}_k) \Big\{ &1 - q_{k|k-1} + q_{k|k-1}(1-p_d)^{L_k} + \\ &\sum_{\Omega\in\mathcal{P}_{1:L_k}(\mathbf{Z}_k)} \psi_k\, \frac{\int \prod_{\mathbf{z}\in\Omega} g_k(\mathbf{z}|\mathbf{x})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x}}{\prod_{\mathbf{z}\in\Omega} \lambda\, c(\mathbf{z})} \Big\} \qquad (66)\end{aligned}$$

with

$$\psi_k = \frac{L_k!}{(L_k-|\Omega|)!}\, \frac{p_d^{|\Omega|}}{(1-p_d)^{|\Omega|-L_k}}. \qquad (67)$$

The subscript k in ψ_k, L_k and $g_k(\mathbf{z}|\mathbf{x})$ is there to emphasize the time varying nature of the number of scatters and the likelihood function.

Solving the Bayesian update equation (50) for $\mathbf{X} = \emptyset$ leads to the update equation for the probability of existence:

$$q_{k|k} = \frac{1-\Delta_k}{1-q_{k|k-1}\Delta_k}\, q_{k|k-1} \qquad (68)$$

where

$$\begin{aligned}\Delta_k = 1 - (1-p_d)^{L_k} - \\ \sum_{\Omega\in\mathcal{P}_{1:L_k}(\mathbf{Z}_k)} \psi_k\, \frac{\int \prod_{\mathbf{z}\in\Omega} g_k(\mathbf{z}|\mathbf{x})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x}}{\prod_{\mathbf{z}\in\Omega} \lambda\, c(\mathbf{z})} \qquad (69)\end{aligned}$$

Similarly, solving (50) for $\mathbf{X} = \{\mathbf{x}\}$ leads to the update equation for the spatial PDF:

$$s_{k|k}(\mathbf{x}) = \frac{(1-p_d)^{L_k} + \sum_{\Omega\in\mathcal{P}_{1:L_k}(\mathbf{Z}_k)} \psi_k \prod_{\mathbf{z}\in\Omega} \frac{g_k(\mathbf{z}|\mathbf{x})}{\lambda\, c(\mathbf{z})}}{1-\Delta_k} \cdot s_{k|k-1}(\mathbf{x}) \qquad (70)$$

If $L_k = 1$, then (69) and (70) reduce to (58) and (59), respectively.

VII. Non-Standard Measurements

Up to this point we implicitly made two assumptions about sensor measurements. First, each measurement is a point in the measurement space $\mathcal{Z}$. Second, the likelihood function of a point measurement, $g_k(\mathbf{z}|\mathbf{x})$, is precisely known. While these two assumptions are widespread, they could be fairly unrealistic in many fields of science and engineering. For example, a natural language statement, such as "the ball is near the center of the field", is an instance of a non-point measurements about the object of interest (the ball). This statement can be translated to a "measurement" which covers a region (an area) around the "center of the field" consisting of an infinite number of points. Imprecise likelihood functions are also very common in practice. For example, object localisation based on the received signal strength in cellular networks involves an imprecise likelihood function, because this likelihood depends on imprecisely known path-loss exponent [45].

When we make inference about a certain system, object or phenomenon, we sometimes have at our disposal prior knowledge expressed in the form of uncertain implication rules. For example, suppose the goal is to localize a suspect and the following piece of intelligence is available for reasoning: "The suspect is often in the *Corner cafe* between 9am and 10am". This piece of intelligence can be expressed as an uncertain implication rule: **if** $y \in Y$, **then** $\mathbf{x} \in \mathbf{X}$, with probability α (y is time, Y is an interval of time, $\mathbf{x}$ is the state, that is the location of the suspect, and $\mathbf{X}$ is the subset of the state space). The rule is characterised as *uncertain* because α can be smaller than 1.

This section reviews the update step of the Bernoulli filter using so-called *non-standard measurements*. By non-standard measurements we mean any combination of the following: (1) imprecise measurements (such as intervals or fuzzy intervals) characterised by precise likelihoods; (2) precise (point) measurements characterised by imprecise likelihoods; (3) uncertain implication rules.

Before we proceed with the treatment of each non-standard measurement separately, we first state the common framework for their processing. The theory is based on [10]. Essentially, all update equations of the Bernoulli filter that we presented so far for traditional measurements (point measurements with precise likelihoods) are valid for non-standard measurements. The only difference is that when we deal with non-standard measurements, the likelihood functions $g_k(\mathbf{z}|\mathbf{x})$ and $c(\mathbf{z})$ need to be replaced with generalised likelihood functions $\tilde{g}_k(\mathbf{z}|\mathbf{x})$ and $\tilde{c}(\mathbf{z})$, respectively. A theoretical justification of the GLF from the measure-theoretic point of view is given in [46]. The generalised likelihood also provides a useful relationship with Dempster-Shafer theory (DST) [47], because by definition it is identical to the plausibility function (which plays an important role in the DST) on singletons [48]. The rest of this section defines the generalised likelihoods for imprecise measurements, imprecise likelihoods and uncertain implication rules.

A. Imprecise measurements

Imprecision is a form of uncertainty distinctly different from randomness. The two kinds of uncertainties have been debated by philosophers under the terms *epistemic uncertainty* (due to lack of knowledge) and *aleatory uncertainty* (due to randomness) [49]. Imprecision or epistemic uncertainty has been studied intensively in the field of expert systems and artificial intelligence [50], but significantly less in statistics [51].

Suppose $h(\mathbf{x})$ is the measurement function which maps the state $\mathbf{x} \in \mathcal{X}$ into a point in the measurement space $\mathcal{Z}$. Furthermore, let's assume that the standard point measurement can be modelled as: $\mathbf{z} = h(\mathbf{x}) + \mathbf{w}$, where $\mathbf{w}$ is additive noise, distributed according to $p_{\mathbf{w}}$.

An imprecise measurement, denoted by $\boldsymbol{\zeta}$, is a subset of the measurement space $\mathcal{Z}$, see Fig.2. The imprecise measurement is therefore modelled by a special type of a random variable, which takes values as closed sets on $\mathcal{Z}$. A rigorous mathematical treatment of random closed sets is beyond the scope of this tutorial, with details in [10, Ch.4-7], [52].

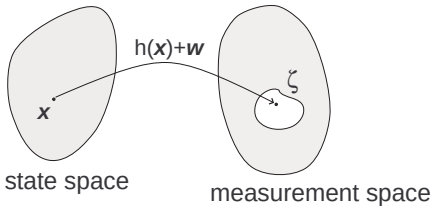


Figure 2. An imprecise measurement $\boldsymbol{\zeta}$ is a subset of the measurement space $\mathcal{Z}$

The GLF of a random set measurement $\boldsymbol{\zeta}$ is defined by [10], [53]:

$$\tilde{g}(\boldsymbol{\zeta}|\mathbf{x}) = P\{\mathbf{z} \in \boldsymbol{\zeta}\} = P\{h(\mathbf{x}) + \mathbf{w} \in \boldsymbol{\zeta}\}. \qquad (71)$$

The imprecise measurement $\boldsymbol{\zeta}$ can be a fuzzy or a crisp set[8] in the measurement space $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$. If $h(\mathbf{x})$ is a proper function, the GLF can be modelled by a single membership function of a fuzzy set on $\mathcal{Z}$. In some cases, however, there is ambiguity in mapping from the state space $\mathcal{X}$ to the measurement space $\mathcal{Z}$, that is a point in the state space $\mathbf{x} \in \mathcal{X}$ maps into multiple regions in $\mathcal{Z}$. For example, suppose a report is received that the object we want to localise has been seen near a traffic light. The source of ambiguity could be the existence of several traffic lights in the surveillance region. In this case the GLF can be represented by a weighted sum of membership functions, describing multiple fuzzy sets on $\mathcal{Z}$ [10, Ch.5], i.e.:

$$\tilde{g}(\zeta|\mathbf{x}) = \sum_i w_i \cdot \mu_i\big(h(\mathbf{x})\big) \qquad (72)$$

where $\mu_i(\mathbf{z})$ and w_i are membership functions and weights, respectively. Eq.(72) represents the most general form of a generalised likelihood.

[8] A fuzzy set is a set whose elements have a degree of membership. The classical set theory deals with crisp sets, whose membership of elements is binary, i.e. an element either belongs to it or not.

The Bernoulli filter for imprecise measurements has been discussed in [54] and [53]. Its application to detection and tracking using spatially referring natural language statements will be presented in Sec.IX-E.

B. Imprecise likelihoods

The traditional measurement $\mathbf{z} \in \mathcal{Z}$ is related to the state $\mathbf{x} \in \mathcal{X}$ via the measurement equation:

$$\mathbf{z} = h(\mathbf{x}; \boldsymbol{\theta}) + \mathbf{w} \tag{73}$$

where $h : \mathcal{X} \to \mathcal{Z}$ is the measurement function, $\boldsymbol{\theta}$ is a precisely known parameter vector and $\mathbf{w}$ is additive noise, distributed according to $p_{\mathbf{w}}$.

But what if the parameter vector $\boldsymbol{\theta}$ is not known precisely [44]? For example, we may know only that $\boldsymbol{\theta} \in [\boldsymbol{\theta}]$, where $[\boldsymbol{\theta}]$ denotes an interval in the parameter space. Then the mapping $h : \mathcal{X} \to \mathcal{Z}$ is not a function any more, because a point from $\mathcal{X}$ maps into an infinite number of points in $\mathcal{Z}$, see Fig.3. The solution proposed in [10, Ch.4-7] is to represent the measurement set $h(\mathbf{x}; [\boldsymbol{\theta}])+\mathbf{w}$ by a random closed set $\mathbf{S}_{\mathbf{x}} \subseteq \mathcal{Z}$. The GLF of a point measurement $\mathbf{z} \in \mathcal{Z}$, characterised by an imprecise measurement function $\mathbf{h}(\mathbf{x}; [\boldsymbol{\theta}])$, is then defined as:

$$\tilde{g}(\mathbf{z}|\mathbf{x}) = P\{\mathbf{z} \in \mathbf{S}_{\mathbf{x}}\} = P\{\mathbf{z} \in h(\mathbf{x}; [\boldsymbol{\theta}]) + \mathbf{w}\} \tag{74}$$

If $\mathbf{w}$ is zero mean white Gaussian with covariance matrix $\boldsymbol{\Sigma}$, that is $p_{\mathbf{w}}(\mathbf{w}) = \mathcal{N}(\mathbf{w}; \mathbf{0}, \boldsymbol{\Sigma})$, the GLF (74) has an analytic solution [44]:

$$\tilde{g}(\mathbf{z}|\mathbf{x}) = \int_{\underline{h}_x}^{\overline{h}_x} \mathcal{N}(h; \mathbf{z}, \boldsymbol{\Sigma})\, dh = \mathcal{C}(\mathbf{z}; \underline{h}_x, \boldsymbol{\Sigma}) - \mathcal{C}(\mathbf{z}; \overline{h}_x, \boldsymbol{\Sigma}) \tag{75}$$

where $\mathcal{C}(\mathbf{z}; \boldsymbol{\mu}_{\mathbf{z}}, \mathbf{P}_{\mathbf{z}}) = \int_{-\infty}^{\mathbf{z}} \mathcal{N}(\mathbf{u}; \boldsymbol{\mu}_{\mathbf{z}}, \mathbf{P}_{\mathbf{z}}) d\mathbf{u}$ is the Gaussian cumulative distribution function (CDF) and $\underline{h}_x = \min_{[\boldsymbol{\theta}]}\{h(\mathbf{x}; [\boldsymbol{\theta}])\}$ and $\overline{h}_x = \max_{[\boldsymbol{\theta}]}\{h(\mathbf{x}; [\boldsymbol{\theta}])\}$ are the limits of the set $h(\mathbf{x}; [\boldsymbol{\theta}])$.

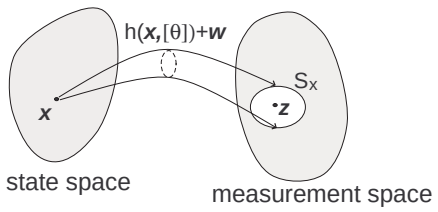


Figure 3. Mapping by an imprecise measurement function $h(\mathbf{x}; \boldsymbol{\theta}) + \mathbf{w}$, where $\boldsymbol{\theta} \in [\boldsymbol{\theta}]$ and $[\boldsymbol{\theta}]$ is an interval in the parameter space, results in a random set $\mathbf{S}_{\mathbf{x}} \subseteq \mathcal{Z}$

In general, partial knowledge of $\boldsymbol{\theta}$ can be represented by multiple non-overlapping and fuzzy intervals. Accordingly, the most general form of the GLF is a weighted sum of membership functions as in (72), i.e. $\tilde{g}(\mathbf{z}|\mathbf{x}) = \sum_i w_i \cdot \mu_i(\mathbf{z})$.

Robust Bayesian estimation using imprecise likelihood functions has been applied to localisation in [44]. The consequence of imprecision is a broader, hence more cautious, posterior PDF (compared to the posterior PDF obtained using the precise values of $\boldsymbol{\theta}$). The support of the posterior, however, is guaranteed to include the true state $\mathbf{x}$.

C. Uncertain implication rules

The first-order implication rule "if $\mathbf{Y}$ then $\mathbf{X}$", mathematically expressed as $\mathbf{Y} \Rightarrow \mathbf{X}$, is a shorter notation for

$$R: \qquad \mathbf{y} \in \mathbf{Y} \subseteq \mathcal{Y} \Rightarrow \mathbf{x} \in \mathbf{X} \subseteq \mathcal{X},$$

where $\mathbf{y}$ is a measurement on the measurement space $\mathcal{Y}$. In the example: "The suspect is often in the *Corner cafe* between 9am and 10am", the measurement space $\mathcal{Y}$ is time. The rule is typically assigned confidence $\alpha \in [0, 1]$.

