Joint Target Tracking and Identification – Part II: Shape Video Computing

Pierre Minvielle
CEA DAM
BP n° 2
33114 Le Barp, France
pierre.minvielle@cea.fr

Alan D. Marrs Simon Maskell
QinetiQ Ltd.
St Andrews Rd, Malvern
WR14 3PS, Wrcs, UK
admarrs@qinetiq.com
s.maskell@signal.qinetiq.com

Arnaud Doucet
Signal Processing Group, CUED
University of Cambridge
Cambridge CB2 1PZ, UK
ad2@eng.cam.ac.uk

Abstract - *This paper describes an application of sequential Monte Carlo model-based approaches to perform joint target tracking and identification. While a geometric shape is moving inside the field of view of a CCD camera, alternatively getting closer and moving away while rotating, the data processing system is confronted to challenging tasks: track the moving shape in real 3D space, i.e. estimate its position and orientation, and at the same time dynamically estimate its dimensions and, if required, identify it. The system is based on class-specific Bayesian filters. More originally, the issue of the fixed hyper-parameter estimation, here the geometric shape dimensions, is solved by combining two different techniques. The first one consists of Markov Chain Monte Carlo moves that rescale both the trajectory and the shape; it benefits from an efficient statistic which summaries the trajectory with regard to moves. The second one is an artificial deformation diffusion of the shape.*

Keywords: Video tracking, identification, particle filtering, parameter estimation, MCMC

1 Introduction

The paper presents a didactic application of digital CCD video tracking of basic geometric shapes (such as cubes, parallelepipeds, spheres and hemispheres). It illustrates the issues developed in the associated paper [1], which deals with sequential Monte Carlo approaches for joint tracking and identification. In the tracking scenario, a shape moves inside the field of view (FOV) of a CCD camera, alternatively getting closer and moving away while rotating (Figure 1).

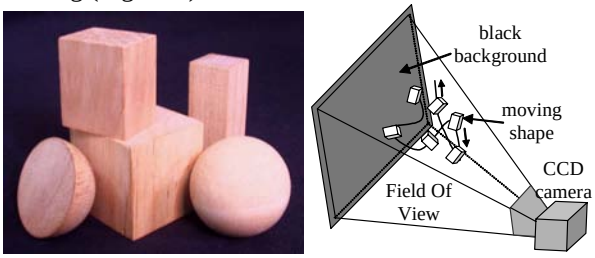


Figure 1. The didactic application

In order to focus on the tracking issues and not to deal with very specific image processing problems, the scenario is simplified:

- The environment is uncluttered (a black cloth has been laid behind the target) ,
- The lighting conditions are correct, providing good conditions for feature extraction.
- There is no occultation of the target (the target is suspended by threads).

The goals of the application are:

(a) Track the shapes in the real 3D space, i.e. estimate simultaneously the position and the orientation of the moving shape (the so-called target pose),

(b) Estimate sequentially also, if they are unknown, the dimensions of the shape (e.g. radius of a sphere, length/width/height of a parallelepiped),

(c) Identify the shape, when there is a prior uncertainty on the class (e.g.: sphere or hemisphere, cube or parallelepiped).

The main characteristics of the shape video joint tracking and identification (JTI) application are:

- The general architecture is based on class-specific filters [1],[2]: each shape model class holds its own inference process to estimate the posterior distributions. Moreover, outputs are collected to perform the estimation of the class probabilities. With this architecture, each shape model class is able to have its own state space (which dimension differs from one class to another class, depending on the symmetry level of the shapes) and its own related dynamic model. While [2] illustrates behaviour-based identification, this application illustrates feature-based identification.
- The inference engines of each model class are based on a recent particle filtering technique [3],[4], that is a sequential adaptation of annealed importance sampling [5] and is also called "bridging densities". For convenience, it will be called subsequently Simulated Annealed Particle Filter (SAPF). Briefly speaking, SAPF avoids the degeneracy phenomenon by computing gradually the posterior distribution and performing resample-MCMC (Markov Chain Monte Carlo) moves [6]. A similar technique has already been used for human body video tracking [7]

- The fixed parameters, which are here dimensions of the shapes, are added to the dynamic state vector, leading to a hybrid state vector. More originally, the dimension estimation is performed by the combined use of two fixed-parameter estimation techniques. The first one consists of MCMC moves that re-scale both the trajectory and the shape; it benefits from an efficient statistic that summarises the trajectory with regard to moves. The second one is an artificial evolution of deformation of the shape that diffuses across the layers of the SAPF.
- The likelihood function uses extracted features (foreground and edges) that are provided by an image processing stage and compares them to an expected projection of the shape requiring a shape model and a camera model.

