A Particle Filter to Track Multiple Objects

Carine Hue Jean-Pierre Le Cadre
IRISA
Campus de Beaulieu
35042 Rennes Cedex, France
chue@irisa.fr lecadre@irisa.fr

Patrick Pérez
Microsoft Research
St George House, 1 Guildhall Street
Cambridge, CB2 3NH, UK
pperez@microsoft.com

Abstract

We address the problem of tracking multiple objects encountered in many situations in signal or image processing. We consider stochastic dynamic systems nonlinearly and uncompletely observed. The difficulty lies on the fact that the estimation of the states requires the assignation of the observations to the multiple targets. We propose an extension of the classical particle filter where the stochastic vector of assignation is estimated by a Gibbs sampler. The merit of the method is assessed in bearings-only context and we present one application in image-based tracking.

1. Introduction

Multiple object tracking (MTT) deals with state estimation of an unknown number of moving targets. Available measurements may both arise from the targets if they are detected, and from clutter. Clutter is generally considered as a model describing false alarms. Its (spatio-temporal) statistical properties are quite different from target ones, which makes possible the extraction of target tracks from clutter. To perform multiple object tracking the observer has at his disposal a huge amount of data, possibly collected on multiple receivers. In signal processing, elementary measurements are receiver outputs, e.g., bearings, ranges, time-delays, Dopplers, etc. In image-based tracking they have to be computed from the images.

But the main difficulty comes from the assignment of a given measurement to a target model. These assignments are generally unknown, as are the true target models. This is a neat departure from classical estimation problems. Thus, two distinct problems have to be solved *jointly*: data association and estimation. As long as the association is considered in a deterministic way, the hypothesis associations must be exhaustively enumerated, which leads to a NP-hard problem (as in JPDAF and MHT algorithms [3] for instance). As soon as the association variables are considered as stochastic variables and moreover statistically independent like in the Probabilistic MHT (PMHT), the complexity is reduced. However, the above algorithms do not cope with non linear state or measurement models and non Gaussian state or measurement noises. Under such assumptions (stochastic state equation and non linear state or measurement equation, non Gaussian noises), particle filters are particularly adapted. They mainly consist in propagating a weighted set of particles which approximates the probability density of the state conditionally to the observations. Particle filtering can be applied under very weak hypotheses, is able to cope with heavy clutter, and is very easy to implement. Numerous versions have been used in various contexts: the bootstrap filter for target tracking in [4], the Condensation algorithm in image analysis [5] are two examples among others. In image analysis a probabilistic exclusion principle has been developed in [7] to track multiple objects but the algorithm is very dependent on the observation model and seems costly to extend for more than two objects. We propose here a quite general algorithm for multiple object tracking applicable both in signal and image analysis.

This work is organised as follows. In section II, we briefly recall the basic particle filter. Section III deals with our extension of the basic filter to multiple objects. Section IV begins with a validation of our algorithm in the passive sonar context, i.e. to estimate the trajectories of multiple "small" targets from their noisy bearings. Then it is used to track pedestrians in a video-sequence.

2. The basic particle filter

For the sake of completeness, the basic particle filter is now briefly reviewed. We consider a dynamic system represented by the stochastic process $(X_t) \in \mathbb{R}^{n_x}$ whose temporal evolution is given by the state equation:

$$X_t = F_t(X_{t-1}, V_t). \tag{1}$$

- Initialization: $\begin{cases} s_0^n \sim p(X_0) \\ q_0^n = 1/N \end{cases} \quad n = 1, \ldots, N.$
- For $t = 1, \ldots, T$:
  - Proposal: sample $\tilde{s}_t^n$ from $f(X_t | X_{t-1} = s_{t-1}^n, Y_t = y_t)$ for $n = 1, \ldots, N$.
  - Weighting: $\begin{cases} \text{Compute un-normalised weights: } \tilde{q}_t^n = q_{t-1}^n \frac{p(\tilde{s}_t^n | s_{t-1}^n)\ l_t(y_t; \tilde{s}_t^n)}{f(\tilde{s}_t^n | s_{t-1}^n, y_t)} \text{ for } n = 1, \ldots, N. \\ \text{Normalise weights: } q_t^n = \frac{\tilde{q}_t^n}{\sum_{n=1}^N \tilde{q}_t^n} \text{ for } n = 1, \ldots, N. \end{cases}$
  - Return $\widehat{\mathbb{E}} g(X_t) = \sum_{n=1}^N q_t^n g(\tilde{s}_t^n)$.
  - Calculate $\hat{N}_{eff} = \frac{1}{\sum_{n=1}^N (q_t^n)^2}$.
  - Resampling: if $\hat{N}_{eff} < N_{threshold}$: $\begin{cases} s_t^n \sim \sum_{k=1}^N q_t^k \delta_{\tilde{s}_t^k} \\ q_t^n = 1/N \end{cases} \quad n = 1, \ldots, N$, else $s_t^n = \tilde{s}_t^n$ for $n = 1, \ldots, N$.

Figure 1. Basic particle filter with adaptive resampling .

