Rao-Blackwellized Monte Carlo Data Association for Multiple Target Tracking

Simo Särkkä
Helsinki University
of Technology
02015 HUT, Finland
simo.sarkka@hut.fi

Aki Vehtari
Helsinki University
of Technology
02015 HUT, Finland
aki.vehtari@hut.fi

Jouko Lampinen
Helsinki University
of Technology,
02015 HUT, Finland
jouko.lampinen@hut.fi

Abstract – *We propose a new Rao-Blackwellized sequential Monte Carlo method for tracking multiple targets in presence of clutter and false alarm measurements. The advantage of the new approach is that Rao-Blackwellization allows the estimation algorithm to be partitioned into single target tracking and data association sub-problems, where the single target tracking sub-problem can be solved by Kalman filters or extended Kalman filters, and the data association by sequential importance resampling. Because the sampled sub-space is finite, it is possible to use the optimal importance distribution explicitly, which significantly reduces the required number of Monte Carlo samples.*

Keywords: multiple target tracking, data association, rao-blackwellization, kalman filter, sequential monte carlo

1 Introduction

The problem of data association makes multiple target tracking much harder task than single target tracking. In multiple target tracking, the algorithm has to estimate which targets produced the measurements, before it is able to use the measurements in actual tracking. If the correct data associations were known, the multiple target tracking problem would reduce to tracking each of the single targets separately.

In this article we propose a new algorithm which is based on separating the multiple target tracking problem into two parts - the estimation of the posterior distribution of the data associations, and the estimation of the single target tracking sub-problems conditional on the data associations. This kind of separation is possible by first solving the joint data association and state estimation problem by sequential Monte Carlo method [1] and then replacing the Monte Carlo integration of the state variables with closed form integration. This Rao-Blackwellization procedure [2] is equivalent to solving the single target tracking problem conditional on the data associations.

The Rao-Blackwellized Monte Carlo data association (RBMCDA) algorithm is derived using Kalman filters [3] as the single target tracking sub-estimators. It can be easily generalized to allow any single target tracking sub-estimator such as the extended Kalman filter (EKF) [4], the unscented Kalman filter (UKF) [5], or the interacting multiple model (IMM) filter [6]. We demonstrate the performance of the proposed algorithm in the presence of the *ghost phenomenon* [7], which refers to the appearance of a ghost target in the virtual crossing of measurements from two angular sensors. Bayesian interpretation [8] of this is that the posterior distribution of target states is multi-modal. Because the data associations are represented as Monte Carlo samples, the multi-modal posterior distributions can be represented without problems.

1.1 Approaches to Multiple Target Tracking

Data association methods in multiple target tracking can be divided into two main classes [7]. *Unique-neighbor data association* methods, such as *multiple hypothesis tracking* (MHT), associate each measurement to one of the previously established tracks. *All-neighbors data association* methods, such as *joint probabilistic data association* (JPDA), use all measurements for updating all the track estimates.

The idea of MHT [7–9] is to associate each measurement to one of the existing tracks, or to form a new track from the measurement. Because this association is not necessarily unique, several hypotheses are continuously formed and maintained. The MHT algorithm calculates the likelihoods of the measurements and the posterior probabilities of the hypotheses, storing only the most probable hypotheses. To enhance the computational efficiency, heuristic methods such as gating, hypothesis merging, clustering and several other strategies can be employed. If some of the target tracks do not get associated measurements for a long period of time, they can be deleted.

JPDA [7,9] approximates the posterior distribution of the targets as separate Gaussian distributions for each target. If the number of targets is T, then T separate Gaussian distributions are maintained. The number of Gaussian distributions is kept constant by integrating over the distribution of data associations of the previous step. This results in algorithm where each of the target estimates gets updated by every measurement with weights that depend on the predicted probabilities of the associations. Gating is used for limiting the number of measurements for each track. If the predicted probabilities are too low for certain targets, those targets are not updated at all. Clutter measurements can be modeled similarly.

Sequential Monte Carlo (SMC) based multiple target tracking methods [10–12] typically belong to the class of *unique-neighbor data association* methods, as they are based on representing the data association and state posteriors as a discrete set of hypotheses. SMC methods can be considered as generalizations of MHT. Instead of maintaining N most probable data association hypotheses, the joint tracking and data association problem is modeled as a Bayesian estimation problem and the *posterior distribution* is estimated with SMC methods. This *particle filtering* approach has the advantage that there are no restrictions for the analytic form of model, although the required number of particles for a given accuracy can be high.

The sequential Monte Carlo data association strategies represented in [10–12] are based on pure particle representation of the joint posterior distribution of states and data associations. In this article we propose how the accuracy and efficiency of SMC methods can be enhanced by closed form marginalization of the state variables. Instead of the pure particle representation this leads to a mixture of Gaussians representation of the joint posterior distribution, which reduces variance and requires less particles for the same accuracy.

2 Optimal Filtering

In this section we review briefly the basic methods of optimal filtering and introduce the notation used. The RBMCDA algorithm, which is based on the reviewed methods, is described later in Sections 3 and 4.

Optimal filtering (see, e.g. [4, 6]), also called Bayesian filtering considers state estimation models of the form

$$\begin{aligned} \mathbf{x}_k &\sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \\ \mathbf{y}_k &\sim p(\mathbf{y}_k \mid \mathbf{x}_k), \end{aligned} \tag{1}$$

where $\mathbf{x}_k$ is the unknown hidden state, which is observable only indirectly through the measurements $\mathbf{y}_k$.