Prior knowledge expressed by the rule R is treated as a non-standard measurement which updates the posterior via the Bayes rule. The expression for the GLF of an uncertain rule R can be found in [10], [55].

VIII. Implementation

A. Bernoulli particle filter

The sequential Monte Carlo method [3]–[8] provides a general framework for the implementation of Bernoulli filters [10]. The resulting Bernoulli-particle filters approximate the spatial PDF $s_{k|k}(\mathbf{x})$ by a particle system $\{w_k^{(i)}, \mathbf{x}_k^{(i)}\}_{i=1}^N$, where $\mathbf{x}_k^{(i)}$ is the state of particle i and $w_k^{(i)}$ is its weight. Since $s_{k|k}(\mathbf{x})$ is a conventional PDF, the weights are normalised, that is $\sum_{i=1}^N w_k^{(i)} = 1$.

Suppose at time $k-1$, the probability of existence is $q_{k-1|k-1}$ and the spatial PDF is approximated by

$$\widehat{s}_{k-1|k-1}(\mathbf{x}) = \sum_{i=1}^{N} w_{k-1}^{(i)}\, \delta_{\mathbf{x}_{k-1}^{(i)}}(\mathbf{x}) \tag{76}$$

where $\delta_{\mathbf{b}}(\mathbf{x})$ is the Dirac delta function concentrated at point $\mathbf{b}$. The computation of the predicted probability of existence $q_{k|k-1}$ is straightforward, see (28). According to (29), the prediction step for the spatial PDF involves the sum of two terms. Hence particle approximation of $s_{k|k-1}(\mathbf{x})$ be written as

$$\widehat{s}_{k|k-1}(\mathbf{x}) = \sum_{i=1}^{N+B} w_{k|k-1}^{(i)}\, \delta_{\mathbf{x}_{k|k-1}^{(i)}}(\mathbf{x}) \tag{77}$$

where the particles are drawn from two proposal distributions [16]:

$$\mathbf{x}_{k|k-1}^{(i)} \sim \begin{cases} \varrho_k(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{Z}_k) & i = 1, \ldots, N \\ \beta_k(\mathbf{x}_k|\mathbf{Z}_k) & i = N+1, \ldots, N+B \end{cases} \tag{78}$$

with weights

$$w_{k|k-1}^{(i)} = \begin{cases} \frac{p_s\, q_{k-1|k-1}}{q_{k|k-1}}\, \frac{\pi_{k|k-1}(\mathbf{x}_{k|k-1}^{(i)}|\mathbf{x}_{k-1}^{(i)}) w_{k-1}^{(i)}}{\varrho_k(\mathbf{x}_{k|k-1}^{(i)}|\mathbf{x}_{k-1}^{(i)}, \mathbf{Z}_k)}, \\ \qquad\qquad i = 1, \ldots, N \\ \frac{p_b(1-q_{k-1|k-1})}{q_{k|k-1}}\, \frac{b_{k|k-1}(\mathbf{x}_{k|k-1}^{(i)})}{\beta_k(\mathbf{x}_{k|k-1}^{(i)}|\mathbf{Z}_k)}\, \frac{1}{B}, \\ \qquad\qquad i = N+1, \ldots, N+B \end{cases} \tag{79}$$

Here B is the number of object-birth particles drawn from the proposal β_k. The design of proposal distributions has been discussed in standard particle filtering references [3]–[8]. One technique we found particularly useful is known as importance sampling with *progressive correction* or *tempering* [56]–[58].

In the absence of any prior knowledge on object birth, it is necessary to assume that the object of interest can appear anywhere in the state space $\mathcal{X}$. One may attempt in this case to model $b_{k|k-1}(\mathbf{x})$ by the uniform distribution over $\mathcal{X}$. The birth proposal β_k in (78) needs to have the same support as $b_{k|k-1}(\mathbf{x})$ (i.e. the entire $\mathcal{X}$) and consequently this approach would lead to a very inefficient algorithm requiring a massive number of birth particles.

A more efficient alternative is to use the measurements to build adaptively the birth density [17]. In this case

$$b_{k|k-1}(\mathbf{x}) \approx \int \pi_{k|k-1}(\mathbf{x}|\mathbf{x}')\, b_{k-1}(\mathbf{x}';\mathbf{Z}_{k-1})d\mathbf{x}' \qquad (80)$$

where $b_{k-1}(\mathbf{x};\mathbf{Z}_{k-1})$ is the object birth density at $k-1$, constructed using the measurement set at $k-1$ and prior knowledge (e.g. max/min speed). Let $\mathbf{x}_{k-1}^{(i)} \sim b_{k-1}(\mathbf{x};\mathbf{Z}_{k-1})$, for $i = N+1,\ldots,N+B$, be a sample representing the birth density at $k-1$. Then (78) and (79) can be written as:

$$\mathbf{x}_{k|k-1}^{(i)} \sim \varrho_k(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)},\mathbf{Z}_k), \qquad (81)$$

for $i = 1,\ldots,N,N+1,\ldots,N+B$ and

$$w_{k|k-1}^{(i)} = \begin{cases} \frac{p_s\, q_{k-1|k-1}}{q_{k|k-1}}\, \frac{\pi_{k|k-1}(\mathbf{x}_{k|k-1}^{(i)}|\mathbf{x}_{k-1}^{(i)})}{\varrho_k(\mathbf{x}_{k|k-1}^{(i)}|\mathbf{x}_{k-1}^{(i)},\mathbf{Z}_k)} \cdot w_{k-1}^{(i)}, \\ \qquad\qquad i = 1,\ldots,N \\ \frac{p_b(1-q_{k-1|k-1})}{q_{k|k-1}}\, \frac{\pi_{k|k-1}(\mathbf{x}_{k|k-1}^{(i)}|\mathbf{x}_{k-1}^{(i)})}{\varrho_k(\mathbf{x}_{k|k-1}^{(i)}|\mathbf{x}_{k-1}^{(i)},\mathbf{Z}_k)}\, \frac{1}{B}, \\ \qquad\qquad i = N+1,\ldots,N+B \end{cases} \qquad (82)$$

respectively.

The predicted Bernoulli PDF at time k is represented by $q_{k|k-1}$ and (77) approximated by $\{w_{k|k-1}^{(i)},\mathbf{x}_{k|k-1}^{(i)}\}_{i=1}^{N+B}$. The implementation of the update step depends on the measurement model, however the basic steps are the same. First, all integrals involving $s_{k|k-1}(\mathbf{x})$ are approximated by sums, due to (77), and computed. This is followed by the calculation of the probability of existence and updated particle weights. Finally, the resampling step is applied [4], [6] to eliminate the particles with small weights and multiply the particles with large weights. At the end of this procedure it is usually necessary to increase the particle diversity, for example by applying the MCMC move step [6], [59].

The pseudo-code of a Bernoulli particle filter (PF) for intensity measurements is presented in Alg. 1. The prediction is carried out in lines 3-5. Lines 6-10 compute the likelihood ratios for each predicted particle. Line 11 approximates, using particles, the integral which features in (40):

$$I_k = \int \ell_k(\mathbf{z}_k|\mathbf{x})\, s_{k|k-}(\mathbf{x})d\mathbf{x} \approx \sum_{i=1}^{N+B} \ell_k(\mathbf{z}|\mathbf{x}_{k|k-1}^{(i)})\, w_{k|k-1}^{(i)} \qquad (83)$$

Unnormalised weights are computed in lines 13-15, followed by normalisation in line 16. Resampling of particles is carried out in lines 17-20. After resampling, the weights of particles are uniform, see line 22. The birth particles, which will be required in the next cycle of the Bernoulli-particle filter, are drawn in line 23 and their weights set in line 24. It is important to note that in line 25 we output only the first N particles, because the particle system $\{w_k^{(i)},\mathbf{x}_k^{(i)}\}_{i=1}^{N}$ approximates the spatial PDF $s_{k|k}(\mathbf{x})$.

The complexity of the Bernoulli-particle filter for intensity measurements generally is higher than that for the standard particle filter. Even though each particle is drawn, updated and resampled as usual, the update for each particles involves a likelihood ratio calculation for all intensity returns, the number of which can be high for example in the case of an image measurement. The resultant cost remains linear in the number of particles, but becomes linear in the number of intensity returns which can be large. The recursion for the existence probability incurs a negligible increase in cost since the value of Bayes normalising constant is already computed and can be reused from the update of the spatial PDF.

Algorithm 1 Pseudo-code of a Bernoulli particle filter: intensity measurement model

1: **function** BERNOULLI PARTICLE FILTER 1
2: Input: $q_{k-1|k-1}$, $\{w_{k-1}^{(i)},\mathbf{x}_{k-1}^{(i)}\}_{i=1}^{N+B}$, $\mathbf{z}_k$
3: Predict existence probability using (28)
4: Draw a sample: $\mathbf{x}_{k|k-1}^{(i)} \sim \varrho_k(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)},\mathbf{z}_k)$ for $i = 1,\ldots,N+B$
5: Predicted weights $w_{k|k-1}^{(i)}$ for $i = 1,\ldots,N+B$ according to (82)
6: Compute likelihood $\varphi_k(\mathbf{z}_k|\emptyset)$, see (32)
7: **for** $i = 1,\ldots,N+B$ **do**
8: Compute likelihoods $\varphi_k(\mathbf{z}_k|\mathbf{x}_{k|k-1}^{(i)})$, see (32)
9: Compute likelihood ratios $\ell_k(\mathbf{z}_k|\mathbf{x}_{k|k-1}^{(i)})$, see (39)
10: **end for**
11: Approximate integral $I_k = \int \ell_k(\mathbf{z}_k|\mathbf{x})\, s_{k|k-}(\mathbf{x})d\mathbf{x}$, using (83)
12: Update existence probability: $q_{k|k} = \frac{I_k\, q_{k|k-1}}{1-q_{k|k-1}+q_{k|k-1}\cdot I_k}$
13: **for** $i = 1,\ldots,N+B$ **do**
14: Update weights, eq.(42): $\tilde{w}_{k|k}^{(i)} = \varphi_k(\mathbf{z}_k|\mathbf{x}_{k|k-1}^{(i)})\, w_{k|k-1}^{(i)}$
15: **end for**
16: Normalise weights: $w_{k|k}^{(i)} = \frac{\tilde{w}_{k|k}^{(i)}}{\sum_{j=1}^{N+B}\tilde{w}_{k|k}^{(j)}}$, for $i = 1,\ldots,N+B$
17: **for** $i = 1,\ldots,N$ **do**
18: Select index $j^{(i)} \in \{1,\ldots,N\}$ with probability $w_{k|k}^{(i)}$
19: $\mathbf{x}_k^{(i)} = \mathbf{x}_{k|k-1}^{(j_i)}$
20: **end for**
21: Particle regularisation (MCMC move)
22: Set weights: $w_k^{(i)} = 1/N$ for $i = 1,\ldots,N$.
23: Draw birth particles: $\mathbf{x}_k^{(i)} \sim b_k(\mathbf{x};\mathbf{z}_k)$, $i = N+1,\ldots,N+B$
24: Birth particle weights: $w_k^{(i)} = 1/B$ for $i = N+1,\ldots,N+B$.
25: Output: $q_{k|k}$, $\{w_k^{(i)},\mathbf{x}_k^{(i)}\}_{i=1}^{N}$
26: **end function**

The pseudo-code of a Bernoulli PF for the detector-output measurement model (point target) of Sec.V is given in Alg. 2. The prediction, carried out in lines 3-5, is the same as in Alg.1. The relevant equations of the filter update are (56), (57) and (59). Integrals involving $s_{k|k-1}(\mathbf{x})$ are approximated by sums as follows:

$$I_1 = \int p_d(\mathbf{x})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x} \approx \sum_{i=1}^{N+B} p_d\left(\mathbf{x}_{k|k-1}^{(i)}\right)\, w_{k|k-1}^{(i)} \qquad (84)$$

$$I_2(\mathbf{z}) = \int p_d(\mathbf{x})\, g_k(\mathbf{z}|\mathbf{x})\, s_{k|k-1}(\mathbf{x})\, d\mathbf{x} \approx \sum_{i=1}^{N+B} p_d\left(\mathbf{x}_{k|k-1}^{(i)}\right) g_k\left(\mathbf{z}|\mathbf{x}_{k|k-1}^{(i)}\right) w_{k|k-1}^{(i)} \qquad (85)$$

Computation of integrals is carried out in lines 6 and 7 of Alg. 2. The expression for Δ_k in (57) is then approximated by:

$$\Delta_k \approx I_1 - \sum_{\mathbf{z}\in\mathbf{Z}_k} \frac{I_2(\mathbf{z})}{\lambda\, c(\mathbf{z})}. \tag{86}$$

Δ_k, which is computed in line 8, is then plugged into (56) to compute $q_{k|k}$, see line 9. Updated weights before normalisation are computed using (59) as follows:

$$\tilde{w}^{(i)}_{k|k} = \left[1 - p_d\left(\mathbf{x}^{(i)}_{k|k-1}\right) + p_d\left(\mathbf{x}^{(i)}_{k|k-1}\right) \sum_{\mathbf{z}\in\mathbf{Z}_k} \frac{g_k\left(\mathbf{z}|\mathbf{x}^{(i)}_{k|k-1}\right)}{\lambda\, c(\mathbf{z})}\right] \cdot w^{(i)}_{k|k-1} \tag{87}$$

for $i = 1, \ldots, N + B$. This is carried out in line 10. The remaining steps of the Bernoulli-particle filter in Alg.2 are identical to those in Alg.1. The number of birth particles B, drawn in lines 18, is typically made dependent on the cardinality of the measurement set $\mathbf{Z}_k$ (e.g. for each $\mathbf{z} \in \mathbf{Z}_k$ one can draw a fixed number of birth particles).