It is observed at discrete times via realizations of the stochastic process $(Y_t) \in \mathbb{R}^{n_y}$ governed by the measurement model:

$$Y_t = H_t(X_t, W_t). \tag{2}$$

The two processes $(V_t) \in \mathbb{R}^{n_v}$ and $(W_t) \in \mathbb{R}^{n_w}$ in (1) and (2) are only supposed to be independent white noises. Moreover, it is to be noted that no linearity hypothesis on F_t and H_t is done. We will denote by $Y_{0:t}$ the sequence of the random variables $(Y_0, \ldots, Y_t)$ and by $y_{0:t}$ one realization of this sequence.

Our problem consists in computing at each time t the conditional density L_t of the state X_t given all the observations accumulated up to t, i.e., $L_t = p(X_t | Y_0 = y_0, \ldots, Y_t = y_t)$ and also in estimating any functional of the state $g(X_t)$ by the expectation $\mathbb{E}(g(X_t)|Y_{0:t})$. The Recursive Bayesian filter, also named Optimal Filter, resolves exactly this problem in two steps at each time t [4].

Suppose we know L_{t-1}. The *prediction step* is done according to the following equation:

$$\begin{aligned} &p(X_t = x_t | Y_{0:t-1} = y_{0:t-1}) = \\ &\int_{\mathbb{R}^{n_x}} p(X_t = x_t | X_{t-1} = x) L_{t-1}(x) dx. \end{aligned} \tag{3}$$

The observation y_t enables us to correct this prediction using the Bayes's rule:

$$L_t(x_t) = \frac{p(Y_t = y_t | x_t) p(X_t = x_t | y_{0:t-1})}{\int_{\mathbb{R}^{n_x}} p(Y_t = y_t | x) p(X_t = x | y_{0:t-1}) dx}. \tag{4}$$

These equations provide a closed-form recursion if we assume restrictive hypothesis such as Kalman Filter's ones. The functions F_t and G_t are then supposed to be linear and the noises V_t and W_t to be Gaussian. Unfortunately this modelling is not appropriate in many problems in signal and image processing, which renders the calculations of the integrals in (3) and (4) infeasible (no closed-form).

The original particle filter, named bootstrap filter [4], proposes to approximate the densities $(L_t)_t$ by a finite weighted sum of N Dirac densities centred on elements of $\mathbb{R}^{n_x}$, named particles. The application of the bootstrap filter requires that one knows how:

- to sample from initial prior marginal $p(X_0)$;
- to sample from $p(V_t)$ for all t;
- to compute $p(Y_t = y_t | X_t = x_t)$ for all t through a known function l_t such that $l_t(y; x) \propto p(Y_t = y | X_t = x)$ where missing normalization must not depend on x.

The algorithm then consists in making evolve the particle set $S_t = (s_t^n, q_t^n)_{n=1,\ldots,N}$, where s_t is the particle position and q_t its weight, and to use it to estimate the density L_t by the density $L_{S_t} = \sum_{n=1}^N q_t^n \delta_{s_t^n}$. The weak convergence of the probability density L_{S_t} towards L_t when $N \to \infty$ with rate $1/\sqrt{N}$ can be proved. To avoid the degeneracy of the particle set, i.e. only few particles with high weights and the others with very small ones, a resampling is done in an adaptive way when the number of effective particles, estimated by $\hat{N}_{eff}$, is under a given threshold [2]. Besides the discretization of the filtering integrals, the use of such particles enables to voice many hypothesis on the position of the object and to keep in the long term only the particles whose position is likely given the sequence of observations. In bootstap filter, the particles are "moved" by sampling from the dynamics (1), and importance sampling theory shows that the weighting is only based on likelihood evaluations. In the most general setting [2], the displacement of particles is obtained by sampling from an appropriate density

f which might depend on the data as well. The complete procedure is summarized in figure 1. The reader will find more details on the different filters in [4], [5] or [2] and on adaptive resampling in [6] and [2]. After these recalls, we propose an extension of this algorithm to multiple-object tracking.

3. Multiple object particle filters

3.1. Notations

Let M be the number of objects to track (first assumed to be known and fixed). The state vector we have to estimate is made by concatenating the state vector of each object. At time t, $X_t = (X_t^1, \ldots, X_t^M)$ follows the state equation (1) decomposed in M partial equations:

$$X_t^i = F_t^i(X_{t-1}^i, V_t^i) \qquad \forall\, i = 1, \ldots, M. \tag{5}$$

The noises (V_t^i) and $(V_t^{i'})$ are only supposed to be white both temporally and spatially, independent for $i \neq i'$. The observation vector at time t is denoted by $y_t = (y_t^1, \ldots, y_t^{m_t})$. Following the seminal ideas of R. Streit and T. Luginbuhl [10], we introduce the stochastic vector $K_t \in \{1, \ldots, M\}^{m_t}$ such that $K_t^j = i$ if y_t^j is issued from the i^{th} object. In this case, y_t^j is a realization of the stochastic process:

$$Y_t^j = H_t^i(X_t^i, W_t^j) \text{ if } K_t^j = i. \tag{6}$$

Again, the noises (W_t^j) and $(W_t^{j'})$ are only supposed to be white noises, independent for $j \neq j'$. We assume that the functions H_t^i are such that they can be associated to functional forms l_t^i defined by $l_t^i(y; x) \propto p(Y_t^j = y | K_t^j = i, X_t^i = x)$.
We make the assumption that one measurement can originate from one object or from the clutter and that one object can produce zero or several measurements at one time. For that, we dedicate the model 0 to false alarms. The false alarms are supposed to be uniformly distributed in the observation area. Their number is assumed to arise from a Poisson density of parameter λV where V is the volume of the surveillance area and λ the number of false alarms by volume unity. Of course, we do not associate any kinematic model to false alarms and then no particles represent their density. Let $\pi_t \in [0, 1]^{M+1}$ defined by $\pi_t^i = \mathbb{P}(K_t^j = i)$ for all $j = 1, \ldots m_t$. This definition implicitly assumes that the probabilities π_t^i are independent of the measurements as their indexation is arbitrary. These assumptions imply that m_t may differ from M and that the association is exclusive and exhaustive. In particular, $\sum_{i=0}^{M} \pi_t^i = 1$. Furthermore, it is assumed that the assignment vector K_t has independent components (see[10]). To estimate the density $L_t = p(X_t = (X_t^1, \ldots, X_t^M) | y_{0:t})$, we propose to use particles whose dimension is the sum of the ones of the individual state spaces corresponding to each object, as in [7]. Each of these concatenated vectors then gives jointly a representation of all objects. Let us present the proposed multiple object particle filter (MOPF).

3.2. The MOPF

The initial particle set $S_0 = (s_0^n, 1/N)_{n=1,\ldots,N}$ is such that each component $s_0^{n,i}$ for $i = 1, \ldots, M$ is sampled from $p(X_0^i)$ independently from the others. Assume we have obtained $S_{t-1} = (s_{t-1}^n, q_{t-1}^n)_{n=1,\ldots,N}$ with $\sum_{n=1}^{N} q_{t-1}^n = 1$. Each particle is a vector of dimension $\sum_{i=1}^{M} n_x^i$ where we denote by $s_{t-1}^{n,i}$ the i^{th} component of s_{t-1}^n and where n_x^i designates the dimension of object i.
The prediction is performed by sampling from some proposal density f. In bootstrap filter case, f coincides with the dynamics (5):

$$\text{For } n = 1, \ldots, N \;\; \tilde{s}_t^n = \begin{pmatrix} F_t^1(s_{t-1}^{n,1}, v_t^{n,1}) \\ \vdots \\ F_t^M(s_{t-1}^{n,M}, v_t^{n,M}) \end{pmatrix} \tag{7}$$

with $(v_t^{n,i})$ being realizations of (V_t^i). Examine now the computation of the likelihood of the observations conditioned by the n^{th} particle. We can write for all $n = 1, \ldots, N$:

$$\begin{aligned} & p(Y_t = (y_t^1, \ldots, y_t^{m_t}) | X_t = \tilde{s}_t^n) = \textstyle\prod_{j=1}^{m_t} p(y_t^j | \tilde{s}_t^n) \\ & \propto \textstyle\prod_{j=1}^{m_t} [\frac{\pi_t^0}{V} + \sum_{i=1}^{M} l_t^i(y_t^j; \tilde{s}_t^{n,i}) \pi_t^i]. \end{aligned} \tag{8}$$

It must be noted that first equality in (8) is true only under the assumption of conditional independence of the measures, which we will make. Moreover, the normalization factors between l_t^i and $p(Y_t^j = y | K_t^j = i, X_t^i = x)$ must be the same for all i to write the second equality in (8).
It remains to estimate the association probabilities $(\pi_t^i)_{i=1,\ldots,M}$ which can be seen as the stochastic coefficients of the $M-$component mixture. To estimate them we propose to use a Gibbs sampler whose principles are briefly recalled (see [1] or [9] for more details). For $\theta = (X_t, K_t, \Pi_t)$, it consists in generating a Markov chain which converges to the stationary distribution $p(\theta | Y_{0:t})$ which cannot be sampled directly. Given a partition $\theta^1, \ldots, \theta^P$ of θ, one samples alternatively from the conditional posterior distribution of each component of the partition. Assume the τ first elements of the Markov chain $(\theta_1, \ldots, \theta_\tau)$ have been drawn. We sample the P components of $\theta_{\tau+1}$ as follows:

Draw $\theta_{\tau+1}^1$ from $p(\theta^1 | Y_{0:t}, \theta_\tau^2, \ldots, \theta_\tau^P)$
Draw $\theta_{\tau+1}^2$ from $p(\theta^2 | Y_{0:t}, \theta_{\tau+1}^1, \theta_\tau^3, \ldots, \theta_\tau^P)$