According to the philosophy of optimal filtering, the purpose of the optimal filter is to form an approximate representation of the posterior distribution of the states. The more accurate this representation is, the closer the algorithm is to the optimal performance.

The equations for computing the posterior distributions of the states sequentially in a recursive manner are called the *optimal filtering* or *Bayesian filtering* equations. The recursion starts from an initial distribution $\mathbf{x}_0 \sim p(\mathbf{x}_0)$ and the successive posteriors can be computed from the equations

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1})\, d\mathbf{x}_{k-1}$$
$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) \propto p(\mathbf{y}_k \mid \mathbf{x}_k)\, p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}).$$

2.1 The Kalman Filter

The Kalman filter (KF) (see, e.g. [4, 6]), which originally appeared in [3], considers the special case of the filtering model (1), in which the dynamic and measurements models are linear Gaussian

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{q}_{k-1} & \mathbf{q}_{k-1} &\sim N(0, \mathbf{Q}_{k-1}) \\ \mathbf{y}_k &= \mathbf{H}_k \mathbf{x}_k + \mathbf{r}_k & \mathbf{r}_k &\sim N(0, \mathbf{R}_k). \end{aligned}$$

If the prior distribution is Gaussian, $\mathbf{x}_0 \sim N(\mathbf{m}_0, \mathbf{P}_0)$, then the optimal filtering equations can be evaluated in closed form and the resulting distributions are Gaussian

$$\begin{aligned} p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) &= N(\mathbf{x}_k \mid \mathbf{m}_k^-, \mathbf{P}_k^-) \\ p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) &= N(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k) \\ p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}) &= N(\mathbf{y}_k \mid \mathbf{H}_k \mathbf{m}_k^-, \mathbf{S}_k). \end{aligned}$$

In this article we use the following notation:

- $\mathrm{KF}_p(\cdot)$ denotes the Kalman filter *prediction step*, that is, the calculation of $p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1})$.
- $\mathrm{KF}_u(\cdot)$ denotes the Kalman filter *update step*, that is, the calculation of $p(\mathbf{x}_k \mid \mathbf{y}_{1:k})$.
- $\mathrm{KF}_{lh}(\cdot)$ the evaluation of *marginalized measurement likelihood* $p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1})$.

2.2 The Extended Kalman Filter

The extended Kalman filter (EKF) (see, e.g. [4,6]) is a nonlinear extension of the Kalman filter. The model is given by

$$\begin{aligned} \mathbf{x}_k &= \mathbf{a}(\mathbf{x}_{k-1}, \mathbf{q}_{k-1}) & \mathbf{q}_{k-1} &\sim N(0, \mathbf{Q}_{k-1}) \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k, \mathbf{r}_k) & \mathbf{r}_k &\sim N(0, \mathbf{R}_k). \end{aligned}$$

By linearization of the model, EKF generates Gaussian approximations of the predictive, posterior and marginal likelihood distributions

$$\begin{aligned} p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}) &\approx N(\mathbf{x}_k \mid \mathbf{m}_k^-, \mathbf{P}_k^-) \\ p(\mathbf{x}_k \mid \mathbf{y}_{1:k}) &\approx N(\mathbf{x}_k \mid \mathbf{m}_k, \mathbf{P}_k) \\ p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}) &\approx N(\mathbf{y}_k \mid \boldsymbol{\mu}_k, \mathbf{S}_k). \end{aligned}$$

2.3 Sequential Importance Resampling

Sequential importance resampling (SIR) (see, e.g. [1, 13]), is a generalization of the *particle filtering* framework for the estimation of generic state space models of the form

$$\begin{aligned} \mathbf{x}_k &\sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \\ \mathbf{y}_k &\sim p(\mathbf{y}_k \mid \mathbf{x}_k). \end{aligned}$$

The algorithm uses a weighted Monte Carlo representation of the posterior state distribution. The set of particles is updated and reweighted using a recursive version of importance sampling.

Additional *resampling* step is used for removing the particles with very low weights and duplicating the particles with high weights. The variance introduced by the resampling procedure can be reduced by a proper choice of the resampling method [13].

The performance of the SIR algorithm is dependent on the importance distribution $\pi(\cdot)$, which is an approximation of the posterior distribution of the states given the state on the previous step. *The optimal importance distribution* is

$$\pi(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{y}_{1:k}) = p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{y}_{1:k}).$$

This importance distribution is optimal in the sense that it minimizes the variance of the importance weights.

The *bootstrap filter* is a variation of SIR, in which the dynamic model $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$ is used as the importance distribution. This makes the implementation of the algorithm very easy, but due to the inefficient importance distribution it may require very high number of Monte Carlo samples for accurate estimation results.

2.4 Rao-Blackwellized Particle Filtering

The idea of Rao-Blackwellized particle filtering (see, e.g. [1]) is that sometimes it is possible to evaluate a part of the filtering equations analytically and the other part by Monte Carlo sampling instead of computing everything by pure sampling. According to the Rao-Blackwell theorem this leads to estimators with less variance than what could be obtained by pure Monte Carlo sampling [2]. An intuitive way of thinking this is that marginalization replaces the finite Monte Carlo particle set representation with an infinite closed form particle set, which is always more accurate than any finite set.