The complexity of the Bernoulli-particle filter for detector-output measurements is generally similar to that of the standard particle filter. Each particle is drawn, updated and resampled, in a similar manner. The main difference lies in the computation of the detector-output likelihood for each particle, which involves a sum over all detections, and results in a cost that is linear in the number of particles and in the number of detections, thus incurring an additional cost compared to the computation of standard likelihood. The increase in computational cost is generally marginal, however it may become noticeable with a high rate of false alarms. There is also the added calculation of the existence probability, which incurs a small and fixed computational cost, and is related to the calculation of the Bayes normalising constant.

Algorithm 2 Pseudo-code of a Bernoulli particle filter: detector-output measurement model (point target)

1: **function** BERNOULLI PARTICLE FILTER 2
2: Input: $q_{k-1|k-1}$, $\{w^{(i)}_{k-1}, \mathbf{x}^{(i)}_{k-1}\}_{i=1}^{N+B}$, $\mathbf{Z}_k$
3: Predict existence probability using eq.(28)
4: Draw a sample: $\mathbf{x}^{(i)}_{k|k-1} \sim \varrho_k(\mathbf{x}_k|\mathbf{x}^{(i)}_{k-1}, \mathbf{Z}_k)$ for $i = 1, \ldots, N + B$
5: Predicted weights $w^{(i)}_{k|k-1}$ for $i = 1, \ldots, N + B$ according to (82)
6: Approximate integral I_1 of (84)
7: For every $\mathbf{z} \in \mathbf{Z}_k$ approximate integral $I_2(\mathbf{z})$ of (85)
8: Compute Δ_k approximation using (86)
9: Update existence: $q_{k|k} = \frac{1-\Delta_k}{1-\Delta_k \cdot q_{k|k-1}} \cdot q_{k|k-1}$
10: Update weights $\tilde{w}^{(i)}_{k|k}$, according to (87), for $i = 1, \ldots, N + B$.
11: Normalise weights: $w^{(i)}_{k|k} = \frac{\tilde{w}^{(i)}_{k|k}}{\sum_{j=1}^{N+B} \tilde{w}^{(j)}_{k|k}}$, for $i = 1, \ldots, N + B$
12: **for** $i = 1, \ldots, N$ **do**
13: Select index $j^{(i)} \in \{1, \ldots, N\}$ with probability $w^{(i)}_{k|k}$
14: $\mathbf{x}^{(i)}_k = \mathbf{x}^{(j_i)}_{k|k-1}$
15: **end for**
16: Particle regularisation (MCMC move)
17: Set weights: $w^{(i)}_k = 1/N$ for $i = 1, \ldots, N$.
18: Draw birth particles: $\mathbf{x}^{(i)}_k \sim b_k(\mathbf{x}; \mathbf{Z}_k)$, $i = N + 1, \ldots, N + B$
19: Birth particle weights: $w^{(i)}_k = 1/B$ for $i = N + 1, \ldots, N + B$.
20: Output: $q_{k|k}$, $\{w^{(i)}_k, \mathbf{x}^{(i)}_k\}_{i=1}^{N}$
21: **end function**

The pseudo-code of the Bernoulli PF for the detector-output measurement model and an *extended object* is a straightforward extension of Alg.2 and is therefore omitted from the tutorial.

If the measurements are non-standard, the only difference with respect to the above Bernoulli PF implementations is in the update of the particle weights. In particular, the (standard) likelihood functions $g(\mathbf{z}|\mathbf{x}^{(i)}_{k|k-1})$ are replaced by the generalised likelihoods $\tilde{g}(\mathbf{z}|\mathbf{x}^{(i)}_{k|k-1})$. The non-standard measurements tend to diffuse the posterior PDF $s_{k|k}(\mathbf{x})$ and consequently require more particles for accurate approximation. This was a motivation to develop a version of the Bernoulli PF for interval measurements, referred to as Bernoulli box-particle filter [53]. Box particle filtering is outside the scope of this tutorial, but further details can be found in [60].

Convergence analysis of the particle implementation of the Bernoulli filter can be found in [61]. In particular, [61] established the conditions for uniform convergence of a general class of filtering algorithms, including the particle Bernoulli filter. These results also hold for other implementations such as the Gaussian mixture.

B. Bernoulli Gaussian sum filter

The Bernoulli Gaussian sum filter (GSF) implementation [18] follows as a straightforward extension of the standard Gaussian sum filter [1], [2], [6], [62]. Under linear Gaussian assumptions, it is possible to propagate the spatial PDF $s_{k|k}(\mathbf{x})$ as a Gaussian sum, in exact closed form, i.e. analytically as a set of weights, means and covariances $\{w^{(i)}_k, \mathbf{m}^{(i)}_k, \mathbf{P}^{(i)}_k\}_{i=1}^{N_k}$. Pruning and merging of Gaussians is naturally required, and discussed briefly after the statement of the recursion.

We present the solution for the case of the detector-output measurement model (point target) of Sec.V. Equations for the extended object case are a straightforward extension and hence omitted. The case of intensity measurements where noise is usually non-Gaussian is best dealt with via the particle filter. As usual, $\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P})$ denotes a Gaussian PDF in the variable $\mathbf{x}$, with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{P}$.

Assume a linear Gaussian transition, likelihood and birth model, i.e.

$$\pi_{k|k-1}(\mathbf{x}|\mathbf{x}') = \mathcal{N}(\mathbf{x}; \mathbf{F}_{k-1}\mathbf{x}', \mathbf{Q}_{k-1}), \tag{88}$$

$$g_k(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \mathbf{H}_k\mathbf{x}, \mathbf{R}_k), \tag{89}$$

$$b_{k|k-1}(\mathbf{x}) = \sum_{i=1}^{N_{b,k}} w^{(i)}_{b,k} \mathcal{N}(\mathbf{x}; \mathbf{m}^{(i)}_{b,k}, \mathbf{Q}^{(i)}_{b,k}) \tag{90}$$

and constant survival and detection probabilities, i.e. $p_s(\mathbf{x}) = p_s$ and $p_d(\mathbf{x}) = p_d$.

Suppose at time $k-1$, the probability of existence is $q_{k-1|k-1}$ and the spatial PDF is given by a Gaussian sum of the form

$$s_{k-1|k-1}(\mathbf{x}) = \sum_{i=1}^{N_{k-1}} w^{(j)}_{k-1} \mathcal{N}(\mathbf{x}; \mathbf{m}^{(i)}_{k-1}, \mathbf{P}^{(i)}_{k-1}), \tag{91}$$

where $\sum_{i=1}^{N_{k-1}} w^{(j)}_{k-1} = 1$, then,

$$q_{k|k-1} = p_b\,(1 - q_{k-1|k-1}) + p_s\, q_{k-1|k-1}, \tag{92}$$

$$s_{k|k-1}(\mathbf{x}) = \frac{p_b(1-q_{k-1|k-1})}{q_{k|k-1}} b_{k|k-1}(\mathbf{x}) + \frac{p_s q_{k-1|k-1}}{q_{k|k-1}} \sum_{i=1}^{N_{k-1}} w_{k-1}^{(i)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}), \quad (93)$$

where $b_{k|k-1}(\mathbf{x})$ is expressed as a Gaussian sum in (90) and

$$\mathbf{m}_{k|k-1}^{(i)} = \mathbf{F}_{k-1}\mathbf{m}_{k-1}^{(i)}, \quad (94)$$

$$\mathbf{P}_{k|k-1}^{(i)} = \mathbf{Q}_{k-1} + \mathbf{F}_{k-1}\mathbf{P}_{k-1}^{(i)}\mathbf{F}_{k-1}^{T}. \quad (95)$$

The predicted spatial PDF (93) can be expressed by a Gaussian sum of the form:

$$s_{k|k-1}(\mathbf{x}) = \sum_{i=1}^{N_{k|k-1}} w_{k|k-1}^{(i)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}), \quad (96)$$

where $\sum_{i=1}^{N_{k|k-1}} w_{k|k-1}^{(i)}$. The update equations of the Bernoulli GSF are then:

$$q_{k|k} = \frac{1-\Delta_k}{1-q_{k|k-1}\Delta_k} q_{k|k-1}, \quad (97)$$

$$s_{k|k}(\mathbf{x}) = \frac{(1-p_d)}{1-\Delta_k} s_{k|k-1}(\mathbf{x}) + \frac{p_d}{1-\Delta_k} \times \sum_{\mathbf{z}\in\mathbf{Z}_k} \sum_{i=1}^{N_{k|k-1}} \frac{w_{k|k-1}^{(i)} q_k^{(i)}(\mathbf{z})}{\lambda\, c(\mathbf{z})} \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k}^{(i)}, \mathbf{P}_{k|k}^{(i)}), \quad (98)$$

$$\Delta_k = p_d \left[1 - \sum_{\mathbf{z}\in\mathbf{Z}_k} \sum_{i=1}^{N_{k-1}} \frac{w_{k-1}^{(i)} q_k^{(i)}(\mathbf{z})}{\lambda c(\mathbf{z})} \right], \quad (99)$$

where

$$q_k^{(i)}(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \eta_{k|k-1}^{(i)}, \mathbf{S}_{k|k-1}^{(i)}), \quad (100)$$

$$\eta_{k|k-1}^{(i)} = \mathbf{H}_k \mathbf{m}_{k|k-1}^{(i)}, \quad (101)$$

$$\mathbf{S}_{k|k-1}^{(i)} = \mathbf{H}_k \mathbf{P}_{k|k-1}^{(i)} \mathbf{H}_k^T + \mathbf{R}_k, \quad (102)$$

$$\mathbf{m}_{k|k}^{(i)}(\mathbf{z}) = \mathbf{m}_{k|k-1}^{(i)} + \mathbf{K}_k^{(i)}(\mathbf{z} - \eta_{k|k-1}^{(i)}), \quad (103)$$

$$\mathbf{P}_{k|k}^{(i)} = \mathbf{P}_{k|k-1}^{(i)} - \mathbf{P}_{k|k-1}^{(i)} \mathbf{H}_k^T [\mathbf{S}_{k|k-1}^{(i)}]^{-1} \mathbf{H}_k \mathbf{P}_{k|k-1}^{(i)} \quad (104)$$

$$\mathbf{K}_k^{(i)} = \mathbf{P}_{k|k-1}^{(i)} \mathbf{H}_k^T [\mathbf{S}_{k|k-1}^{(i)}]^{-1}. \quad (105)$$

The above results specify a closed form recursion for the spatial PDF in terms of a Gaussian sum, i.e. a recursion for the weights, means, and covariances $\{w_{k|k}^{(i)}, \mathbf{m}_{k|k}^{(i)}, \mathbf{P}_{k|k}^{(i)}\}_{i=1}^{N_k}$ from $\{w_{k-1|k-1}^{(i)}, \mathbf{m}_{k-1|k-1}^{(i)}, \mathbf{P}_{k-1|k-1}^{(i)}\}_{i=1}^{N_{k-1}}$. As expected, the number of Gaussian components N_k required to represent the spatial PDF grows without bound over time. Implementation thus requires pruning of components with insignificant weights (i.e. deletion of components with low weights), and possibly merging of components which are closely spaced (i.e. replacement of several closely spaced ones, via a Mahalonbis distance, with a single matching mean and covariance). Adaptations of the Gaussian sum solution to accommodate mild non-linearities are also possible, typically via the approach of the extended Kalman filter (linearisation) or the unscented Kalman filter (sigma point approximation), however these are outside the scope of this tutorial.

IX. Selected applications

This section presents five applications of the Bernoulli filter which also demonstrate the choice of the appropriate measurement model.

A. Detection and tracking using a sensor network

Sensor networks are used in many applications, such as pollution monitoring, battlefield surveillance, machine fault detection, etc. Sensors are typically spatially distributed and measure physical or environmental properties, such as sound, vibration, pressure, pollution, temperature, to name a few [63]. We consider the problem of simultaneous detection and tracking of a moving object through a region of interest, in the ground surveillance context. The assumption is that the moving object produces energy (e.g. acoustic energy) and that the surveillance region is populated by sensors (e.g. microphones) which sample the energy field at their respective locations. Sensor locations are arbitrary and known to the fusion centre, which collects and processes the measurements in order to detect and track the moving objects. This problem has been studied by many authors, see for example [37], [64], [65]. In this application we deal with intensity measurements, whose model was described in Sec. IV-A. We consider two cases: in case 1 the measurement function $h_k^{(s)}(\mathbf{x}_k)$ in (31) is precisely known; in case 2, the measurement function is known only partially, as in Sec. VII-B.

The measurement model is adopted from [37], [65]. A measurement at sensor $s = 1, \ldots, n$ at discrete-time $k = 1, 2, \ldots$ is modelled by (31) with the measurement function expressed as [37]:

$$h_k^{(s)}(\mathbf{x}) = \hbar_s \cdot A_k \cdot \left[\frac{d_0}{\|\mathbf{p}_s - \mathbf{p}_k\|} \right]^{\alpha_s} \quad (106)$$

where A_k is emitted energy by the object of interest at reference distance d_0; $\hbar_s$ is the gain factor of sensor s; α_s is the propagation loss factor of sensor s, which in general depends on the environment, and $\|\mathbf{p}_s - \mathbf{p}_k\|$ is the Euclidean distance between the position of the moving object at time k, $\mathbf{p}_k = [x_k \quad y_k]^\intercal$, and the location of sensor s, $\mathbf{p}_s = [x_s \quad y_s]^\intercal$. Measurement noise in (31) is modelled as $w_k^{(s)} \sim \mathcal{N}(w; \mu_w, \sigma_w^2)$ for all sensors, see [37], [65]. The propagation loss factor in open space is $\alpha_s = 2$, but in practical applications, due to multipath and shadowing, it can take any value in the interval $[2, 4]$ [66], [45].