- Initialization: $\begin{cases} s_0^n \sim p(X_0) \\ q_0^n = 1/N \end{cases} \quad n = 1, \ldots, N.$
- For $t = 1, \ldots, T$:
  - Proposal: sample $\tilde{s}_t^n$ from $f(X_t|X_{t-1} = s_{t-1}^n, Y_t = y_t)$ for $n = 1, \ldots, N$.
  - Weighting:
    1. Initialization of the Gibbs sampler: $\begin{cases} \pi_{t,0}^i = \frac{1-\pi_t^0}{M} & i = 1, \ldots, M; \\ X_{t,0}^i = \sum_{n=1}^N q_{t-1}^n \tilde{s}_t^{n,i} & i = 1, \ldots, M. \end{cases}$
    2. For $\tau = 0, \ldots, \tau_{end}$:
       a. $K_{t,\tau+1}^j \sim \quad p(K_{t,\tau+1}^j = i) \propto \begin{cases} \pi_{t,\tau}^i l_t^i(y_t^j; x_{t,\tau}^i) & \text{if } i = 1, \ldots, M; \\ \pi_t^0/V & \text{if } i = 0. \end{cases}$
       b. $\pi_{t,\tau+1}^{1:M} \sim \mathcal{D}((1 + n^i(K_{t,\tau+1}))_{i=1,\ldots,M})$, $n^i(K) \triangleq \sharp\{j : K^j = i\}$.
       c. For each i such that $\exists j^1, \ldots, j^i / K_{t,\tau+1}^{j^l} = i$,
       ▶ $\begin{cases} \sigma_{\tau+1}^n = \tilde{s}_t^{n,i} \\ \chi_{\tau+1}^n = \frac{p(y_t^{j^1}, \ldots, y_t^{j^i} | X_t^i = \sigma_{\tau+1}^n)\, q_{t-1}^n}{\sum_{n=1}^N p(y_t^{j^1}, \ldots, y_t^{j^i} | X_t^i = \sigma_{\tau+1}^n)\, q_{t-1}^n} \end{cases} \quad n = 1, \ldots, N.$
       ▶ $X_{t,\tau+1}^i \sim \sum_{n=1}^N \chi_{\tau+1}^n \delta_{\sigma_{\tau+1}^n}$
       d. For each i such that $\nexists j / K_t^j = i,\ X_{t,\tau+1}^i \sim \sum_{n=1}^N q_{t-1}^n \delta_{\tilde{s}_t^{n,i}}$.
    3. $\hat{\pi}_t^i = \frac{1}{\tau_{beg} - \tau_{end}} \sum_{\tau=\tau_{beg}}^{\tau_{end}} \pi_{t,\tau}^i \quad i = 1, \ldots, M.$
    4. $l_t(Y_t = (y_t^1, \ldots, y_t^{m_t})|\tilde{s}_t^n) = \prod_{j=1}^{m_t} [\frac{\pi_t^0}{V} + \sum_{i=1}^M l_t^i(y_t^j; \tilde{s}_t^{n,i})\hat{\pi}_t^i] \quad n = 1, \ldots, N.$
    5. $q_t^n \propto q_{t-1}^n \frac{p(\tilde{s}_t^n)|s_{t-1}^n)\ l_t(Y_t|\tilde{s}_t^n)}{f(\tilde{s}_t|s_{t-1}, y_t)} \quad n = 1, \ldots, N.$
  - Return $\widehat{\mathbb{E}}g(X_t) = \sum_{n=1}^N q_t^n g(\tilde{s}_t^n)$.
  - Calculate $\hat{N}_{eff} = \frac{1}{\sum_{n=1}^N (q_t^n)^2}$.
  - Resampling: if $\hat{N}_{eff} < N_{threshold}$: $\begin{cases} s_t^n \sim \sum_{k=1}^N q_t^k \delta_{\tilde{s}_t^k} \\ q_t^n = 1/N \end{cases} \quad n = 1, \ldots, N$, else $s_t^n = \tilde{s}_t^n$ for $n = 1, \ldots, N$

Figure 2. MOPF: multiple object particle filter with adaptive resampling.

```
        ⋮                ⋮
Draw θ^P_{τ+1} from p(θ^P|Y_{0:t}, θ^1_{τ+1}, ..., θ^{P-1}_{τ+1})
```

In our case, at a given instant t, the partitionning of θ is:

$$\begin{cases} \theta^j = K_t^j & \text{for } j = 1, \ldots, m_t; \\ \theta^{m_t+i} = \pi_t^i & \text{for } i = 1, \ldots, M; \\ \theta^{m_t+M+i} = X_t & \text{for } i = 1, \ldots, M. \end{cases} \quad (9)$$

We now detail the different steps of this Gibbs sampler.
1. The initialization of Gibbs sampler consists in assigning uniform association probabilities, i.e., $\pi_{t,0}^i = \frac{1-\pi_0}{M}$ for all $i = 1, \ldots, M$, and taking $X_{t,0} = \sum_{n=1}^N q_{t-1}^n \tilde{s}_t^n$, i.e., the centroid of the predicted particle set. Then, suppose that at instant t we have already simulated $(\theta_{t,1}, \ldots, \theta_{t,\tau})$.
2. The $\tau + 1^{th}$ iteration is handled as follows.

a. As the $(K_t^j)_{j=1,\ldots,m_t}$ are supposed to be independent, their individual conditional density reads:

$$p(K_t^j|Y_{0:t}, X_t, (K_t^l)_{l\neq j}, \Pi_t) = p(K_t^j|Y_t^j, X_t, \Pi_t). \quad (10)$$