3 Tracking Multiple Targets in Clutter

The Rao-Blackwellized Monte Carlo data association (RBMCDA) algorithm proposed here estimates data associations with a SIR filter and the other parts with a Kalman filter. This idea can be directly used in the multiple-targets-in-clutter case, where the dynamic and measurement models of the targets are linear Gaussian, and the only non-linearity is due to the uncertainty in the data associations. If the measurement or dynamic model is non-linear, it is possible to use approximate non-linear filters such as EKF, UKF or IMM instead of the Kalman filter. Of course, in this case the filtering is no longer theoretically optimal.

3.1 Description of the Problem

All targets $j = 1, \ldots, T$ are assumed to have linear dynamic and measurement models

$$\begin{aligned} \mathbf{x}_{j,k} &= \mathbf{A}_{j,k-1}\mathbf{x}_{j,k-1} + \mathbf{q}_{j,k-1} \\ \mathbf{y}_k &= \mathbf{H}_{j,k}\mathbf{x}_{j,k} + \mathbf{r}_{j,k}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{q}_{j,k-1} &\sim N(\mathbf{0}, \mathbf{Q}_{j,k-1}) \\ \mathbf{r}_{j,k} &\sim N(\mathbf{0}, \mathbf{R}_{j,k}). \end{aligned}$$

We do not know which measurements were generated by the targets and which by clutter, and the measurement origins form a random sequence such as:

$$\begin{array}{lcl} \mathbf{y}_1 & : & \text{target 1} \\ \mathbf{y}_2 & : & \text{target 4} \\ \mathbf{y}_3 & : & \text{target 1} \\ \mathbf{y}_4 & : & \text{clutter} \\ \mathbf{y}_5 & : & \text{target 2} \\ \mathbf{y}_6 & : & \text{target 4} \\ \mathbf{y}_7 & : & \text{clutter} \\ & \vdots & \end{array}$$

Our goal is to estimate the *posterior distributions* of the states of all the targets at each time step k.

3.2 Assumptions

The assumptions in this section state the conditions when the algorithm works best, that is, closest to the optimal filter. The performance of the algorithm depends on how well they are met. These assumptions are not very restrictive in practice, because they are quite much the same as the typical ones in tracking literature [9] and they can be overcome by extending the algorithm (and model), for example, by adding the adaptiveness of certain model parameters [14].

In order to derive the algorithm, we need to make some general assumptions about the nature of the estimation problem:

- The posterior distribution of the association history can be represented as a weighted sample of finite number of association histories. This is probably a quite good approximation, because there actually exists *exactly one* true association history. From the estimation point of view, however, our goal is *not* to estimate that one truth, but to form a representation of the information that can be inferred from the measurements – the posterior distribution.

- Given a state and an association history, every new measurement is conditionally independent of the old measurement history. This is a common assumption in statistical inference, including estimation theory [6]. It can, in principle, always be achieved by redefining the model such that it includes the assumed correlations and other dependencies.

- Every new association is independent of the old association histories. This simply states that even if we knew the associations up to this point, we do not know what the next measurement will be associated with. The time dependencies could be allowed with quite small modification to the algorithm, but this is not discussed in this article.

- We are not restricted to synchronized sensors. Instead, we use the more general and realistic asynchronous sensor assumption.

- Measurements are processed one at a time in sequential fashion instead of parallel fashion. The sequential and parallel update schemes are mathematically equivalent [9].

- For the above reason, we do not perform inference on the number of clutter measurements at one time step. Thus, we also do not need a Poisson approximation for the number of clutter measurements as in [9].

- Dynamic models are treated as discrete-time models because integration of a continuous time model from measurement to measurement leads to discrete-time Markov models [4]. We only need to derive algorithms for discrete Markov dynamics and this easily generalizes to the continuous dynamics case.

- As the measurements are assumed to arrive at irregular, discrete instances of time, we have to assume that our dynamic models may depend on time (for example, $\mathbf{A}_k$ instead of $\mathbf{A}$).

The additional explicit mathematical assumptions used in the algorithm derivation are the following:

- The clutter originated measurements are uniformly distributed in the measurement space with volume V. The probability density is given as

$$p(\mathbf{y}_k \mid \mathbf{X}_k,\ \mathbf{y}_k \text{ is clutter}) = 1/V,$$

  where $\mathbf{X}_k$ denotes the stacked vector of all target states at the time step k.

- The target originated measurements are linear functions of the state with normally distributed measurement errors. The probability density is

$$\begin{aligned} &p(\mathbf{y}_k \mid \mathbf{X}_k,\ \mathbf{y}_k \text{ is from target } j) \\ &\quad = N(\mathbf{y}_k \mid \mathbf{H}_{j,k}\mathbf{x}_{j,k}, \mathbf{R}_{j,k}). \end{aligned}$$

- Target dynamics can be modeled as Gauss-Markov random sequences

$$p(\mathbf{x}_{j,k} \mid \mathbf{x}_{j,k-1}) = N(\mathbf{x}_{j,k} \mid \mathbf{A}_{j,k-1}\mathbf{x}_{j,k-1}, \mathbf{Q}_{j,k-1}).$$

- The prior distributions of the target states are Gaussian

$$p(\mathbf{x}_{j,0}) = N(\mathbf{x}_{j,0} \mid \mathbf{m}_{j,0}\mathbf{P}_{j,0}).$$

  These prior distributions do not need to be very informative, but if there is information available on the initial states of the targets, it can be encoded into the prior distributions.