According to the described measurement model, the measurement function $\mathbf{h}_k(\mathbf{x}_k)$ is parametrised by vector $\boldsymbol{\theta}$, which includes α_s, $\hbar_s$, x_s, y_s, for $s = 1, \ldots, n$. Case 1 corresponds to the perfectly calibrated sensor network where $\boldsymbol{\theta}$ is known precisely. Case 2 considers a more realistic situation where $\boldsymbol{\theta}$ is only partially known, as an interval value (see Sec. VII-B). Thus $\alpha_s \in [2, 4]$, $\hbar_s \in [\underline{\hbar}, \overline{\hbar}]$, $x_s \in [\hat{x}_s - \epsilon_x, \hat{x}_s + \epsilon_x]$ and $y_s \in [\hat{y}_s - \epsilon_y, \hat{y}_s + \epsilon_y]$, for $s = 1, \ldots, n$. Here $\hat{x}_s$ and $\hat{y}_s$ are the nominal sensor coordinates in the Cartesian coordinate system with ϵ_x and ϵ_y being their confidence bounds. The

fusion centre knows only the nominal sensor coordinates and ϵ_x and ϵ_y.

The state vector in this application is $\mathbf{x}_k = [x_k \ \dot{x}_k \ y_k \ \dot{y}_k \ A_k]^\intercal$. Its dynamics is described by model (16) with $\pi_{k|k-1}(\mathbf{x}|\mathbf{x}') = \mathcal{N}(\mathbf{x}; \mathbf{F}_k\mathbf{x}', \mathbf{Q}_k)$, where

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{G}_k & \mathbf{0}_{22} & \mathbf{0}_{21} \\ \mathbf{0}_{22} & \mathbf{G}_k & \mathbf{0}_{21} \\ \mathbf{0}_{12} & \mathbf{0}_{21} & 1 \end{bmatrix} \quad \mathbf{Q}_k = \begin{bmatrix} \Xi_k & \mathbf{0}_{22} & \mathbf{0}_{21} \\ \mathbf{0}_{22} & \Xi_k & \mathbf{0}_{21} \\ \mathbf{0}_{12} & \mathbf{0}_{21} & \varpi_2 T_k \end{bmatrix}. \tag{107}$$

Here $\mathbf{0}_{nm}$ is an $n \times m$ zero-matrix,

$$\mathbf{G}_k = \begin{bmatrix} 1 & T_k \\ 0 & 1 \end{bmatrix}, \quad \Xi_k = \begin{bmatrix} \frac{\varpi_1 T_k^3}{3} & \frac{\varpi_1 T_k^2}{2} \\ \frac{\varpi_1 T_k^2}{2} & \varpi_1 T_k \end{bmatrix}, \tag{108}$$

$T_k = t_{k+1} - t_k$ is the sampling interval and ϖ_1 and ϖ_2 denote the intensity of process noise [67, p.269].

The objective of the fusion centre is to detect and track the object using received sensor measurements and prior knowledge of (measurement and dynamic) models and their parameters: μ_w, σ_w, d_0, $\{\hbar_s, \alpha_s, x_s, y_s\}_{s=1}^n$, ϖ_1, ϖ_2, T_k, p_b and p_s.

The Bernoulli PF for this application is implemented according to Alg.1. The likelihoods functions, which are computed in line 6, are defined by (32). In case 1, when $\boldsymbol{\theta}$ is precisely known, we have:

$$\begin{aligned} g_1^{(s)}\left(z_k^{(s)}|\mathbf{x}_{k|k-1}^{(i)}\right) &= \mathcal{N}\left(z; h_k^{(s)}\left(\mathbf{x}_{k|k-1}^{(i)}\right) + \mu_w, \sigma_w^2\right), \\ g_0^{(s)}\left(z_k^{(s)}\right) &= \mathcal{N}\left(z; \mu_w; \sigma_w^2\right) \end{aligned}$$

for $i = 1, \ldots, N+B$ and $s = 1, \ldots, n$. In case 2, when the measurement model is imprecise, we need to compute the generalised likelihood $\tilde{g}_1^{(s)}(z_k^{(s)}|\mathbf{x}_k)$ which follows from (75). Thus $\underline{h}_x$ and $\overline{h}_x$, which feature in (75), are computed for each particle $\mathbf{x}_{k|k-1}^{(i)}$, $i = 1, \ldots, N+B$ and measurement $z_k^{(s)}$ from sensor $s = 1, \ldots, n$ as follows:

$$\begin{aligned} \underline{h}_x^{(s)}(\mathbf{x}_{k|k-1}^{(i)}) &= \min\left\{\hbar_s A_{k|k-1}^{(i)} \frac{d_0^{\alpha_s}}{\|\mathbf{p}_s - \mathbf{p}_{k|k-1}^{(i)}\|^{\alpha_s}} + \mu_x\right\} \\ \overline{h}_x^{(s)}(\mathbf{x}_{k|k-1}^{(i)}) &= \max\left\{\hbar_s A_{k|k-1}^{(i)} \frac{d_0^{\alpha_s}}{\|\mathbf{p}_s - \mathbf{p}_{k|k-1}^{(i)}\|^{\alpha_s}} + \mu_x\right\} \end{aligned}$$

with minimisation/maximisation over $\hbar_s, \alpha_s, x_s, y_s$. Since the likelihood ratio (39), which is computed in line 9 of Alg.1, does not have units, we need to adopt the generalised likelihood even for the noise-only case. For this purpose we used in simulations

$$\tilde{g}_0^{(s)}\left(z_k^{(s)}\right) = \frac{\mathcal{N}\left(z; \mu_w; \sigma_w^2\right)}{\max_u \left\{\mathcal{N}\left(u; \mu_w; \sigma_w^2\right)\right\}}. \tag{109}$$

Fig.4 shows the placement of $n = 40$ sensors, the moving object trajectory (ground truth, gray solid line) and two estimated trajectories obtained by the two Bernoulli PFs (using the precise and imprecise models). As expected, the precise (perfectly calibrated) measurement model results in the more accurate estimate of the trajectory (red line).

The moving object starts emitting at $k = 5$ and continues to do so until $k = 45$. The total observation time is $K = 50$. The initial state of the moving objects at $k = 5$ is $\mathbf{x}_5 = [25\text{m}\ 1.2\text{m/s}\ 80\text{m}\ -1.65\text{m/s}\ 2000]^\intercal$. One half of the sensors is characterised by $\alpha = 2.3$ and $\hbar = 1.02$, while for the other half $\alpha = 2.9$ and $\hbar = 1.08$. The true placement of sensors is also random with $x_s \sim \mathcal{U}[\hat{x}_s - \epsilon_x, \hat{x}_s + \epsilon_x]$, $y_s \sim \mathcal{U}[\hat{y}_s - \epsilon_y, \hat{y}_s + \epsilon_y]$ and $\epsilon_x = \epsilon_y = 0.3$m. Other parameters were selected as: $\mu_w = 1$, $\sigma_w = 0.1$, $d_0 = 1$m, $T_k = 1$s, $\varpi_1 = 0.04$, $\varpi_2 = 1$. Both Bernoulli PFs used $N = B = 2000$ particles and applied the progressive correction in line 4 of Alg.1. Other algorithmic parameters were: $p_b = 0.02$, $p_s = 0.98$. Birth density $b_k(\mathbf{x}; \mathbf{z})$ in line 18 was selected as follows. Birth particles in (x, y) were drawn from a Gaussian density whose mean is the weighted mean of the locations of three sensors with the strongest readings, with appropriately chosen variance. The birth particles in velocity were drawn from a uniform density in $\dot{x}$ and $\dot{y}$, which spans from -5m/s to $+5$m/s. Finally, the birth particles in amplitude were drawn from a uniform density which spans from 5 to 8000. The Bernoulli PF using the imprecise measurement model was based on partial prior knowledge of $\alpha_s \in [2, 4]$, $\hbar_s \in [0.9, 1.1]$ for $s = 1, \ldots, n$ and $\epsilon_x = \epsilon_y = 0.3$m. The point estimates shown in Fig.4 were obtained using the expected value of the particle approximated posterior PDF.

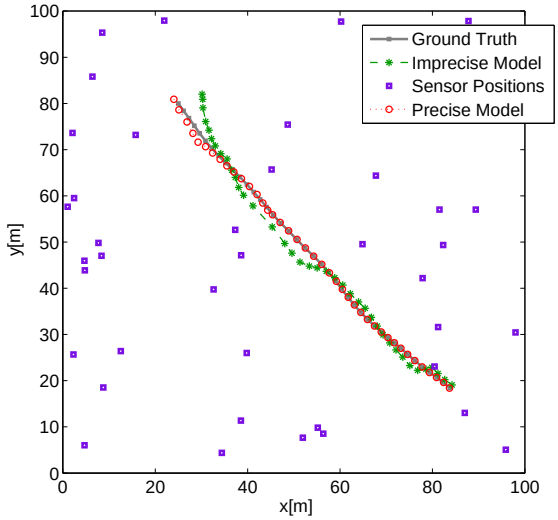


Figure 4. Sensor network application: moving object true trajectory (gray sold line), estimated trajectory using the precise correct model (red), estimated trajectory using the imprecise model (green); sensor locations indicated by squares.

Estimated probabilities of existence $q_{k|k}$ of two Bernoulli PFs are shown in Fig.5 for precise and imprecise measurement models. Finally, Fig.6 displays a zoomed-in particle approximation of the posterior PDF $s_{k|k}(\mathbf{x})$ at $k = 30$, both for the precise and imprecise measurement models.

The numerical results show that both Bernoulli PFs successfully detect the presence of the object and track its motion through the state space. In accordance with intuition, the Bernoulli PF using the precise (and correct) measurement model is far more accurate in estimation of the posterior PDF $s_{k|k}(\mathbf{x})$. However, considering that the perfect calibration of a sensor network is difficult and often impossible, an imprecise

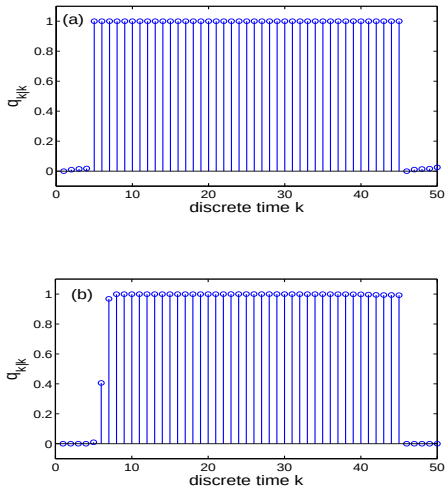


Figure 5. Estimated probability of existence $q_{k|k}$ over time: (a) using the precise model; (b) using the imprecise model

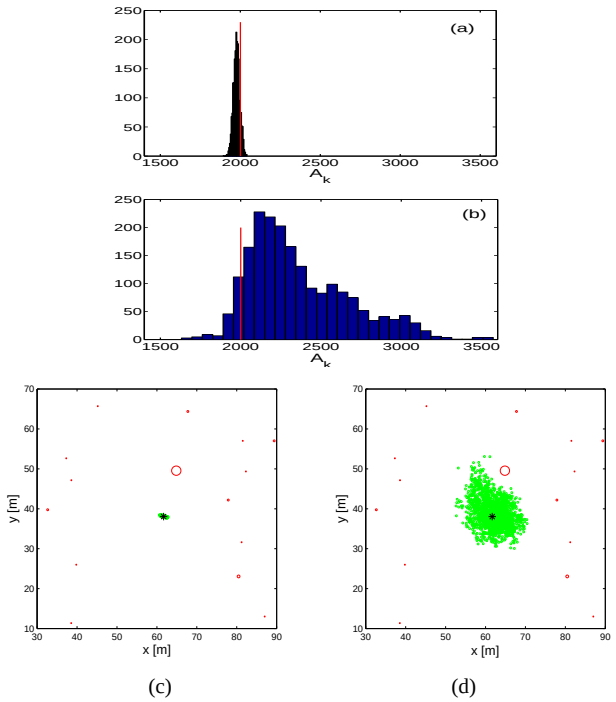


Figure 6. Particle approximation of the posterior PDF $s_{k|k}(\mathbf{x})$ at $k = 30$: (a) histogram in A_k space, precise model; (b) histogram in A_k space, imprecise model; (c) (x_k, y_k) space, precise model; (d) (x_k, y_k) space, imprecise model. The true value in (a) and (b) indicated by a vertical red line. The true values in (c) and (d) indicated by asterisk. The centre of each red circle in (c) and (d) represents the location of a sensor, while the radius of the circle is proportional to the measured value.

measurement model becomes a practical alternative. While its estimated posterior PDF is much broader (than the posterior obtained using the precise correct model), its support will contain the true object state, provided that the particle filter is properly implemented. This has been confirmed numerically in [44], [53]. Finally, note that using the precise but incorrect measurement model (e.g. by selecting wrongly the values of gains, propagation factors and sensor locations) typically leads to divergence of the filter.

B. Bearings-only tracking

The problem of autonomous bearings-only target tracking has been studied over many years due to its tremendous importance in passive surveillance. The basic problem is the sequential estimation of target location, speed and heading, from noise corrupted detector-output measurements of target line-of-sight (LOS) bearing. Due to the nonlinear measurement equation, the optimal solution in the sequential Bayesian framework results in a non-Gaussian posterior PDF. The consequence of this nonlinear/non-Gaussian context is the absence of a closed form solution even if detection is perfect (no false and missed detections). Various approximate implementations of the sequential Bayesian estimator have been considered in the ideal detection case, see reviews in [68] and [69].