(K_t^j) are discrete variables and we can write:

$$\begin{aligned} &\mathbb{P}(K_t^j = i|Y_t^j = y_t^j, X_t, \Pi_t) \\ &= \frac{p(Y_t^j = y_t^j|K_t^j = i, X_t, \Pi_t)\mathbb{P}(K_t^j = i|X_t, \Pi_t)}{p(Y_t^j = y_t^j|X_t, \Pi_t)} \\ &\propto \begin{cases} \pi_t^i l_t^i(y_t^j; x_t^i) & \text{if } i = 1, \ldots, M; \\ \pi_t^0/V & \text{if } i = 0. \end{cases} \end{aligned} \quad (11)$$

The realizations $k_{t,\tau+1}^j$ of the vector $K_{t,\tau+1}$ are then sampled according to the weights $p_{t,\tau+1}^{j,0} = \pi_t^0/V$, and $p_{t,\tau+1}^{j,i} = \pi_{t,\tau}^i l_t^i(y_t^j; x_{t,\tau}^i)$ for $i = 1, \ldots, M$.

b. Mixture proportion vector $\Pi_{t,\tau+1}^{1:M}$ is drawn from the conditional density:

$$p(\Pi_t^{1:M}|K_{t,\tau+1},X_{t,\tau},Y_{0:t}) = \\ p(\Pi_t^1,\dots,\Pi_t^M|K_{t,\tau+1}^1,\dots,K_{t,\tau+1}^M,X_{t,\tau},Y_{0:t}) \propto \\ p(K_{t,\tau+1}^1,\dots,K_{t,\tau+1}^M|\Pi_t^1,\dots,\Pi_t^M)p(\Pi_t^1,\dots,\Pi_t^M) \\ = (1-\pi_0)\mathcal{D}(\Pi_t|M,\{n^i(K_{t,\tau+1})_{i=1,\dots,M}\}) \tag{12}$$

where we denote by $n^i(K)$ the number of k^j equal to i and $\mathcal{D}$ stands for Dirichlet distribution.
$X_{t,\tau+1}$ has to be sampled according to the density

$$p(X_t|Y_{0:t},K_{t,\tau+1},\Pi_{t,\tau+1}) = \prod_{i=1}^{M} p(X_t^i|Y_{0:t},K_{t,\tau+1},\Pi_{t,\tau+1}) \tag{13}$$

The values of $K_{t,\tau+1}$ can imply that one object is associated with zero or several measurements that is why we decompose the preceding product in two products:

$$\prod_{i/\exists j^1,\dots,j^i/K_{t,\tau+1}^j=i} p(X_t^i|Y_{0:t-1},y_t^{j^1},\dots,y_t^{j^i},\Pi_{t,\tau+1}) \\ \prod_{i/\forall j K_{t,\tau+1}^j\neq i} p(X_t^i|Y_{0:t-1},\Pi_{t,\tau+1}) \tag{14}$$

c. Let i be an integer in the first product. We can write

$$p(X_t^i|Y_{0:t-1},y_t^{j^1},\dots,y_t^{j^i},\Pi_{t,\tau+1}) = \\ \frac{p(y_t^{j^1},\dots,y_t^{j^i}|X_t^i)p(X_t^i|Y_{0:t-1})}{p(y_t^{j^1},\dots,y_t^{j^i}|Y_{0:t-1})} \tag{15}$$

We are not able to sample directly from the density $\frac{p(y_t^{j^1},\dots,y_t^{j^i}|X_t^i)p(X_t^i|Y_{0:t-1})}{p(y_t^{j^1},\dots,y_t^{j^i}|Y_{0:t-1})}$, for the same reasons as those exposed in section 2 to justify the use of the particle filter (intractability of the integrals). We propose to build the particle set $\Sigma_{\tau+1} = (\sigma_{\tau+1}^n,\chi_{\tau+1}^n)_{n=1,\dots,N}$ whose weights $\chi_{\tau+1}^n$ measure the likelihood of the observations affected by $K_{t,\tau+1}$ to object X_t^i . More precisely, we let:

$$\begin{cases} \sigma_{\tau+1}^n = \tilde{s}_t^{n,i}; \\ \chi_{\tau+1}^n = \frac{p(y_t^{j^1},\dots,y_t^{j^i}|X_t^i=\sigma_{\tau+1}^n)\,q_{t-1}^n}{\sum_{n=1}^N p(y_t^{j^1},\dots,y_t^{j^i}|X_t^i=\sigma_{\tau+1}^n)\,q_{t-1}^n}. \end{cases} \tag{16}$$

The density $\Lambda_{\tau+1} = \sum_{n=1}^N \chi_{\tau+1}^n\delta_{\sigma_{\tau+1}^n}$ converges weakly to the density $p(X_t^i|y_t^{j^1},\dots,y_t^{j^i},Y_{0:t-1})$. Not being able to sample from this last density, $X_{t,\tau+1}^i$ is drawn as a realization from $\Lambda_{\tau+1}$.
d. Now let i be an integer in the second product. As we do not have any measure to correct the predicted particles we draw a realization from the density $\sum_{n=1}^N q_{t-1}^n\delta_{\tilde{s}_t^n}$ for $X_{t,\tau+1}^i$.
3. After a finite number of iterations, we estimate the vector π_t by the average of its last realizations:

$$\hat{\pi}_t^i = \frac{1}{\tau_{beg}-\tau_{end}}\sum_{\tau=\tau_{beg}}^{\tau_{end}} \pi_{t,\tau}^i. \tag{17}$$

Finally the weights are computed according to (8) using the estimation $\hat{\pi}_t^i$ of π_t^i. Figure 2 summarizes the whole procedure.