- If the prior probabilities of clutter and target associations are known, they can be included in the model. For example, if the clutter density is 10% and the 3 targets have equal prior association probabilities, these prior probabilities are

$$\begin{aligned} p(c_k = 0) &= 0.1 \\ p(c_k = 1) &= 0.3 \\ p(c_k = 2) &= 0.3 \\ p(c_k = 3) &= 0.3. \end{aligned}$$

  If there is no information, the prior probabilities can be set to be the same for all targets and clutter.

4 Derivation of the Algorithm

In order to implement a SIR filter for the data associations, we need the following:

- The *likelihood of a measurement* conditional on all previous measurements, previous data associations and the current data association

$$p(\mathbf{y}_k \mid c_k, \mathbf{y}_{1:k-1}, c_{1:k-1}).$$

- The *predictive probability*

$$p(c_k \mid c_{1:k-1}),$$

  which gives the prior probability of the data associations given the old data associations.

- The *optimal importance distribution* is useful for constructing an efficient SIR filter. The optimal importance distribution is given by

$$p(c_k | \mathbf{y}_{1:k}, c_{1:k-1}).$$

  As we shall see later, in this case we can sample from the optimal importance distribution directly.

Once we have the distributions above, we can implement SIR as follows:

for $i = 1, \ldots, N$

- *Sample* a new association from the optimal importance distribution:

$$c_k^{(i)} \sim p(c_k^{(i)} \mid \mathbf{y}_{1:k}, c_{1:k-1}^{(i)}).$$

- *Calculate* new weights as

$$\begin{aligned} &w_k\ (i) \propto w_{k-1}^{(i)} \\ &\times \frac{p(\mathbf{y}_k \mid c_k^{(i)}, \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)})\, p(c_k^{(i)} \mid c_{1:k-1}^{(i)})}{p(c_k^{(i)} \mid \mathbf{y}_{1:k}, c_{1:k-1}^{(i)})}. \end{aligned}$$

end for

Resample if needed, that is, if the observed variance of the weights is too high [1].

4.1 Data Representation

An association event c_k is represented with an integer variable with $T + 1$ values

$$\begin{aligned} c_k = 0 &\ \Rightarrow\ \text{clutter association at time step } k \\ c_k = 1 &\ \Rightarrow\ \text{target 1 association at time step } k \\ &\ \vdots \\ c_k = T &\ \Rightarrow\ \text{target } T \text{ association at time step } k. \end{aligned}$$

Each of the N particles implicitly contain the state means and covariances for each target on time step k, the whole association history up to time step k, and an importance weight. In practice, because the old associations have no direct effect on the new associations, we only need to store the means, covariances and weights:

$$\begin{aligned} &\text{particle 1}:\ \{\mathbf{m}_{1,k}^{(1)}, \mathbf{P}_{1,k}^{(1)}, \ldots \mathbf{m}_{T,k}^{(1)}, \mathbf{P}_{T,k}^{(1)},\ w_k^{(1)}\} \\ &\qquad\qquad \vdots \\ &\text{particle N}:\ \{\mathbf{m}_{1,k}^{(N)}, \mathbf{P}_{1,k}^{(N)}, \ldots \mathbf{m}_{T,k}^{(N)}, \mathbf{P}_{T,k}^{(N)},\ w_k^{(N)}\}. \end{aligned}$$

4.2 Likelihood of Measurement

In order to derive the SIR filter for the data associations, we need to calculate the measurement likelihood

$$p(\mathbf{y}_k \mid c_k, \mathbf{y}_{1:k-1}, c_{1:k-1}).$$

For each data association history particle i and for each target j at time step k, we have the state distribution

$$p(\mathbf{x}_{j,k-1} \mid \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)}) = N(\mathbf{x}_{j,k-1} \mid \mathbf{m}_{j,k-1}^{(i)}, \mathbf{P}_{j,k-1}^{(i)}).$$

The predicted state distribution for the target j at the time step k is

$$\begin{aligned} p(\mathbf{x}_{j,k} \mid \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)}) &= \\ &= \int p(\mathbf{x}_{j,k} \mid \mathbf{x}_{j,k-1})\, p(\mathbf{x}_{j,k-1} \mid \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)})\, d\mathbf{x}_{j,k-1} \\ &= \int N(\mathbf{x}_{j,k} \mid \mathbf{A}_{j,k-1}\mathbf{x}_{j,k-1}, \mathbf{Q}_{j,k-1}) \\ &\quad \times N(\mathbf{x}_{j,k-1} \mid \mathbf{m}_{j,k-1}^{(i)}, \mathbf{P}_{j,k-1}^{(i)})\, d\mathbf{x}_{j,k-1}. \end{aligned} \tag{2}$$