The target state vector at discrete-time k is adopted as:

$$\mathbf{x}_k^t = \begin{bmatrix} x_k^t & \dot{x}_k^t & y_k^t & \dot{y}_k^t \end{bmatrix}^\mathsf{T} \tag{110}$$

where (x_k^t, y_k^t) and $(\dot{x}_k^t, \dot{y}_k^t)$ determine the target position and velocity in the two-dimensional Cartesian coordinate-system. The ownship state vector $\mathbf{x}_k^o$, which is known, is similarly defined. The motion model is then written for the relative state vector, defined as:

$$\mathbf{x}_k := \mathbf{x}_k^t - \mathbf{x}_k^o = \begin{bmatrix} x_k & \dot{x}_k & y_k & \dot{y}_k \end{bmatrix}^\mathsf{T}. \tag{111}$$

Target motion is modelled by a nearly constant velocity model, that is

$$\pi_{k|k-1}(\mathbf{x}|\mathbf{x}') = \mathcal{N}(\mathbf{x}; \mathbf{F}_k\mathbf{x}' - \mathbf{U}_{k+1,k}, \mathbf{Q}_k) \tag{112}$$

where $\mathbf{F}_k$ is the transition matrix, $\mathbf{U}_k$ is a known deterministic matrix taking into account the effect of observer accelerations. Relevant matrices in (112) are: $\mathbf{F}_k = \mathbf{I}_2 \otimes \mathbf{G}_k$, $\mathbf{Q}_k = \mathbf{I}_2 \otimes \Xi_k$, and

$$\mathbf{U}_{k+1,k} = \begin{bmatrix} x_{k+1}^o - x_k^0 - T_k\dot{x}_k^o \\ \dot{x}_{k+1}^o - \dot{x}_k^o \\ y_{k+1}^o - y_k^0 - T_k\dot{y}_k^o \\ \dot{y}_{k+1}^o - \dot{y}_k^o \end{bmatrix}, \tag{113}$$

where $\otimes$ is the Kroneker product, $\mathbf{I}_2$ is 2×2 identity matrix, while T_k, $\mathbf{G}_k$ and Ξ_k have been defined in Sec.IX-A. By adopting $T_k = T = const$, notation simplifies to $\mathbf{F}_k = \mathbf{F}$ and $\mathbf{Q}_k = \mathbf{Q}$. The model of target appearance and disappearance has been described in Sec.III-A.

The measurement model is described in Sec. V-A. The measurement space is the interval of bearings measurements $\mathcal{Z} = (-\pi, \pi]$, with the (conventional) likelihood function $g_k(z|\mathbf{x})$ for every $z \in Z_k = \{z_{k,1}, \ldots, z_{k,m_k}\}$ adopted as:

$$g_k(z|\mathbf{x}_k) = \mathcal{N}(z; h(\mathbf{x}_k), \sigma_w^2) \tag{114}$$

where the measurement function

$$h(\mathbf{x}_k) = \text{atan2}(x_k, y_k) \tag{115}$$

is the four-quadrant inverse tangent function, taking values in $\mathcal{Z}$. The probability of target detection is assumed independent of the state, i.e. $p_d(\mathbf{x}) = p_d$. The model of false detections is given by (45).

The Bernoulli PF has been implemented for this application following the pseudo-code in Alg.2. The importance density in line 4 of Alg.2 is the transitional prior (hence the Bernoulli PF will require a large number of particles). Target birth density $b_k(\mathbf{x};\mathbf{Z}_k)$, which features in line 18 of Alg.2, is approximated by an equally weighted Gaussian mixture. Each component of this mixture density corresponds to a bearing measurement $z \in Z_k$, and is designed using the standard initialisation technique for bearings only tracking, explained for example in Sec. 4.1.1 of [69]. The initialisation technique is based on three parameters: the initial target range $\bar{r}$, its standard deviation $\sigma_{\bar{r}}$, and target velocity standard deviation $\sigma_{\dot{x}} = \sigma_{\dot{y}} = \sigma_v$.

The testing scenario is shown in Fig. 7. Its total duration is 30 minutes, the sampling interval is $T = 30$s (i.e. $k = 1,\ldots,60$). The target exists throughout the observation period. Initially it is 8 km away from the observer, it maintains the course of -130° and travels at a constant speed of 7 knots. The observer is traveling at the speed of 5 knots and its course during the first leg of its trajectory is 140°. At the end of the first leg (after 30 scans), the observer makes a manoeuvre and chooses a new course of 20°. The measurement standard deviation, which features in (114), is $\sigma_w = 1^\circ$, the probability of detection $p_d = 0.9$ and the average number of false detections per scan is $\lambda = 1$. The model of clutter spatial distribution is uniform, i.e. $c(z) = (2\pi)^{-1}$.

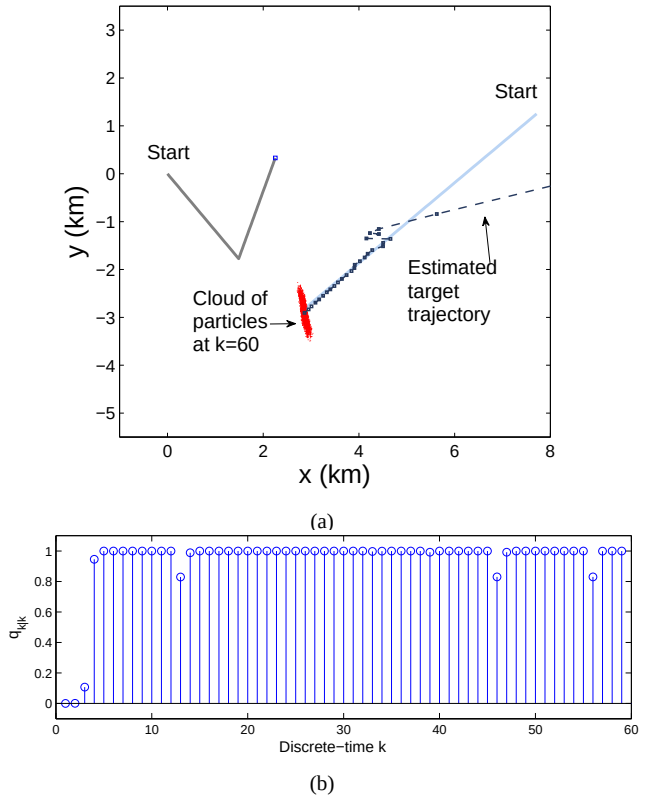


Figure 7. Bearings-only tracking application: (a) observer/target trajectories and the Bernoulli PF estimated trajectory at $k = 60$; (b) the probability of target existence $q_{k|k}$ versus k

The parameters of the Bernoulli PF are as follows: probability of survival $p_s = 0.98$, birth probability $p_b = 0.01$, process noise intensity $\varpi = 0.2$, number of particles $N = B = 5000$, the parameters of $\beta_k(\mathbf{x}|\mathbf{Z}_k)$ are $\bar{r} = 11$ km, $\sigma_{\bar{r}} = 3.5$ km and $\sigma_v = 2.6$ m/s. The estimated target trajectory indicates a good agreement with the ground truth after the observer has manoeuvered. The probability of existence grows to 1 after only a few time steps and remains high throughout the observation interval. Occasionally, when the target detection is missing, it drops to about 0.8. The occasional false detections do not affect significantly the performance of the Bernoulli PF. More technical details about this application, including the observer motion control, can be found in [17].

C. Video tracking of an extended object

There are many features one can use for video tracking, such as colour, shape, motion, body/head pedestrian detections, etc [40]. In this section we explore tracking of a moving object using, as features, corner detections [70]. The dataset (PETS 2000, [71]) consist of 227 image frames, each of size 768×576 pixels. The goal is to detect and track a moving car, which enters the scene from the right and during the observation period changes its size and aspect ratio. The object of interest begins entering the scene at frame $k = 12$ and leaves the scene at frame $k = 225$. OpenCV [41, Ch.10] implementation of the Shi-Tomasi corner detector was used in creating detection sets $\mathbf{Z}_k$, $k = 1, 2, \ldots, 227$. Detections are filtered using a clutter map created from 30 frames of background images. In this way, detections which arise in areas deemed to be part of the background, are likely to be removed. Four image frames of the sequence are shown in Fig.8.

Corner detections, shown as green dots in Fig.8, are instances of detector-output measurements for an extended target, discussed in Sec.VI. The state vector consists of the object centroid position $\mathbf{p} = [x, y]^\intercal$, its velocity vector $\boldsymbol{\upsilon} = [\dot{x}, \dot{y}]^\intercal$ and the ellipsoidal shape defined by vector $\boldsymbol{\theta} = [a, b, c]^\intercal$. Similarly to [72], $\boldsymbol{\theta}$ defines the elements of the covariance matrix which determines the spread of a Gaussian PDF with mean $\mathbf{p}$. Thus a and b are the diagonal elements of this covariance matrix, while c determines the cross-covariance term (effectively the orientation of the ellipse). The state vector thus consists of 7 components: $\mathbf{x} = [\mathbf{p}^\intercal \ \boldsymbol{\upsilon}^\intercal \ \boldsymbol{\theta}^\intercal]^\intercal$. Target dynamics is described by the transitional density $\pi_{k|k-1}(\mathbf{x}|\mathbf{x}') = \mathcal{N}(\mathbf{x}; \mathbf{F}\mathbf{x}', \mathbf{Q})$, where $\mathbf{F}$ and $\mathbf{Q}$ are selected so that the centroid of the target moves with nearly constant velocity, while the shape parameters evolve according to random walk [67]. The model of target appearance and disappearance has been described in Sec.III-A.

Measurements $\mathbf{z} \in \mathbf{Z}_k$ represent coordinates of corner detections on image frame k. The (conventional) likelihood function $g(\mathbf{z}|\mathbf{x}_k)$, which features in update equations of the Bernoulli filter (69) and (70), is replaced by the generalised likelihood $\tilde{g}(\mathbf{z}|\mathbf{x}_k)$. This generalised likelihood is expressed as an indicator function $\tilde{g}(\mathbf{z}|\mathbf{x}_k) = P\{\mathbf{z} \in \mathbf{S}_{\mathbf{x}_k}\} = \mathbb{I}_{\mathbf{S}_{\mathbf{x}_k}}(\mathbf{z})$, where $\mathbf{S}_{\mathbf{x}_k}$ is the random closed set specified by the ellipsoid in the image plane, whose center is $\mathbf{p}_k$ and whose contour is determined by $\boldsymbol{\theta}_k$. This practically means that if a corner detection is inside the extended object (specified by the ellipse), its likelihood is one, otherwise it is zero. The probability of

target detection p_d is assumed independent of the state; the model of false detections is given by (45).

The Bernoulli PF for this application has been implemented following the mains steps of Alg.2, but using update equations (69) and (70). The location of birth particles at k is randomly generated from a Gaussian distribution whose mean equals the mean of the measurements in subsets $\Omega \in \mathcal{P}(\mathbf{Z}_k)$. The velocities of target-birth particles are randomly sampled from a uniform distribution from $-v_{\max}$ to $+v_{\max}$. The shape vector of target-birth particles is also Gaussian distributed with the mean corresponding to the elements of the covariance matrix of measurements in $\Omega \in \mathcal{P}(\mathbf{Z}_k)$. If Ω is a singleton, then the shape vector represents a circle with a small random diameter.

The number of scattering points, that is *corners* in this context, L_k, is estimated from as follows:

$$\widehat{L}_k = \min \left\{ \left\lfloor \frac{|\mathbf{Z}_k| - \lambda}{p_d} \right\rceil, L^* \right\}. \tag{116}$$

where $\lfloor \cdot \rceil$ denotes the nearest integer function, λ and p_d have been defined in Sec. VI, and $L^* > 0$ is adopted as a trade-off between the computational speed and accuracy.

In order to reduce the computational load of the algorithm (which is heavy because the algorithm deals with all partitions of the measurement set), validation or *gating* of measurements is introduced. This means that a subset $\mathbf{Z}_k^* \subseteq \mathbf{Z}_k$ is selected as follows: $\mathbf{z} \in \mathbf{Z}_k$ is included in $\mathbf{Z}_k^*$ if $\sum_{i=1}^{N+B} \tilde{g}_k(\mathbf{z}|\mathbf{x}_{k|k-1}^{(i)}) > \eta$, where $\eta \geq 0$ is a user specified threshold. The update step of the Bernoulli filter is then carried out using $\mathbf{Z}_k^*$.

The values used in the Bernoulli PF for this applications are: $p_b = 0.01$, $p_s = 0.95$, $\lambda = 0.5$, $p_d = 0.65$, $L^* = 5$, $\eta = 50$, $N = 5000$, $B = 250 \cdot |\mathcal{P}_{1:L_k}(\mathbf{Z}_k)|$, $c(\mathbf{z})$ is the uniform density over the image frame and $v_{\max} = 50$. While the density of false detections was very low and therefore relatively easy to handle, the main challenges with this dataset are: (1) the variation of the size of the object over time (especially as the object grows as it enters the scene); (2) the variation of object speed (starts at approximately 150 pixels/s, to drop towards the end to only 5 pixels/s). In order to deal with this variability the algorithm was applied with large amount of process noise.