The organisation of the paper is as follows. Section 2 presents the SAPF filter and the model-based approach. Next Section 3 describes the combination of two methods to estimate target features. Section 4 gives results for two scenarios. Finally, Section 5 gives the concluding remarks.

2 Simulated Annealed Particle Filter

2.1 Principles

Simulated Annealed Particle filters (SAPF), known as "bridging densities" [3],[4], are a sequential adaptation of annealed importance sampling [5]. SAPF turns out to be useful when no good importance density is available, meaning when the likelihood is centred far away from the points sampled from the importance distribution. For example, this is the case when, through lack of an efficient importance density, the prior $p(x_k|x_{k-1})$ is used but has a much broader distribution than the likelihood $p(y_k|x_k)$. In such a situation, resample-move techniques [6], which add MCMC iterations in order to avoid sample impoverishment, might have a very few accepted moves and consequently the procedure might be very computationally expensive. The basic idea of SAPF algorithms is then to use intermediate distributions in order that each step is small enough to avoid the degeneracy problem and that the particles progressively migrate and concentrate in the likely parts of the state space. Furthermore, replacing a single large transition with a series of smaller transitions is supposed to reduce the Monte Carlo variations of the posterior distribution.

Between the initial distribution $\Pi_0(x_k) \propto q(x_k|x_{k-1}, y_k)$ ($q(\cdot)$ is the importance density) and the final distribution $\Pi_{M+1}(x_k) \propto p(x_k|y_{1:k})$, $M-1$ intermediate distributions $\Pi_m(x_k)$ are defined according to an appropriate schedule, such as:

$$\Pi_m(x_k) \propto q(x_k|x_{k-1}, y_k)^{\alpha_m} \cdot p(x_k|y_{1:k})^{1-\alpha_m} \qquad (1)$$

The sequence $\{\alpha_m\}_{m=0,...,M+1}$, with $\alpha_0 = 1 < \alpha_1 < ... < \alpha_M < \alpha_{M+1} = 0$, determines the cooling schedule of this simulated annealing method. Using the prior distribution as the importance density, (1) can be rewritten as:

$$\Pi_m(x_k) \propto p(x_k|y_{1:k-1}) \cdot p(y_k|x_k)^{1-\alpha_m} \qquad (2)$$

Notice that the gradual introduction of the likelihood is then evident.

The SAPF aims to move the set of particles through this sequence of distributions while respecting Bayesian inference; the resulting weighting samples are eventually distributed according the posterior distribution $p(x_k|y_{1:k})$.

2.2 The SAPF algorithm

The pseudo-code of a SAPF recursion is shown below:

$$\left[\{x_k^i, w_k^i\}_{i=1}^{N_S}\right] = \text{SAPF}\left(\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_S}, y_k\right)$$

- Draw $\left\{x_k^{i\,(0)}, w_k^{i\,(0)}\right\}_{i=1}^{N_S}$ from the importance distribution $q(x_k|x_{k-1}^i, y_k)$

  (*) Choosing the prior distribution, $q(x_k|x_{k-1}^i, y_k)$

- For $m = 1,...,M+1$ (layer of simulated annealing)

  - *Increment step*

  Increment the importance weights:

$$x_k^{i\,(m)} = x_k^{i\,(m-1)}, \quad w_k^{i\,(m)} \propto w_k^{i\,(m-1)} \cdot \frac{\Pi_m\left(w_k^{i\,(m-1)}\right)}{\Pi_{m-1}\left(w_k^{i\,(m-1)}\right)}$$

  (*) Choosing $\Pi_m(x_k) \propto p(x_k|y_{1:k-1}) \cdot p(y_k|x_k)^{1-\alpha_m}$, that leads to:

$$w_k^{i\,(m)} = w_k^{i\,(m-1)} \cdot p\left(y_k \middle| x_k^{i\,(m-1)}\right)^{\alpha_m - \alpha_{m-1}}$$