The last term in Eq. (2) is the Kalman Filter prediction step for the target j

$$p(\mathbf{x}_{j,k}|\mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)}) = N(\mathbf{x}_{j,k} \mid \mathbf{m}_{j,k}^{-(i)}, \mathbf{P}_{j,k}^{-(i)}),$$

where

$$[\mathbf{m}_{j,k}^{-(i)}, \mathbf{P}_{j,k}^{-(i)}] = \mathrm{KF}_p(\mathbf{m}_{j,k-1}^{(i)}, \mathbf{P}_{j,k-1}^{(i)}, \mathbf{A}_{j,k-1}, \mathbf{Q}_{j,k-1}).$$

If the measurement originated from target j, the measurement likelihood is

$$\begin{aligned} p(\mathbf{y}_k \mid c_k = j, \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)}) &= \\ &= \int p(\mathbf{y}_k \mid c_k = j, \mathbf{x}_{j,k}) p(\mathbf{x}_{j,k} \mid \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)}) d\mathbf{x}_{j,k} \\ &= \int N(\mathbf{y}_k \mid \mathbf{H}_{j,k}\mathbf{x}_{j,k}, \mathbf{R}_{j,k}) N(\mathbf{x}_{j,k} \mid \mathbf{m}_{j,k}^{-(i)}, \mathbf{P}_{j,k}^{-(i)}) d\mathbf{x}_{j,k}. \end{aligned} \tag{3}$$

The last term in Eq. (3) is the Kalman Filter likelihood for the target

$$\begin{aligned} &p(\mathbf{y}_k \mid c_k = j, \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)}) \\ &\quad = \mathrm{KF}_{lh}(\mathbf{y}_k, \mathbf{m}_{j,k}^{-(i)}, \mathbf{P}_{j,k}^{-(i)}, \mathbf{H}_{j,k}, \mathbf{R}_{j,k}). \end{aligned}$$

A clutter measurement has the likelihood

$$p(\mathbf{y}_k \mid c_k = 0, \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)}) = 1/V.$$

Thus, the state independent measurement likelihood term is

$$\begin{aligned} &p(\mathbf{y}_k \mid c_k, \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)}) \\ &= \begin{cases} 1/V & \text{if } c_k = 0 \\ \mathrm{KF}_{lh}(\mathbf{y}_k, \mathbf{m}_{1,k}^{-(i)}, \mathbf{P}_{1,k}^{-(i)}, \mathbf{H}_{1,k}, \mathbf{R}_{1,k}) & \text{if } c_k = 1 \\ \quad\vdots & \\ \mathrm{KF}_{lh}(\mathbf{y}_k, \mathbf{m}_{T,k}^{-(i)}, \mathbf{P}_{T,k}^{-(i)}, \mathbf{H}_{T,k}, \mathbf{R}_{T,k}) & \text{if } c_k = T, \end{cases} \end{aligned} \tag{4}$$

where for $j = 1, \ldots, T$

$$[\mathbf{m}_{j,k}^{-(i)}, \mathbf{P}_{j,k}^{-(i)}] = \mathrm{KF}_p(\mathbf{m}_{j,k-1}^{(i)}, \mathbf{P}_{j,k-1}^{(i)}, \mathbf{A}_{j,k-1}, \mathbf{Q}_{j,k-1}),$$

and $\mathbf{m}_{j,k-1}^{(i)}, \mathbf{P}_{j,k-1}^{(i)}$ are the mean and covariance of target j in particle i, which is conditioned on the state history $c_{1:k-1}^{(i)}$.

4.3 Predictive Probability

Because the associations are independent of time, we have

$$p(c_k \mid c_{1:k-1}) = p(c_k).$$

The prior distribution $p(c_k)$ gives the relative probabilities of the target and clutter associations.

4.4 Optimal Importance Distribution

The optimal importance distribution is given by

$$p(c_k \mid \mathbf{y}_{1:k}, c_{1:k-1}^{(i)})$$

for each particle i.

We already know the measurement likelihood $p(\mathbf{y}_k \mid c_k, \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)})$, which is given by Eq. (4). The posterior distribution of c_k can be calculated using the Bayes' rule

$$p(c_k \mid \mathbf{y}_{1:k}, c_{1:k-1}^{(i)}) \propto p(\mathbf{y}_k \mid c_k, \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)})\, p(c_k),$$

where we have used the fact that an association c_k does not depend on the previous measurements $\mathbf{y}_{1:k-1}$ or the associations $c_{1:k-1}^{(i)}$.

We can sample from the optimal importance distribution as follows:

1. Compute the unnormalized clutter association probability
$$\hat{\pi}_0^{(i)} = p(\mathbf{y}_k \mid c_k^{(i)} = 0, \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)})\, p(c_k^{(i)} = 0).$$
2. Compute the unnormalized target association probabilities for each target $j = 1, \ldots, T$
$$\hat{\pi}_j^{(i)} = p(\mathbf{y}_k \mid c_k^{(i)} = j, \mathbf{y}_{1:k-1}, c_{1:k-1}^{(i)})\, p(c_k^{(i)} = j).$$
3. Normalize the importance distribution:
$$\pi_j^{(i)} = \frac{\hat{\pi}_j^{(i)}}{\sum_{j'=0}^{T} \hat{\pi}_{j'}^{(i)}}, \quad j = 0, \ldots, T.$$
4. Sample a new association $c_k^{(i)}$ with the following probabilities:
   - Draw $c_k^{(i)} = 0$ with probability $\pi_0^{(i)}$
   - ...
   - Draw $c_k^{(i)} = T$ with probability $\pi_T^{(i)}$.