The estimated probability of object existence $q_{k|k}$, shown in Fig.9.(a), demonstrates reliable detection of both appearance and disappearance of the object of interest. Fig.9.(b) displays the estimated trajectory of the centroid of the object of interest (blue solid line), as well as all corner detections accumulated over the observation period of 227 frames. Due to the large amount of process noise, the trajectory is not smooth, but it follows fairly accurately the cluster of detections. This can be also seen in Fig.8, which displays four image frames of the video sequence, with overlayed object size/shape estimates. As the object enters the scene (frames 14), both the object and its the estimate grow rapidly. As the object moves far away, both its speed and size slowly reduce, which is correctly reflected

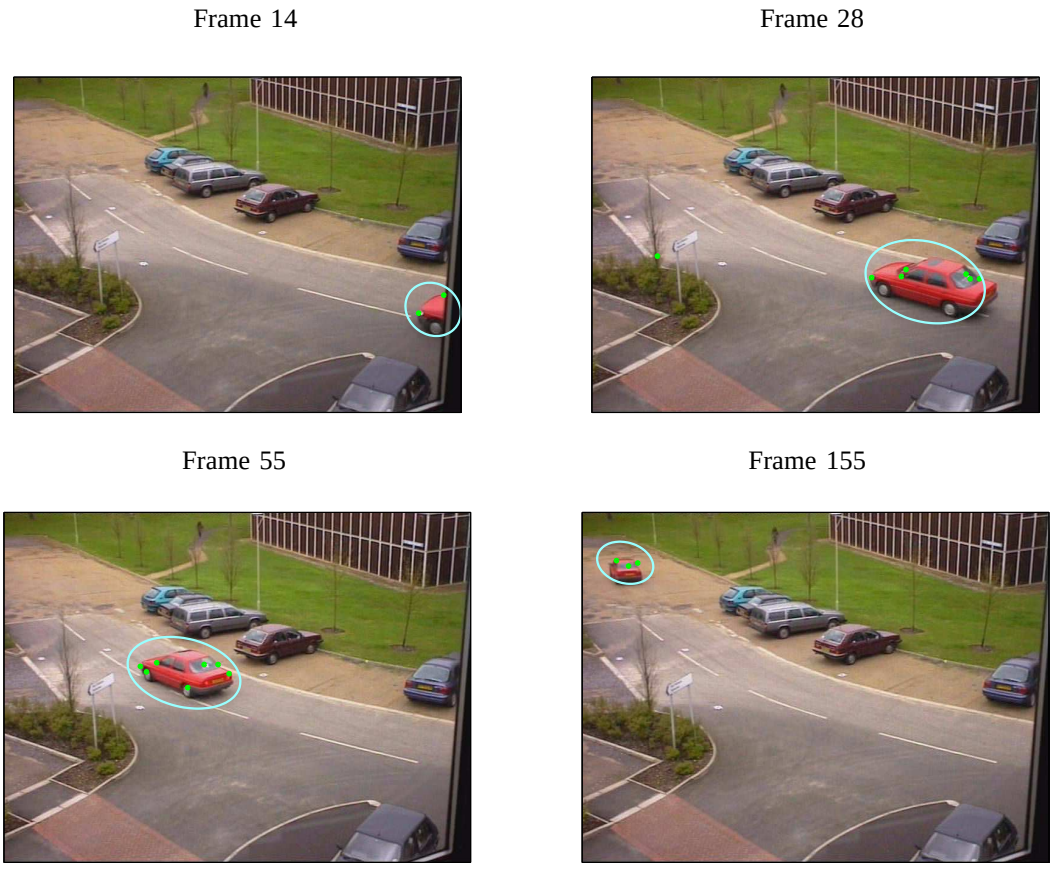
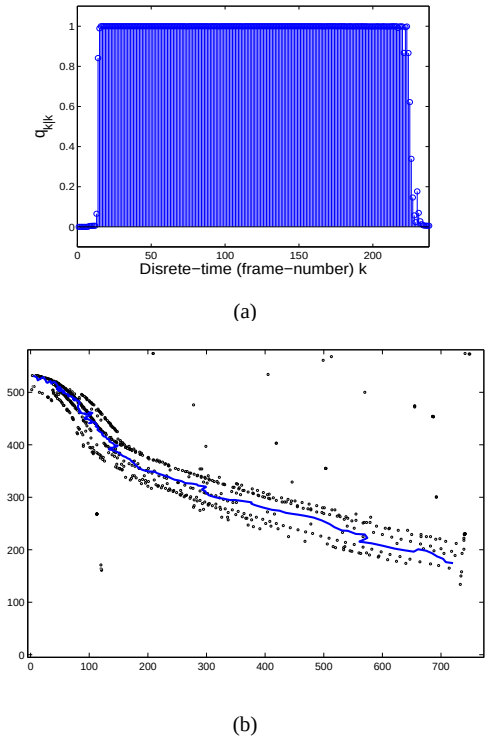


Figure 8. Four out of 227 image frames demonstrating detection and tracking of the red car (PETS 2000 dataset): green dots are corner detections; each ellipse (cyan solid line) represents the estimate of the extended object.

by the estimates.



(a)

(b)

Figure 9. Video surveillance sequence:(a) Estimated probability of existence $q_{k|k}$ over time; (b) Estimated trajectory of the object centrod (blue solid line) and all corner detections (circles)

D. Detection and prediction of an epidemic

The progress of an epidemic in many cases can be described by mathematical models that involve only a few parameters. One class of such models, called compartmental models, is based on a premise that population can be divided into the following classes in relation to the disease: susceptible (S), infectious (I) and recovered or removed (R). Susceptible individuals have never come into contact with the disease. They are able to catch the disease and thus to move to compartment I. All infectious individuals eventually recover (or die) and thus move into compartment R.

Let the number of susceptible, infectious and recovered be denoted by S, I and R, respectively, so that $S+I+R=P$, where P is the total population size. The dynamic model of an epidemic progression in time can be expressed by two stochastic differential equations [73], [74] and the "conservation" law for the population as follows:

$$\frac{ds}{dt} = -\beta\, i\, s^{\nu} + \sigma_q \xi, \tag{117}$$

$$\frac{di}{dt} = \beta\, i\, s^{\nu} - \gamma i - \sigma_q \xi + \sigma_\gamma \zeta, \tag{118}$$

$$r = 1 - s - i. \tag{119}$$

where: $s = S/P$, $i = I/P$, $r = R/P$ are normalised compartment sizes; ξ, ζ are two uncorrelated white Gaussian noise processes, both with zero mean and unity variance; γ is the recovery time (i.e. a reciprocal of the average infectious period for the disease); $\nu \geq 0$ is the population 'mixing' parameter and β is the *contact rate* parameter. The case $\nu = 1$ corresponds to the 'well-mixed' approximation, while for $\nu = 0$ we obtain a linear system without any social interactions (self-isolation). The *contact rate* is a product $\beta = \rho{\cdot}\gamma$, where ρ is the *basic reproductive number* (an important parameter for epidemiologists, representing the average number of people infected by a direct contact with a sick person). The model in (117)-(119) is referred to as the modified SIR (mSIR) epidemic model. When we deal with a large number of individuals P, standard deviations of demographic noise can be approximated as [75], [74]:

$$\sigma_q \approx \sqrt{\frac{\beta\, i\, s^{\nu}}{P}}, \quad \sigma_\gamma \approx \sqrt{\frac{\gamma i}{P}}. \tag{120}$$

Fig.10 shows by a solid line a realisation of the dynamic system described by (117)-(119) during an interval of 200 days. The system was implemented in discrete-time using the Euler approximation with integration interval $\tau = 1/48$. One person in a community of $P = 10000$ people was infected by a contagious disease on the 30th day of the observation period. The parameters used in simulations were: $\nu = 1.56$, $\gamma = 0.098$ (i.e. infectious period of 10.2 days) and $\beta = 0.2549$ (i.e. $\rho = 2.6$). Fig.10 displays the normalised number of infected people over time.

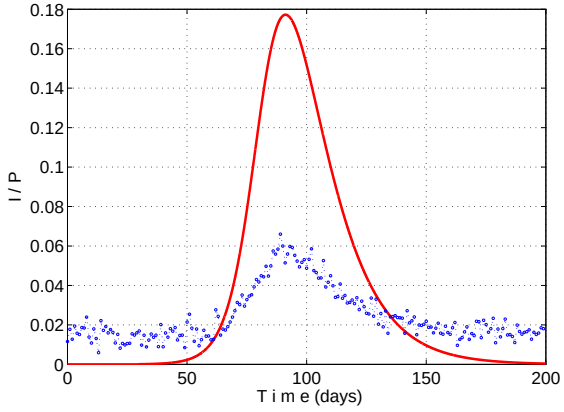


Figure 10. The normalised number of infected people over time (solid red line) and a sample of non-medical measurements (blue dotted line)

The measurements related to the number of infected are assumed to come from non-medical sources, such as Google searches and/or twitter messages with particular keywords [76], [77]. Following [76] one can assume that a power law relationship holds for odds-ratios of observations and the number of infected people:

$$\frac{z_k^{(s)}}{1 - z_k^{(s)}} \propto \left(\frac{i_k}{1 - i_k}\right)^{h_s}, \tag{121}$$

where $z_k^{(s)}$ is the measurement with index (e.g. keyword) $s = 1, \ldots, n$ (normalised by population size P), i_k is the normalised number of infected at discrete-time k and h_s is the power-law exponent. Since at the initial stage of epidemic $i_k \ll 1$, $z_k^{(s)} \ll 1$, eq.(121) can be reduced to a simple power-law model: $z_k^{(s)} = b_s\, i^{h_s} + w_k^{(s)}$, where $b_s > 0$ and $w_k^{(s)}$ is

measurement noise. The described measurement model was treated in Sec.IV-A, with the likelihood function given by (32). The parameters of measurement noise $w_k^{(s)}$ and coefficients b_s and h_s are typically determined during the calibration phase (see for example the results of the linear regression fits in [76]). We will assume a precise measurement model (although in practice an imprecise model could be more appropriate); for measurement noise we adopt Gaussian distribution with mean $\mu_s > 0$ and variance σ_s^2, so that the likelihoods in (32) are given by:

$$\zeta_1^{(s)}\left(z_k^{(s)}|\mathbf{x}\right) = \mathcal{N}(z;\mu_s + b_s\, i^{h_s}, \sigma_s^2) \quad (122)$$

$$\zeta_0^{(s)}\left(z_k^{(s)}\right) = \mathcal{N}(z;\mu_s, \sigma_s^2). \quad (123)$$

Fig.10 shows by a dotted blue line a sample of measurements collected over the observation interval of 200 days. The measurements are collected once per day, with measurement equation parameters: $b_1 = 0.2$, $h_1 = 0.9$, $\mu_1 = 0.015$ and $\sigma_1 = \mu_1/4$.

Numerical results using the normalised number of infected people from Fig.10 as the truth and $n = 2$ sources of measurements, with parameters $b_1 = b_2 0.2$, $h_1 = h_2 = 0.9$, $\mu_1 = \mu_2 = 0.015$ and $\sigma_1 = \sigma_2 = \mu_1/4$, are shown in Fig.11. The Bernoulli PF was implemented according to the pseudo-code in Alg.1, using $N = B = 2000$ particles and TPM parameters $p_s = 0.9$ and $p_b = 0.01$. The state vector contains five components, i_k, s_k and dynamic model parameters β_k, γ_k and ν_k. The model parameters are included in the state vector because their values may not be known precisely (for example, if the disease is caused by a new virus, then β and γ are unknown; ν depends on social interactions in a community and is difficult to measure). Newborn particles are drawn as follows. Particles $i_k^{(i)}$, $i = N+1,\ldots,N+B$ are drawn from a Gaussian with mean $[(z_k^{(s)} - \mu_s)/b_s]^{1/h_s}$ and standard deviation $(\sigma_s/b_s)^{1/h_s}$, with all negative samples set to zero. Particles $s_k^{(i)}$ simply equal to $1 - i_k^{(i)}$. Model parameters are drawn from uniform densities as: $\beta_k^{(i)} \sim \mathcal{U}[0.156, 0.39]$, $\gamma_k^{(i)} \sim \mathcal{U}[0.06, 0.15]$ and $\nu_k^{(i)} \sim \mathcal{U}[0.9, 2]$. Fig.11.(a) displays the estimated probability of existence of the epidemic, $q_{k|k}$, using the described non-medical measurements. The epidemic is confidently detected at day 68. Prediction of the size of epidemic is shown in Fig.11.(b) after day $k = 75$. The short green line in Fig.11.(b) indicates the estimate of i_k, bearing in mind that estimation starts from the time epidemic was detected, i.e. $k = 68$. The dashed red line is the true curve for i_k (the same as in Fig.10). The two solid lines in Fig.10.(b), starting from day 75, indicate the 95% confidence band for the prediction of the epidemic. Note how accurately the timing of the epidemic peak is predicted. The width of the confidence band depends on the uncertainty in model parameters β_k, γ_k and ν_k. More technical details of this application can be found in [78].

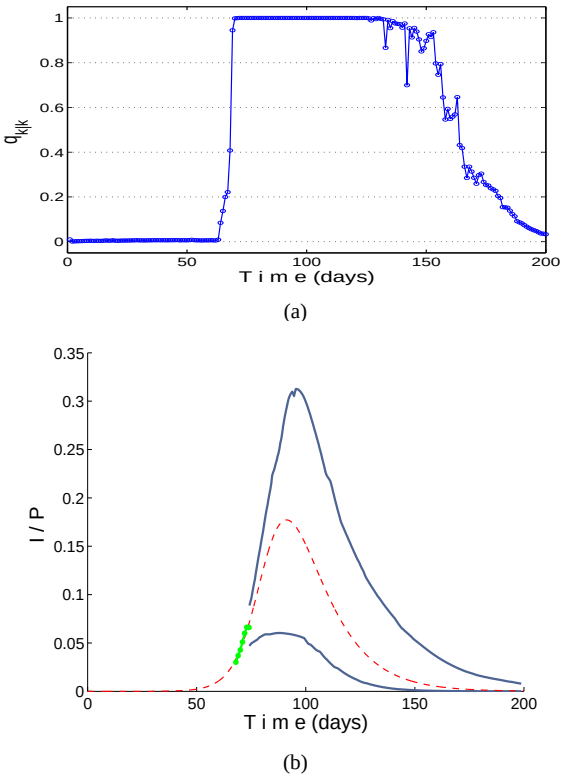


Figure 11. Detection and prediction results: (a) the probability of existence of the epidemic over time; (b) The prediction of epidemic from day $k = 75$: dashed line is the true curve (the same as in Fig.10); the two solid lines starting from day 75 indicate the 95% confidence band

E. Detection and tracking using natural language statements

The problem is detection, positioning and tracking of a dynamic object using measurements expressed by spatially referring natural language (NL) statements. The NL statements are inherently imprecise and ambiguous as they depend greatly on the context and grounding of the referenced subjects [79], [80]. Suppose $Z_k = \{\zeta_{k,1},\ldots,\zeta_{k,m_k}\}$ denotes a set of spoken propositions concerning the state of the object of interest at time k. In order to illustrate the concept, consider a very simple form of a NL statement:

$$\zeta_a = \text{the } \textit{target} \text{ is } \textit{near} \text{ the } \textit{anchor} \text{ a}, \quad (124)$$

where $a \in \mathcal{A}$ is an anchor whose location $\mathbf{r}_a$ is assumed known ($\mathcal{A}$ is the set of anchors). Here the spatial relationship is expressed using the word *near*. Other spatially referring relationships can be added in a more realistic case, such as *behind*, *in front of*, *inside*.