  - *Selection step*

  If necessary ($\hat{N}_{eff} \overset{\Delta}{=} \frac{1}{\sum_{i=1}^{N_S}\left(w_k^{i\,(m)}\right)^2} < N_T$) to resample,

$$\left[\left\{x_k^{i\,(m)}, w_k^{i\,(m)}\right\}_{i=1}^{N_S}\right] = \text{RESAMPLE}\left(\left\{x_k^{i\,(m)}, w_k^{i\,(m)}\right\}_{i=1}^{N_S}\right)$$

  [8],[9],[10]

  (**) *Without normalisation (model selection): all the particle weights have now the same weight, but* $\sum_{i=1}^{N} w_k^{i\,(m)}$ *must be unchanged.*

  - *MCMC step*

Apply MCMC moves with a Markov chain transition kernel that keeps $\Pi_m(x_k)$ invariant.

- End for
- Eventually assign $\left\{x_k^i, w_k^i\right\}_{i=1}^{N_S} = \left\{x_k^{i\,(M+1)}, w_k^{i\,(M+1)}\right\}_{i=1}^{N_S}$

After the increment step that modifies the weights, $\left\{x_k^{i\,(m)}, w_k^{i\,(m)}\right\}_{i=1}^{N_S}$ are distributed according to the intermediate distribution $\Pi_m(x_k)$. The following steps do not change that distribution. The resampling step reduces the degeneracy of the samples and the MCMC moves diversify the samples and explore the probable areas of state space. The algorithm is illustrated in Figure 2 with a number of simulated annealing layers equal to 3. Starting from the importance distribution Π_0 (dark grey), the likelihood (clear grey) is progressively introduced, so that at each layer the particles concentrate more and more in the probable parts of the state space. Eventually, the particles are distributed according to the posterior distribution Π_3.

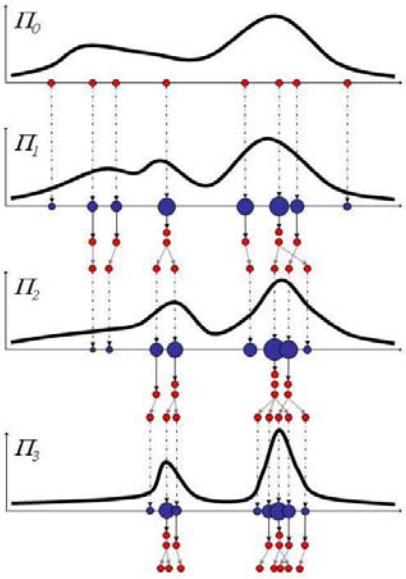


Figure 2. SAPF principles

Remark: In the JTI video application, it has been chosen to perform the resampling step systematically. As the resampling step helps to regain the entire efficient particle set, it improves the exploration of the state-space for the next move step. Notice that if $M = 0$, there is no bridging densities and the algorithm corresponds to the "Resample-Move" algorithm.

2.3 Interests for the JTI application

- Through the lack of a good importance distribution, the SAPF algorithm is able to deal with a multi-modal likelihood without becoming misguided by local maxima and being at a complete standstill in a specific mode. It just so happened with the likelihood of the application, which is described further. Besides, the high dimension state space induces rather peaked likelihood functions and posterior distributions. SAPF is adapted to deal with high dimensional state spaces [3],[4] and moreover to output good maximum a posteriori (MAP) estimates.
- Since SAPF is derived from resample-MCMC move techniques, it is rather straightforward and natural to add target attribute moves inside the MCMC move step in order to perform fixed-parameter estimation. As the dimension of the state space becomes higher and the search of the good modes more challenging, an appropriate cooling schedule and suitable proposal distributions are even more required to lead to efficient annealed algorithms.
- Concerning the model selection goal, SAPF is able to receive whether reversible jump MCMC inside the MCMC move step or to compute the outputs needed by the class-specific filters approach to estimate the class probabilities. The latest approach is chosen for the JTI video application: the particle weights must not be normalised in the resampling step (**) so that $\sum_{i=1}^{N} w_k^i$ eventually approximates the predictive density of the measurement y_k (given the model class). The expected reduction of the Monte Carlo variations should benefit the accuracy of the class probability estimates.