4. Application to bearings-only problems

We first deal with the classical bearings-only problem using synthetic data. The objects are then "point-objects" in the $x-y$ plane. Their state vector X_t represents the coordinates and the velocities in the $x-y$ plane: $X_t = (x_t,y_t,vx_t,vy_t)$. The following multitarget scenario has been considered: three targets follow a near-constant-velocity model defined by (18). The discretized state equation associated with time period Δt is:

$$X_{t+\Delta t} = \begin{pmatrix} Id_2 & \Delta t\, Id_2 \\ 0 & Id_2 \end{pmatrix} X_t + \begin{pmatrix} \frac{\Delta t^2}{2} Id_2 \\ \Delta t\, Id_2 \end{pmatrix} V_t, \tag{18}$$

where Id_2 is the identity matrix in dimension 2 and V_t is a Gaussian zero-mean vector of covariance matrix $\Sigma_V = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$. Let $\hat{X}_t$ be the estimation of X_t computed by the MOPF with $g(x) = x$, i.e. , $\hat{X}_t = \sum_{n=1}^N q_t^n \tilde{s}_t^n$. For this application we use a bootstrap filter, i.e. , the importance function f is in fact the prior law $p(x_t|x_{t-1})$. Each object produces one measurement at each time period according to (19) except during the time interval $[600\ 700]$ where the first object does not produce any measurement and the second produces two y^1 and y^2 according to:

$$Y_t = \arctan\left(\frac{x_t - x_t^{obs}}{y_t - y_t^{obs}}\right) + W_t, \tag{19}$$

where W_t is a zero-mean Gaussian noise of covariance σ_w^2 independent of V_t. x_{obs} and y_{obs} are the Cartesian coordinates of the observer, which are known. The trajectories of the three targets and of the observer are plotted in figure 3.1 and the differences between the three couples of bearings simulated are plotted in figure 4. As soon as the difference between two bearings issued from two different targets is lower than the standard deviation of the observation noise, the two measures cannot be distinguished, which makes this scenario very difficult. This difficulty is increased by the detection gap for the first object. One particular run of the particle filter with 5000 particles is presented in figure 3.2. The plot of the three estimated trajectories shows that the data association is overcome. There is no trajectories reversal and the estimations are quite satisfactory. Figure 5 shows the results of the estimation of the three components of π_t and figure 3.3 represents the average of each component π_t^i over successive intervals of 100 time steps and over the 20 trials. When there is an ambiguity about the origin of the measurements (i.e., when the differences between the bearings are lower than the standard deviation noise), the

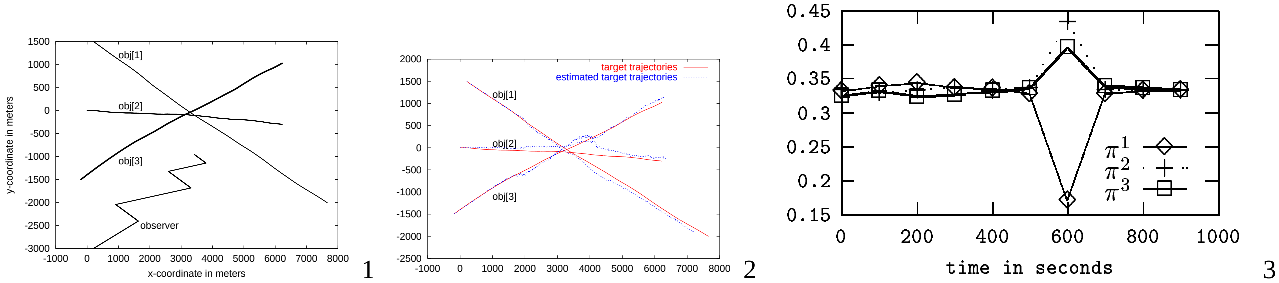

Figure 3. (1): Trajectories of the three targets and of the observer (2): The target trajectories and their estimation with 5000 particles (3): The average of the estimated components of the vector π_t over the ten consecutive time intervals of length 100 and over 20 trials.

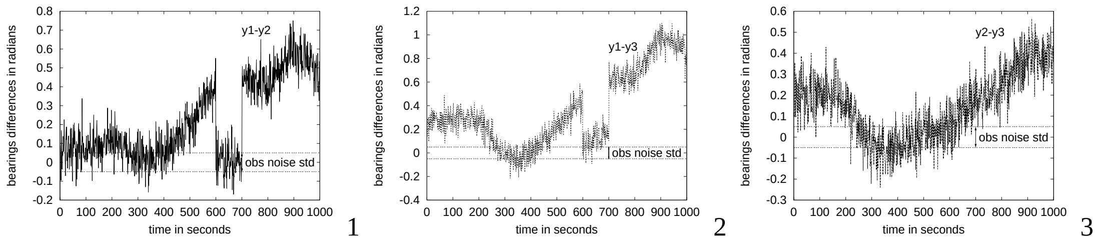

Figure 4. Differences between the three couples of target bearings at each time period compared to the standard deviation of the observation noise: (1) measurements 1 and 2 (2) 1 and 3 (3) 2 and 3.