4.5 Generalization to Non-Linear Kalman Filters

Any *sufficient statistics* based state estimation algorithm can be used as the sub-estimator instead of the linear Kalman filter. This kind of state estimators are, for example, the extended Kalman filter (EKF), the unscented Kalman filter (UKF) and multiple model based filters such as the interacting multiple model (IMM) filter. The Kalman filter prediction step $\mathrm{KF}_p(\cdot)$, the update step $\mathrm{KF}_u(\cdot)$ and the evaluation of marginalized measurement likelihood $\mathrm{KF}_{lh}(\cdot)$ are replaced by the corresponding steps of the estimator.

5 Simulation Results

5.1 Tracking a Sine Signal in Clutter

In this scenario the true signal is the sine signal

$$x(t) = \sin(\omega t), \tag{5}$$

where the angular velocity ω is only approximately known. Half of the measurements are corrupted by additive Gaussian noise and half of them are completely corrupted so that they can take any value in the sensor's dynamic range, which in this case is $[-2, 2]$.

Assuming a sampling period of Δt, the true signal (5) can be approximately modeled by a discretized white noise acceleration model [6]

$$\begin{pmatrix} x_k \\ \dot{x}_k \end{pmatrix} = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{k-1} \\ \dot{x}_{k-1} \end{pmatrix} + \mathbf{q}_{k-1}.$$

where the Gaussian white noise process $\mathbf{q}_{k-1}$ has moments

$$\begin{aligned} \mathrm{E}[\mathbf{q}_{k-1}] &= \mathbf{0} \\ \mathrm{E}[\mathbf{q}_{k-1}\mathbf{q}_{k-1}^T] &= \begin{pmatrix} \frac{1}{3}\Delta t^3 & \frac{1}{2}\Delta t^2 \\ \frac{1}{2}\Delta t^2 & \Delta t \end{pmatrix} q, \end{aligned}$$

where q is the spectral density of the noise. The state is defined as $\mathbf{x}_k = (x_k \; \dot{x}_k)^T$ where x_k is the value of signal at time step $t_k = t_0 + k\Delta t$ and $\dot{x}_k$ is the derivative of signal at the same time step.

The likelihood of the measurement y_k can be modeled by defining a latent variable c_k, which has the value of 0 if measurement is a corrupted measurement (clutter) and 1 if it is a measurement from the signal. If the measurement is clutter, it is assumed to be evenly distributed in the measurement space $[-2, 2]$ (which is the dynamic range of sensor). The likelihood is

$$p(y_k \mid \mathbf{x}_k, c_k) = \begin{cases} 1/4 & , \quad \text{if } c_k = 0 \\ N(y_k \mid \mathbf{H}\mathbf{x}_k, R) & , \quad \text{if } c_k = 1 \end{cases},$$

where $\mathbf{H} = (1 \; 0)$.

The prior distributions of the signal and its derivative were chosen to be $x_0 \sim N(0, \sigma^2)$ and $\dot{x}_0 \sim N(1, \sigma^2)$ with $\sigma^2 = 0.1$.

Table 1 shows the RMSE results of tracking the simulated sine signals with the following methods:

Table 1: Root mean squared error values for the different methods for tracking a sine signal in 50% clutter. The means (RMSE) and standard deviations (STD) from 10 different simulated data sets are given in the table. The same data sets were used with all the methods.

Method	*RMSE*	*STD*
RBMCDA, 10 particles	0.16	0.02
RBMCDA, 100 particles	0.15	0.01
Bootstrap filter, 1000 particles	2.07	2.31
Bootstrap filter, 10000 particles	0.16	0.02
Kalman filter, assuming no clutter	0.39	0.02
Kalman filter, clutter modeled	0.32	0.03
Kalman filter, perfect associations	0.11	0.01

- *RBMCDA, 10 particles*: Rao-Blackwellized Monte Carlo data association algorithm with 10 particles.
- *RBMCDA, 100 particles*: Rao-Blackwellized Monte Carlo data association algorithm with 100 particles.
- *Bootstrap filter, 1000 particles*: Bootstrap filter with adaptive resampling and 1000 particles, such that the joint distribution of states and data associations is represented as a set of weighted Monte Carlo samples. The high RMSE values are due to filter divergence in many of the test cases.
- *Bootstrap filter, 10000 particles*: The same bootstrap filter as above with 10000 particles.
- *Kalman filter, assuming no clutter*: Kalman filter with the assumption that there are no clutter measurements at all.
- *Kalman filter, clutter modeled*: Kalman filter with increased measurement variance such that the presence of 50% clutter is taken into account.
- *Kalman filter, perfect associations*: Kalman filter with perfect data association knowledge, such that clutter measurements are simply thrown away as would an ideal data association algorithm do.

Typical conditional means of the estimated marginal state distributions when the Rao-Blackwellized Monte Carlo data association method is used are shown in Fig. 1. It can be seen that the estimate follows the true signal trajectory quite well despite the high number of clutter measurements.

5.2 Multiple Target Bearings Only Tracking

Now, we shall consider a classical bearings only multiple target tracking problem, which frequently arises in the context of passive sensor tracking.