We demonstrate the Bernoulli PF for this application in the context of a detector-output measurement model of Sec.V-A with imprecise measurements of Sec.VII-A. The assumption is that a speech recognition system and a parser are available to automatically transform the spoken propositions of the form (124) into their corresponding generalised likelihood functions, which are then used by the Bernoulli filter. The demonstration is based on a simulation which mimics a busy public space (e.g. airport, railway station) populated by many pedestrians and observers who report their sightings of a person of interest (the target) using the NL statements such as (124). Observer statements are received continuously, but processed every T seconds. The observers sometimes make false detections and occasionally miss the target. The scenario

is illustrated in Fig.12.(a): the solid lines represent the corridor walls while the anchors are indicated by blue circles. The anchors can be, for example, the red bench, the news stand, the clock, etc.

The Bernoulli PF was implemented for this application based on pseudo-code in Alg.2. The state vector of the moving object at time k is $\mathbf{x}_k = \begin{bmatrix} x_k & \dot{x}_k & y_k & \dot{y}_k \end{bmatrix}^\intercal$, where $\mathbf{p}_k = \begin{bmatrix} x_k & y_k \end{bmatrix}^\intercal$ and $\boldsymbol{\upsilon}_k = \begin{bmatrix} \dot{x}_k & \dot{y}_k \end{bmatrix}^\intercal$ denote object position and velocity, respectively, in the two-dimensional Cartesian coordinate system. Target motion was modelled by the transitional density $\pi_{k|k-1}(\mathbf{x}|\mathbf{x}') = \mathcal{N}(\mathbf{x}; \mathbf{F}_k\mathbf{x}', \mathbf{Q}_k)$, where $\mathbf{F}_k = \mathbf{I}_2 \otimes \mathbf{G}_k$, $\mathbf{Q}_k = \mathbf{I}_2 \otimes \Xi_k$. The model of target appearance and disappearance has been described in Sec.III-A.

Each NL statement $\zeta_a \in Z_k$ of the form (124) is represented by the GLF of (71). For this example we adopt a very simple GLF expressed as an indicator function, i.e.

$$\tilde{g}_k(\zeta_a|\mathbf{x}_k) = \mu_a(\mathbf{x}_k) = \begin{cases} 1, & \text{if } \| \mathbf{p}_k - \mathbf{r}_a \| < 2d_a/3 \\ 0, & \text{otherwise,} \end{cases}$$

where d_a is the distance from anchor a to its nearest neighbour. The results of a single run of the Bernoulli PF are shown in Fig.12. The probability of correct detection was $p_d = 0.9$, the average number of false reporting per time interval $T = 15$s was $\lambda = 0.15$. The target "appears" at $k = 5$ and "disappears" at $k = 60$. Its position at $k = 5$ is $\mathbf{p}_5 = \begin{bmatrix} 295 & 990 \end{bmatrix}^\intercal$ and velocity $\boldsymbol{\upsilon}_5 = \begin{bmatrix} 0.125 & -1.25 \end{bmatrix}^\intercal$. The Bernoulli PF was implemented using $N = B = 5000$ particles with $p_s = 0.98$ and $p_b = 0.02$. The estimated trajectory is shown in Fig.12.(a) by a black solid line. The figure also shows the particles representing the posterior spatial PDF at $k = 15$, $k = 25$ and $k = 55$. The true target state is always inside the support of the posterior spatial PDF. The probability of existence in Fig.12.(b) reliably reflects the period of time when the target is present in the scene.

The tracking accuracy can be improved by incorporating implication rules. For example, we may know that if the alarm is turned on, the target moves south, with probability 0.9. This rule is uninformative if the alarm is off, but when the antecedent is satisfied, then the particles whose velocity vector points south are assigned the weight 0.9, while the remaining particles are assigned weight 0.1. Further technical details related to this application and implication rules can be found in [55].

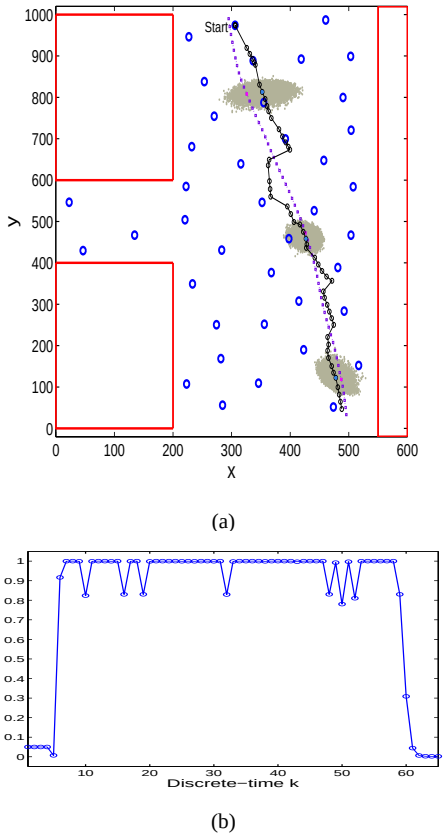


Figure 12. Detection and tracking using NL statements: (a) the scenario with anchors (blue circles) and true and estimated trajectories; (b) the probability of existence

X. Advanced topics for further research

The Bernoulli filter is a relatively recent discovery with many open problems for further research. Some of them are discussed briefly in this section.

A. Fusion of multiple sensors

Sensor network application in Sec. IX-A considered a simple case where fusion was carried out in a centralised manner assuming conditional independence of intensity measurements. The multi-sensor Bernoulli filter for centralised fusion of detector output measurements of a point target has been discussed in [18] and applied to multi-static Doppler-only tracking in [81].

Fusion in large data networks, however, is typically implemented in a distributed (rather than centralised) manner, where each node provides it local state estimate to its neighbours. The estimates from the different nodes are not conditionally independent and, if optimal fusion is to occur, common information has to be "cancelled out" [82]. In most networks, computing the common information is prohibitively expensive and a suboptimal fusion techniques, such as covariance intersection (CI), have attracted significant interest [83]. A theoretical formulation of the CI method for distributed Bernoulli filters, adopting the detector output measurement model, has been discussed in [84]. Other methods for distributed fusion of Bernoulli filters remain largely unexplored.

B. Bernoulli smoothing

The concept of stochastic smoothing [85], [86] is distinguished from stochastic filtering, in the sense that estimation at a specific point in time is to be determined from a batch of measurements, wherein part of the batch can contain data at a *later* time-step than the current. Consequently, there is a delay in producing the state estimate, but more observations on the system are now available.

In the realm of random sets and FISST, smoothing for multi-target states has recently been considered for the detector-output measurement model [87]. In particular, the Bernoulli

forward-backward smoother has been derived in [16]. It consists of forward filtering followed by backward smoothing. It has been shown that both forward filtering and backward smoothing preserve the Bernoulli FISST PDF form. Implementations of the Bernoulli forward-backward smoother have been proposed, via particle filters [16] and Gaussian sum techniques [87].

Efficient Bernoulli particle filters and smoothers for various measurement models described in the paper need to be developed. This includes both the design of the proposal density and the birth density. The rich literature on standard particle filters serves as a guideline for further research [3]–[8].

C. Parameter estimation

The Bernoulli filter, as a sequential Bayesian estimator, depends on mathematical models. In particular, the dynamic motion model, the appearance/disappearance model (Sec. III) and the appropriate measurement model (discussed in Secs. IV, V, VI, VII). These models typically depend on many parameters, such as the probability of birth p_b, probability of survival p_s, process noise standard deviation, probability of detection, false alarm rate, etc.

One way to deal with imprecise measurement model parameters is by using the generalised likelihood functions, as described in Secs. VII-B and IX-A. This approach does not attempt to improve the prior PDF on model parameters. Another approach is to perform the estimation of imprecise model parameters using the measurements, either recursively or as a batch method, during the calibration phase. The topic of parameter calibration for the Bernoulli filter remains unexplored until now, however, useful insights can be found in the literature on particle filters for parameter estimation of dynamic systems [7], [88], [89]. Model parameter estimation, in the context of random finite sets and PHD filters, have also been discussed in [90].

D. Multi-target tracking

The Bernoulli filter, as an exact Bayesian single-object detector/tracker, can be extended to track multiple objects. In the case of the detector-output model and point targets, an approximate Multi-Bernoulli filter was proposed in [10], [25]. Related works using various data association techniques for measurement-to-track assignment were also proposed in [91], [92]. The resulting multi-object trackers are suboptimal, but can be adequate for a number of practical applications [92]. Further developments for Multi-Bernoulli filtering with image measurements were proposed in [29]. The latest development is the exact closed form recursion or conjugate result known as the Generalized Labeled Multi-Bernoulli filter in [14], [93]. Their performance accuracy remains to be compared with other traditional and modern multi-target tracking algorithms.

E. Sensor control

Modern sensors are capable of a variety of actions, such as looking in certain directions, moving to other locations, using different modes of operation, etc. The objective of sensor control is the on-line selection of actions to be taken by individual sensors, in order to maximise the overall utility of the surveillance system.

This class of problems has been studied in the framework of partially observed Markov decision processes (POMDPs) [94]. The elements of a POMDP include the current information state, the set of admissible sensor control vectors (or actions) and the reward function associated with each action. By adopting the information theoretic approach to sensor control, the uncertain information state at time k is represented by the predicted FISST PDF of a Bernoulli RFS $f_{k|k-1}(\mathbf{X}_k|\boldsymbol{\Upsilon}_{1:k-1})$, introduced in Sec.III-B, while the reward function is a measure of *information gain* associated with each action. A particularly useful information measure is the Rényi divergence [95]. The single-step ahead reward function based on the Rényi divergence, in the context of the Bernoulli filter and the detector-output measurement model, was derived in [17]. This reward function was also used for the purpose of sensor selection in the multi-static Doppler-only radar [81]. Other approaches to sensor control involving different measurements models remain to be considered.

F. Performance Bounds

Theoretical performance limits, such as the posterior Cramer-Rao bound, would be of fundamental importance for the Bernoulli filter. Recently a posterior Cramer-Rao type bound for Bernoulli filtering with the detector-output (random finite set) measurement model has been studied in [96], using the Optimal Sub-Pattern Assignment (OSPA) distance [97]. Extending this result to the recursive posterior Cramer-Rao lower bound [98], as well as to other measurement models, are venues for further research. The main difficulty is that the notion of estimation error for a set-valued state is not a straightforward extension of the standard root mean squared error based on the Euclidean distance.

G. Bernoulli filter for switching dynamic models

Once the system is turned on, it may follow one of several dynamic modes of operation, with random switching between them. The switching is typically modelled by a finite state homogeneous Markov chain with known transitional probabilities. Various practical solutions have been proposed in the context of the standard Bayes filter [67, Ch.11], [99, Ch.10], however, to the best our knowledge, a Bernoulli filter for switching dynamic models has not been considered so far.

XI. Summary

The Bernoulli filter is a recent discovery of potential importance in many fields of science and engineering where stochastic dynamic systems (objects, phenomena) of interest can appear and disappear. The key feature of the Bernoulli filter is that it jointly estimates the posterior PDF of the system state and the probability of its existence. This tutorial article reviewed the theory, implementation and several applications of the Bernoulli filter. The update equations of the Bernoulli filter have been derived for various measurement models

encountered in practice. These models have been illustrated by several applications of the Bernoulli filter for detection, estimation and prediction of dynamic systems. The selected applications are diverse in order to demonstrate the power of random set models. We are now in the position to apply the optimal Bayes solution to a much wider spectrum of problems, some of them previously unsolved. The most general implementation of the Bernoulli filter is based on the sequential Monte Carlo method, resulting in a class of Bernoulli particle filters. Finally the article presented an outlook on Bernoulli filters with open problems for further research.

REFERENCES

[1] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Prentice-Hall, 1979.

[3] A. Doucet, J. F. G. de Freitas, and N. J. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[4] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for non-linear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[5] P. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. Bugallo, and J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, pp. 19–38, Sep. 2003.

[6] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle filters for tracking applications*. Artech House, 2004.

[7] O. Cappé, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proc. IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[8] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," Department of Statistics, University of British Columbia, Tech. Rep., Dec. 2008.

[9] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.

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

[11] H. Sidenbladh and S. L. Wirkander, "Tracking random sets of vehicles in terrain," in *Proc. 2nd IEEE Workshop on Multi-Object Tracking*, Madison, WI, USA, June 2003.

[12] T. Zajic and R. Mahler, "A particle-systems implementation of the PHD multitarget tracking filter," in *Proc. SPIE*, vol. 5096, April 2003, pp. 291–299.

[13] 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, Oct. 2005.

[14] B.-T. Vo and B.-N. Vo, "A random finite set conjugate prior and application to multi-target tracking," in *Proc. 7th IEEE Intern. Conf. Intelligent Sensors, Sensor Networks and Information Proc. (ISSNIP)*, Adelaide, Australia, Dec. 2011, pp. 431 –436.

[15] B.-T. Vo, "Random Finite Sets in Multi-Object Filtering," Ph.D. dissertation, School of Electrical, Electronic and Computer Engineering, University of Western Australia, Oct. 2008.

[16] B.-T. Vo, D. Clark, B.-N. Vo, and B. Ristic, "Bernoulli forward-backward smoothing for joint target detection and tracking," *IEEE Trans. Signal Processing*, vol. 59, no. 9, pp. 4473–4477, Sept. 2011.