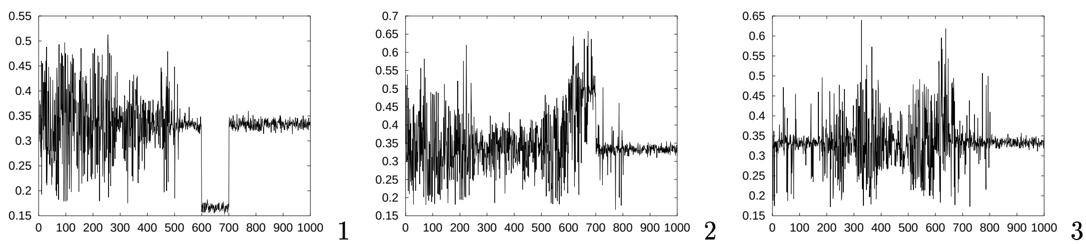

Figure 5. The estimated components of the vector π_t obtained with 5000 particles: (1) $\hat{\pi}_t^1$, (2) $\hat{\pi}_t^2$, (3) $\hat{\pi}_t^3$.

components of π vary in average around $1/3$ for $M = 3$ objects and it stabilizes at uniform estimation ($1/3$ for $M = 3$ objects) when the ambiguity disappears. The momentary measurement gap for the first object is correctly handled as the first component π_t^1 is instantaneously estimated as 0.15 from instant 600 to 700.
The vector π_t can then in turn help the estimation: an object leaving the surveillance or vision area can be detected by a drop of its π_t component. On the other hand, the appearing of a new object might be related to an observation whose likelihood is low whatever object it is associated with. The statistic framework of our algorithm, through an interaction between the estimation of the objects trajectories and of the data association, seems then very promising in order to deal with the appearing and disappearing of objects of the study area, i.e. to deal with a varying number of objects.

5. Application in image-based tracking

We focus on a video sequence where three persons are moving according to unknown dynamics. As they are crossing in the 2D-plane of the images, the tracker must solve partial and complete occlusions as shown on figures 6.{1,2,3}. The modelization of the tracking problem consists in defining a state vector, its evolution model and an observation model. Such a design stage is quite easy in signal processing because the models and the different measures are well defined in the literature. It is not the case for the problem of video-based tracking: we have usually few knowledge on the a priori motion of the objects and any

Figure 6. $(1, 2, 3)$: Frames 7, 30 and 67 of a pedestrian video-sequence with occlusions; (4): Outline of the motion area detected around one of the moving person; (5): Fourier contour obtained by inverse Fourier transform of the truncated discrete Fourier transform to the five first coefficients of the previous outline.

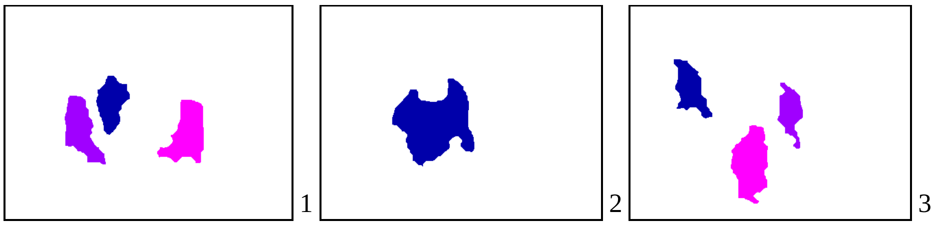

Figure 7. Maps of motion-based segmentation associated to the frames $19, 39$ and 79.

Figure 8. Estimated outlines of the three pedestrians with 100 particles and 9 Fourier coefficients for the frames $19, 39$ and 79.

computation on the image can constitute a measurement.

5.1. State space model

In order to keep reasonable dimension of the state space, one have to come up with a compact representation of the object silhouette unless strong prior on the shape of the object of interest is available. An appealing and generic approach consists in using Fourier descriptors obtained by inverse Fourier transform of the truncated discrete Fourier transform of the initial object contour. For instance, the outline of one moving pedestrian (as observed through a motion segmentation map, see section 5.2) associated to the five first Fourier coefficients is presented in figure 6.5. For each object, the state vector is then composed of $2N_U + 1$ Fourier coefficients Xc_t^i and of the 2D translation vector Xv_t^i of the object center between $t - \Delta t$ and t:

$$X_t^i = (Xc_t^i, Xv_t^i). \tag{20}$$

Notice that the object center is captured by the first Fourier coefficient denoted $Xc_t^{i,0}$. We use the following dynamics prior:

$$\begin{cases} Xv_t^i = W_t^i; \\ Xc_t^{i,0} = Xc_{t-\Delta t}^{i,0} + \Delta t Xv_t^i + V_t^{i,0}; \\ Xc_t^{i,k} = Xc_{t-\Delta t}^{i,k} + V_t^{i,k} \text{ for } k = -N_U, \ldots, -1, +1, \ldots, N_U, \end{cases} \tag{21}$$

where V_t is a Gaussian zero-mean noise of standard deviation σ_v and W_t is uniform on $[-V_{max}; V_{max}]$.