The dynamics of target j with the state vector $\mathbf{x}_{j,k} = (x_{j,k} \; y_{j,k} \; \dot{x}_{j,k} \; \dot{y}_{j,k})^T$ can be modeled with a discretized Wiener velocity model

$$\begin{pmatrix} x_{j,k} \\ y_{j,k} \\ \dot{x}_{j,k} \\ \dot{y}_{j,k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{j,k-1} \\ y_{j,k-1} \\ \dot{x}_{j,k-1} \\ \dot{y}_{j,k-1} \end{pmatrix} + \mathbf{q}_{k-1},$$

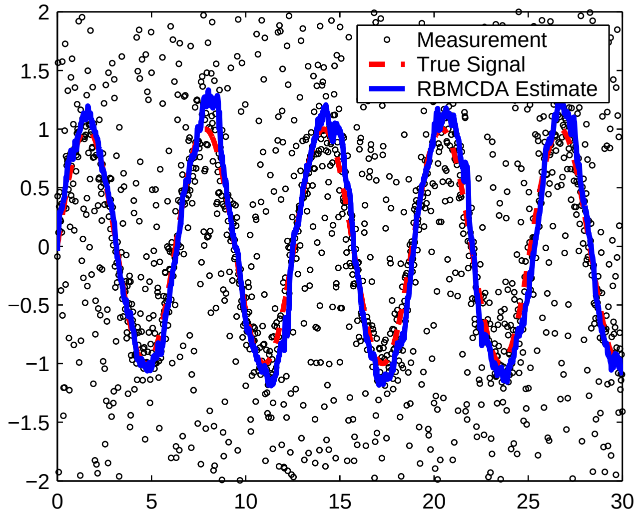

Fig. 1: Result of tracking a sine signal in the presence of 50% clutter measurements with RBMCDA and 100 particles.

where $\mathbf{q}_{k-1}$ is the process noise. The noise in an angular measurement from known target j made by sensor i can be modeled as Gaussian

$$\hat{\theta}_k = \arctan\left(\frac{y_{j,k} - s_y^i}{x_{j,k} - s_x^i}\right) + r_k,$$

where (s_x^i, s_y^i) is the position of sensor i and $r_k \sim N(0, \sigma^2)$.

Because the measurement model is non-linear we replace the Kalman filter in the data association algorithm with sub-optimal EKF. The uncertainty in data associations can be modeled by defining a variable c_k, which has the value $c_k = j$ if the measurement at time step k is associated to target j.

The evolution of the tracking is illustrated in Figs. 2 – 4. The particles in the figures are a random sample drawn from the posterior distribution estimate, used for visualizing the distribution. The actual posterior distribution estimate is a mixture of Gaussians which is hard to visualize directly. The number of Monte Carlo samples used in the Rao-Blackwellized Monte Carlo data association method was 100.

The prior distribution was selected on purpose such that all the four crossings of measurements from the two sensors contain some probability mass, and the distributions of the targets are two-modal as shown in Fig. 2. In the beginning of the tracking the multi-modality can be still seen in the estimated posterior distribution (Fig. 3). This phenomenon is often called the *ghost phenomenon* in tracking literature [7]. When the tracking proceeds, the implicit trajectory restrictions set by the dynamic model force the false modes to become much less probable than the other, causing them to disappear, as can be seen in Fig. 4.

Fig. 5 shows the final tracking result, where the multi-modality can clearly be seen. Again, particles are used for visualizing the distribution, although the true posterior distribution estimate is a mixture of Gaussians. Fig. 5 also illustrates the disadvantage of using the conditional mean

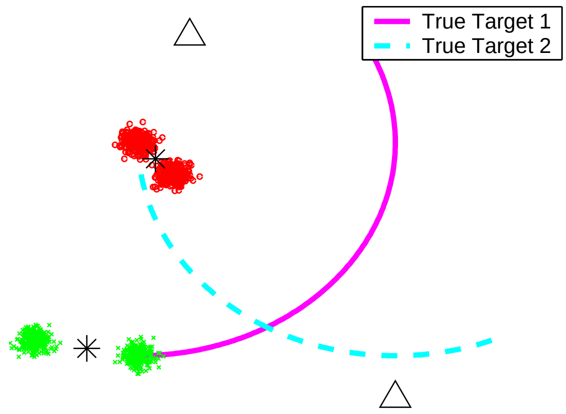

Fig. 2: The prior distributions of the targets. Half of the prior probability mass is located in the ghost sensor measurement crossings.

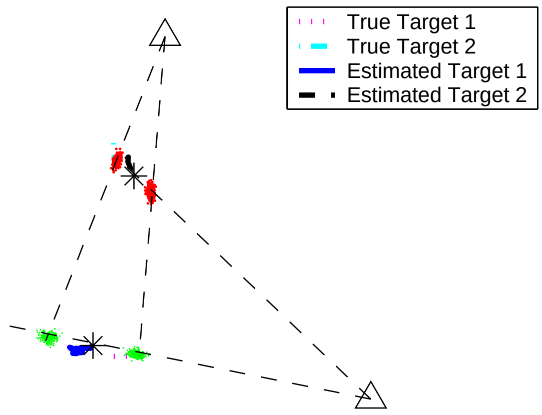

Fig. 3: At the start of tracking the multi-modality of posterior distribution can clearly be seen.