[17] B. Ristic and S. Arulampalam, "Bernoulli particle filter with observer control for bearings only tracking in clutter," *IEEE Trans. Aeropace & Electronic Systems*, vol. 48, no. 3, pp. 1–11, July 2012.

[18] B.-T. Vo, C.-M. See, N. Ma, and W. T. Ng, "Multi-sensor joint detection and tracking with the Bernoulli filter," *IEEE Trans. Aerospace & Electronic Systems*, vol. 48, no. 2, pp. 1385–1402, 2012.

[19] R. P. S. Mahler, "Multi-target Bayes filtering via first-order multi-target moments," *IEEE Trans. Aerospace & Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, 2003.

[20] 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.

[21] N. P. Whiteley, S. S. Singh, and S. J. Godsill, "Auxiliary particle implementation of the probability hypothesis density filter," *IEEE Trans. on Aerospace & Electronic Systems*, vol. 46, no. 3, pp. 1437–1454, July 2010.

[22] B. Ristic, D. Clark, B.-N. Vo, and B.-T. Vo, "Adaptive target birth intensity for PHD and CPHD filters," *IEEE Trans. Aerospace & Electronic Systems*, vol. 48, no. 2, pp. 1656–1668, April 2012.

[23] R. P. S. Mahler, "PHD filters of higher order in target number," *IEEE Trans. Aerospace & Electronic Systems*, vol. 43, no. 4, pp. 1523–1543, 2007.

[24] 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, 2007.

[25] B.-T. Vo, B. 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, 2009.

[26] S. B. Colegrove, A. W. Davis, and J. K. Ayliffe, "Track initiation and nearest neighbours incorporated into probabilistic data association," *Journal of Electrical and Electronics Engineers, Australia*, vol. 6, no. 3, pp. 191–198, Sept. 1986.

[27] D. Mušicki, R. Evans, and S. Stankovic, "Integrated probabilistic data association," *IEEE Trans. Automatic Control*, vol. 39, no. 6, pp. 1237–1240, June 1994.

[28] Y. Boers, F. Ehlers, W. Koch, T. Luginbuhl, L. D. Stone, and R. Streit, "Track before detect algorithms," *EURASIP Journal on Advances in Signal Processing*, 2008, editorial article for a special issue, article ID 413932.

[29] B.-N. Vo, B.-T. Vo, N.-T. Pham, and D. Suter, "Joint detection and estimation of multiple objects from image observations," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5129–5241, 2010.

[30] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME - Journal of Basic Engineering*, vol. 82 (Series D), pp. 35–45, 1960.

[31] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME - Journ. Basic Engineering*, vol. 83 (Series D), pp. 95–108, 1961.

[32] R. L. Stratonovich, "Conditional Markov processes," *Theory Prob. Appl.*, vol. 5, no. 2, pp. 156–178, 1960.

[33] H. J. Kushner, "On the dynamical equations of conditional probability density functions with applications to optimal stochastic control theory," *Journ. Math. Anal.*, vol. 8, pp. 332–344, 1964.

[34] D. Stoyan, D. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & sons, 1995.

[35] J. Czyz, B. Ristic, and B. Macq, "A particle filter for joint detection and tracking of color objects," *Image and Vision Computing*, vol. 25, pp. 1271–1281, 2007.

[36] Y. Boers and J. N. Driessen, "Multitarget particle filter track before detect application," *IEE Proc. Radar, Sonar Nav.*, vol. 151, no. 6, pp. 351–357, 2004.

[37] X. Sheng and Y.-H. Hu, "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 44–53, Jan. 2005.

[38] O. Rabaste, C. Riché, and A. Lepoute, "Long-time coherent integration for low SNR target via particle filter in track-before-detect," in *Proc. 15th Int. Conf. Information Fusion*, Singapore, July 2012.

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

[40] L. G. Shapiro and G. C. Stockman, *Computer vision*. Prentice Hall, 2001.

[41] G. Bradski and A. Kaehler, *Lerning OpenCV*. O'Reilly Media, 2008.

[42] A. Swain and D. Clark, "The single-group PHD filter: an analytic solution," in *Proc. 14th Int. Conf. Information Fusion*, Chicago, USA, July 2011.

[43] B.-T. Vo, B.-N. Vo, and A. Cantoni, "Bayesian filtering with random finite set observations," *IEEE Trans. Signal Processing*, vol. 56, no. 4, pp. 1313–1326, 2008.

[44] B. Ristic, "Bayesian estimation with imprecise likelihoods: Random set approach," *IEEE Signal Processing Letters*, vol. 18, no. 7, pp. 395–398, July 2011.

[45] N. Patwari, J. N. Ash, S. Kyperountas, R. L. M. A. O. Hero III, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, July 2005.

[46] R. Mahler and A. El-Fallah, "The random set approach to nontraditional measurements is rigorously Bayesian," in *Proc. SPIE*, vol. 8392, 2012.

[47] A. P. Dempster, "The Dempster-Shafer calculus for statisticians," *Int. Jour. Approximate Reasoning*, vol. 48, no. 2, pp. 365–377, 2008.

[48] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.

[49] T. O'Hogan, "Dicing with the unknown," *Significance*, vol. 1, no. 3, pp. 132–133, Sep. 2004.

[50] G. J. Klir and M. J. Wierman, *Uncertainty-based information: Elements of generalized information theory*, 2nd ed. New York: Physica-Verlag, 1999.

[51] A. Gelman, "The boxer, the wrestler and the coin flip: A paradox of robust Bayesian inference and belief functions," *The American Statistician*, vol. 60, no. 2, pp. 146–150, 2006.

[52] I. R. Goodman, R. P. S. Mahler, and H. T. Nguyen, *Mathematics of Data Fusion*. Kluwer, 1997.

[53] A. Gning, B. Ristic, and L. Mihaylova, "Bernoulli particle/box-particle filters for detection and tracking in the presence of triple measurement uncertainty," *IEEE Trans. Signal Processing*, vol. 60, no. 5, pp. 2138–2151, May 2012.

[54] B. Ristic, A. Gning, and L. Mihaylova, "Nonlinear filtering using measurements affected by stochastic, set-theoretic and association uncertainty," in *Proc. 14th Int. Conf. Information Fusion*, Chicago, USA, July 2011.

[55] A. Bishop and B. Ristic, "Fusion of spatially referring natural language statements with random set theoretic likelihoods," *IEEE Trans. Aerospace & Electronic Systems*, 2012, in print.

[56] C. Musso, N. Oudjane, and F. LeGland, "Improving regularised particle filters," in *Sequential Monte Carlo methods in Practice*, A. Doucet, N. deFreitas, and N. J. Gordon, Eds. New York: Springer-Verlag, 2001, ch. 12.

[57] M. R. Morelande and B. Ristic, "Radiological source detection and localisation using Bayesian techniques," *IEEE Trans. Signal Processing*, vol. 57, no. 11, pp. 4220–4231, Nov. 2009.

[58] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, 2nd ed. Springer, 2004.

[59] W. R. Gilks and C. Berzuini, "Following a moving target - Monte Carlo inference for dynamic Baysian models," *Journal of the Royal Statistical Society, B*, vol. 63, pp. 127–146, 2001.

[60] A. Gning, B. Ristic, L. Mihaylova, and A. Fahed, "Introduction to box particle filtering," *IEEE Signal Processing Magazine*, 2012, in press.

[61] F. Caron, P. D. Moral, M. Pace, and B.-N. Vo, "On the stability and the approximation of branching distribution flows, with applications to nonlinear multiple target filtering," *Journal of Stochastic Analysis and Applications*, vol. 29, no. 6, pp. 951–997, 2011.

[62] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.

[63] W. Dargie and C. Poellabauer, *Fundamentals of Wireless Sensor Networks: Theory and Practice*. Wiley, 2010.

[64] F. Zhao, J. Shin, and J. Reich, "Information driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, pp. 61–72, March 2002.

[65] P. M. Djuric, M. Vemula, and M. Bugallo, "Target tracking by particle filtering in binary sensor networks," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2229–2238, 2009.

[66] J. Figueiras and S. Frattasi, *Mobile positioning and tracking*. Wiley, 2010.

[67] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, 2001.

[68] A. Farina, "Target tracking with bearings only measurements," *Signal Processing*, vol. 78, pp. 61–78, 1999.

[69] S. Arulampalam, B. Ristic, N. Gordon, and T. Mansell, "Bearings-only tracking of manoeuvring targets using particle filters," *EURASIP Journal of Applied Signal Processing*, vol. 15, pp. 2351–2365, 2004.

[70] J. Shi and C. Tomasi, "Good features to track," in *Proc. 9th IEEE Conference on Computer Vision and Pattern Recognition*, June 1994, pp. 593–600.

[71] "PETS: performance evaluation of tracking and surveillance," http://www.cvg.rdg.ac.uk/slides/pets.html.

[72] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Trans Aerospace & Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, 2008.

[73] R. M. Anderson and R. M. May, "Population biology of infectious diseases: Part 1," *Nature*, vol. 280, pp. 361–367, 1979.

[74] C. E. Dangerfield, J. V. Ross, and M. J. Keeling, "Integrating stochasticity and network structure into an epidemic model," *J. R. Soc. Interface*, vol. 6, no. 38, pp. 761–774, 2009.

[75] O. Ovaskainen and B. Meerson, "Stochastic models of population extinction," *Trends in Ecology and Evolution*, vol. 25, no. 11, pp. 643–652, 2010.

[76] J. Ginsberg, M. Mohebbi, R. S. Patel1, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, pp. 1012–1015, 2009.

[77] A. Signorini, A. M. Segre, and P. M. Polgreen, "The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic," *PLoS One*, vol. 6, no. 5, p. e19467, 2011.

[78] A. Skvortsov and B. Ristic, "Monitoring and prediction of an epidemic outbreak using syndromic observations," *Mathematical Biosciences*, vol. 240, pp. 12–19, 2012.

[79] W. Hayward and M. Tarr, "Spatial language and spatial representation," *Cognition*, vol. 55, no. 1, pp. 39–84, 1995.

[80] J. Bateman, J. Hois, R. Ross, and T. Tenbrink, "A linguistic ontology of space for natural language processing," *Artificial Intelligence*, 2010.

[81] B. Ristic and A. Farina, "Target tracking via multi-static Doppler shifts," *IET Proc. Radar, Sonar Navig.*, 2012, (in review).

[82] C.-Y. Chong, S. Mori, and K.-C. Chang, "Distributed multitarget multisensor tracking," in *Multitarget-Multisensor Tracking: Advanced Applications*, Y. Bar-Shalom, Ed.

[83] S. J. Julier and J. K. Uhlmann, "Using covariance intersection for SLAM," *Robotics and Autonomous Systems*, vol. 55, no. 7, pp. 3–20, 2007.

[84] D. Clark, S. Julier, R. Mahler, and B. Ristic, "Robust multi-object sensor fusion with unknown correlations," in *Proc. Sensor Signal Processing for Defence (SSPD 2010, IEEE)*, London, UK, Sep. 2010, pp. 1–5.

[85] J. S. Meditch, "A survey of data smoothing," *Automatica*, vol. 9, pp. 151–162, 1973.

[86] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1989.

[87] B.-N. Vo, B.-T. Vo, and R. P. S. Mahler, "Closed form solutions to forward-backward smoothing," *IEEE Trans. Signal Processing*, vol. 60, no. 1, pp. 2–17, Jan. 2012.

[88] C. Andrieu, A. Doucet, and R. Holenstein, "Particle Markov chain Monte Carlo methods," *Journ. Royal Statistical Soc. B*, vol. 72, no. Part 3, pp. 269–342, 2010.

[89] F. Lindsten and T. B. Schon, "On the use of backward simulation in the particle Gibbs sampler," in *Int. Conf. Acoustic, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012, pp. 3845–3848.

[90] B. Ristic, D. Clark, and N. Gordon, "Calibration of tracking systems using detections from non-cooperative targets," in *Proc. IEEE Workshop on Sensor Data Fusion*, Bonn, Germany, Sep. 2012, pp. 25–30.

[91] J. L. Williams, "Experiments with graphical model implementations of multiple target multiple Bernoulli filters," in *Proc. 7th IEEE Intern. Conf. Intelligent Sensors, Sensor Networks and Information Proc. (ISSNIP)*, Adelaide, Australia, Dec. 2011, pp. 532–537.

[92] B. Ristic, J. Sherrah, and A. F. Garcia-Fernandez, "Performance evaluation of random set based pedestrian tracking algorithms," in *Proc. 8th IEEE Intern. Conf. Intelligent Sensors, Sensor Networks and Information Proc. (ISSNIP)*, Melbourne, Australia, April 2013, (Submitted).

[93] B.-T. Vo and B.-N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Trans. Signal Processing*, vol. XX, no. X, p. submitted, Jun. 2012.

[94] D. A. Castanon and L. Carin, "Stochastic control theory for sensor management," in *Foundations and Applications of Sensor Management*, A. O. Hero, D. A. Castanon, D. Cochran, and K. Kastella, Eds. Springer, 2008, ch. 2, pp. 7–32.

[95] A. O. Hero, C. M. Kreucher, and D. Blatt, "Information theoretic approaches to sensor management," in *Foundations and applications of sensor management*, A. O. Hero, D. Castanon, D. Cochran, and K. Kastella, Eds. Springer, 2008, ch. 3, pp. 33–57.

[96] M. Rezaeain and B.-N. Vo, "Error bounds for joint detection and estimation of a single object with random finite set observation," *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1493–1504, Mar. 2010.

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

[98] A. Farina, B. Ristic, and L. Timmoneri, "Cramer-rao bounds for nonlinear filtering with $p_d < 1$ and its application to target tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 8, pp. 1316–1324, Aug. 2002.

[99] F. Gustafsson, *Adaptive Filtering and Change Detection*. Wiley, 2000.