5.2. Measurements

We use two types of measurements related to position and velocity of moving objects in the scene. By a motion-based segmentation of the image (see [8]), we first compute connected components of motion Cc_t^j for $j = 1, \ldots, m_t$ as shown in figure 7. Such a segmentation provides information on the localization (and shape) of the objects in motion w.r.t. the camera. The likelihood of the connected component Cc_t^j conditionnaly to the i^{th} component of the n^{th} particle $Sc_t^{n,i}$ is chosen as:

$$\frac{|\mathring{S}c_t^{n,i} \cap Cc_t^j|}{\sqrt{|\mathring{S}c_t^{n,i}|}} \tag{22}$$

where $\mathring{S}c$ denotes the inside of the closed contour defined by Fourier descriptors Sc, and $|.|$ stands for set cardinal. The denominator prevents a bias toward large contours. We then estimate the translation in the 2D-plane of each connected component between the frame at current time and the previous one. These translations will be noted T_t^j for $j = 1, \ldots, m_t$ and are supposed to arise from the model:

$$T_t^j = Xv_t^i + Z_t^j \text{ if } K_t^j = i \tag{23}$$

where Z_t^j is a zero-mean Gaussian noise of std σ_z. As the velocity prior is very weak (uniform distribution, in (21)), we use these motion measurements to construct an importance function that will perform better than standard bootstrap-type choice based only on the dynamics, to guide particles toward regions of high likelihood.

5.3. Importance function

Given a partcile s, a translation estimation is considered available when only one of the object centers belongs to the connected component under consideration. Otherwise it means that several objects are observed through the same motion region (as in figure 7.2) and its global motion results in general from an intricate mix of individual motions. Then, if $s_{t-1}^{n,i}$ is the only object from s_{t-1}^n whose center lies in Cc_t^j, we propose the displacement of the i-th object in the new particle according to:

$$\tilde{S}v_t^{n,i} = T_t^j + Z_t^{n,i} \tag{24}$$

where Z_t^i is the same noise as in (23) . The other components $\tilde{S}c_t^{n,i}$ are predicted according to the dynamic prior in (21).

If no translation motion estimation is available, the new particle is displaced according to:

$$\tilde{S}v_t^{n,i} = Sv_{t-1}^{n,i} + Z_t^{n,i}. \tag{25}$$

5.4. Importance weights

The weight of the n^{th} particle is proportionnal to

$$q_{t-1}^n \frac{p(\tilde{S}c_t^n|S_{t-\Delta t}^n)\, p(\tilde{S}v_t^n|S_{t-\Delta t}^n)\, p(T_t|\tilde{S}_t^n)\, p(Cc_t|\tilde{S}_t^n)}{p(\tilde{S}c_t^n|S_{t-\Delta t}^n)\, f(\tilde{S}v_t^n|T_t)} \tag{26}$$

As the prior on the motion is uniform, it does not affect the weights. When we use the current translation estimation, the choice of the importance function implies that $p(T_t|X_t) = p(T_t|Xv_t)$ is equal to $f(Xv_t|T_t)$. If this is the case for all the objects of the predicted particle, the new weight is obtained by multiplying the previous one by the likelihood of the connected component likelihood of the motion-based segmentation.

Result samples on multiple people tracking are presented in figure 8.

6. Conclusion

We proposed multiple object tracking based on a mix of particle filtering and Gibbs sampling. Target state vectors and association probabilities are estimated jointly without enumeration, pruning or gating, by means of particle sets representing the joint *a posteriori* law of the target states. We have demonstrated the relevance of the approach both in bearings-only tracking and image-based tracking. We are currently investigating its extension to a varying number of objects, using parameters statistics.

References

[1] J. Diebolt and C. P. Robert. Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society series B*, 56:363–375, 1994.

[2] A. Doucet. On sequential simulation-based methods for Bayesian filtering. Technical report, CUED/F-INFENG/TR 310, Signal Processing Group, Departement of Engінerring, University of Cambridge, 1998.

[3] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *IEEE Journal of Oceanic Engineering*, 8:173–184, July 1983.

[4] N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc.F, Radar and signal procesing*, 140(2):107–113, 1993.

[5] M. Isard and A. Blake. CONDENSATION — conditional density propagation for visual tracking. *Int. J. Computer Vision*, 29(1):5–28, 1998.

[6] A. Kong, J. Liu, and W. Wong. Sequential imputation method and Bayesian missing data problems. *J. Amer. Statist. Assoc.*, 89:278–288, 1994.

[7] J. MacCormick and A. Blake. A probabilistic exclusion principle for tracking multiple objects. In *Proc. Int. Conf. Computer Vision*, pages 572–578, 1999.

[8] J.-M. Odobez and P. Bouthémy. Direct incremental model-based image motion segmentation for video analysis. *Signal Processing*, 66(2):143–155, Apr. 1998.

[9] M. Stephens. *Bayesian Methods for Mixtures of Normal Distributions*. PhD thesis, Magdalen College, Oxford, 1997.

[10] R. L. Streit and T. E. Luginbuhl. Maximum likelihood method for probabilistic multi-hypothesis tracking. In *Proceedings of SPIE International Symposium, Signal and Data Processing of Small Targets 1994*, volume 2335, 1994.