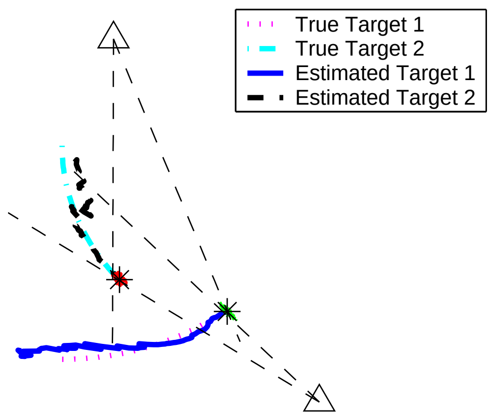

Fig. 4: After a while the posterior distribution becomes unimodal due to the restrictions set by the dynamic model.

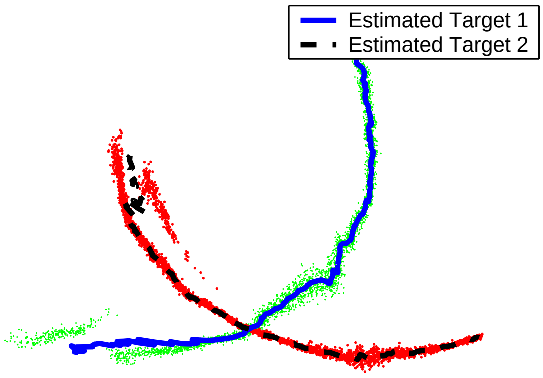

Fig. 5: Filter estimates for each time step. In the beginning of the trajectory the posterior distribution is multi-modal and using the posterior mean as the summary statistic can be misleading.

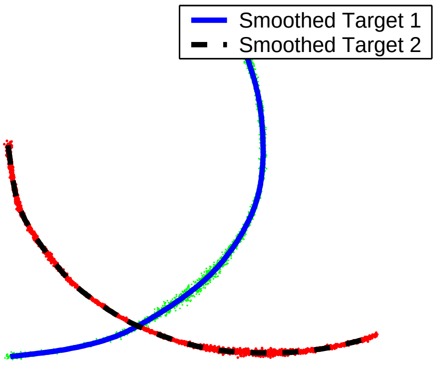

Fig. 6: Smoothed estimates do not have the ambiguity in the beginning of the trajectory, because the later measurements have resolved it. This is because the trajectories with discontinuities are much less probable than the ones without them. These discontinuities are not taken into account in the filter estimates, because filter cannot know about the discontinuities before they occur.

as the state summary in the multi-modal case. When the posterior distribution is two-modal, the conditional mean is *between* the modes, which is a place where the target is known *not* to be.

Fig. 6 shows the *smoothed* tracking result, which is an estimate where the distributions of all time steps are conditioned on all the measurements. This kind of an estimate can be easily calculated with Kalman smoothers [4, 6] and particle smoothers [13] also in the Rao-Blackwellized particle filtering case. Conditional on all the measurements the trajectory no longer contains the ghost phenomenon, because the dynamic model forces the discontinuous trajectories to be much less probable than the continuous ones. These discontinuities can be detected after seeing the future measurements, not before.

6 Conclusions

In this article we have proposed a new multiple target tracking method, which is based on Rao-Blackwellization of the sequential Monte Carlo estimator formulated for the joint tracking and data association problem. The performance of the algorithm was demonstrated in the case of a very high number of clutter measurements, and in the case where the posterior distribution of targets is multi-modal due to the ghost phenomenon in bearings only tracking.

Acknowledgments

Authors would like to thank Toni Tamminen for his comments on the manuscript.

References

[1] Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[2] George Casella and Christian P. Robert. Rao-Blackwellisation of sampling schemes. *Biometrika*, 83(1):81–94, 1996.

[3] Rudolph E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering*, 82:34–45, March 1960.

[4] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[5] Simon J. Julier, Jeffrey K. Uhlmann, and Hugh F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *Proceedings of the 1995 American Control, Conference, Seattle, Washington*, pages 1628–1632, 1995.

[6] Yaakov Bar-Shalom, Xiao-Rong Li, and Thiagalingam Kirubarajan. *Estimation with Applications to Tracking and Navigation*. Wiley Interscience, 2001.

[7] Samuel Blackman and Robert Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House Radar Library, 1999.

[8] Lawrence D. Stone, Carl A. Barlow, and Thomas L. Corwin. *Bayesian Multiple Target Tracking*. Artech House, Boston, London, 1999.

[9] Yaakov Bar-Shalom and Xiao-Rong Li. *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS, 1995.

[10] Neil Gordon. A hybrid bootstrap filter for target tracking in clutter. *IEEE Transactions on Aerospace and Electronic Systems*, 33(1):353–358, January 1997.

[11] Rickard Karlsson and Fredrik Gustafsson. Monte Carlo data association for multiple target tracking. In *IEEE Target tracking: Algorithms and applications, The Netherlands*, Oct 2001.

[12] Carine Hue, Jean-Pierre Le Cadre, and Patrick Perez. The (MR)MTPF: particle filters to track multiple targets using multiple receivers. In *4th International Conference on Information Fusion, Montreal, Canada*, 2001.

[13] Genshiro Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5:1–25, 1996.

[14] Simo Särkkä, Aki Vehtari, and Jouko Lampinen. Bayesian adaptive Kalman filtering and smoothing by separable approximations of posterior distributions. *Submitted*.