Remark: An adaptive cooling schedule [4] monitors the size of the successive steps $\Delta\alpha_m = \alpha_m - \alpha_{m-1}$ so that the rate of efficient particles $\frac{\hat{N}_{eff}}{N_T}$ remains roughly constant in the SAPF resampling step and the degeneracy is under control ($\Delta\alpha_m$ can be determined by any numerical optimisation routine). Consequently, the number of steps varies, depending on the difference between the proposal distribution (here the prior for the video application) and the posterior distribution. Thus, the first measurement requires the most important computational effort and the highest number of steps since the prior knowledge $p(x_0, \theta)$ is then vaguely informative, while the next measurements starting from a quite informative prior $p(x_k, \theta | y_{1:k-1})$ require less computation.

2.4 MCMC proposal distribution

The MCMC move step used in the video application is based on the MH scheme with a symmetric (Gaussian) random walk proposal (Metropolis algorithm). It consists in a transition move K_T, simultaneously on the position, the velocity and the orientation of the target. The variance of the proposal distribution decreases at each layer of the SAPF recursion as the important moves become less and less possible and the distribution freezes. The variance evolution is tuned in order to assure more or less a rate of

acceptance move around 30%. Other moves dedicated to target features are presented further.

2.5 Model-based approach

In this section, we briefly describe the models chosen for this specific application.

Hybrid state space model

The state space is defined according to the shape class. A hybrid state vector $\boldsymbol{\chi}_k = [\mathbf{x}_k \quad \boldsymbol{\Theta}_S]^T$ describes the target state at time k.

- $\mathbf{x}_k$ is the dynamic part that includes the position vector of the target (3 coordinates), its velocity vector (3 coordinates and its orientation. The target orientation is described by the Euler angles. Depending on the level of symmetry of the shape, none of them are required (sphere) to the full three Euler angles are required (parallelepiped).
- $\boldsymbol{\Theta}_S$ is the attribute state of the target, i.e. the shape dimensions. Basically it depends on the shape: $\boldsymbol{\Theta}_S = [r]$ (radius) for spheres and hemispheres, $\boldsymbol{\Theta}_S = [l]$ (length) for cubes, $\boldsymbol{\Theta}_S = [l\ w\ h]^T$ (length, width, height) for parallelepipeds. Of course, this state part is fixed.

Dynamic model

The dynamic model is a (second order) piecewise constant white acceleration model [11] on the position component and a random walk on the orientation component.

Observation model

Similarly to [7], a pre-processing step of feature extraction is applied to the current image. Two image features are extracted: the edges and the foreground (cf. Figure 3). Notice that both the lighting effects (part of the shape not very exposed) and the blurring effects may cause some of the edges to be undetected, altering the information given by the image.

Figure 3 - Feature extraction

Associated with a shape particle and its state vector $\boldsymbol{\chi}_k = [\mathbf{x}_k \quad \boldsymbol{\Theta}_S]^T$ at time k, a shape model and a standard pinhole camera model are used to determine the expected image features of such a particle. The shape model generates edge and foreground sampling points, according to the class of the shape and its pose.

The likelihood function $p(y|\boldsymbol{\chi}_k)$ is then the product of a likelihood $L_y^E(\boldsymbol{\chi}_k)$ based on the edges and a likelihood $L_y^F(\boldsymbol{\chi}_k)$ based on the foreground. Both are computed from the above mentioned sampling points.

3 Shape feature estimation

Combined techniques are used to perform the estimation of the shape features. The first technique is a genuine resample-MCMC move that is here applied to re-scale both the dimensions of the target and its whole trajectory. Then, as described for the point 2D bearings-only tracking problem [12], a sufficient statistic can sum up the whole trajectory so that the computational and memory requirements of the re-scaling moves are reasonable.

In order to be able to reach any point of the hyper-parameter space, other moves are needed. They consist in deformation moves that change the shape dimensions without keeping the proportions. Since no efficient statistics are here available and since anyway the degeneracy of the past of the trajectory would strongly bias the evaluation and the resulting estimates, an artificial evolution is applied, consisting in a deformation diffusion through the layers of the SAPF.

3.1 Re-scaling MCMC Move K_λ

Denoting the whole trajectory $\boldsymbol{\Lambda}_k$ (states at all time points and parameters) at time k, the re-scaling move K_λ consists in:

$\boldsymbol{\Lambda}_k = [\mathbf{x}_0 \ \ldots\ \mathbf{x}_k \ \boldsymbol{\Theta}_\mathbf{S}] \xrightarrow{K_\lambda} \boldsymbol{\Lambda}_k^* = [\mathbf{x}_0^* \ \ldots\ \mathbf{x}_k^* \ \boldsymbol{\Theta}_\mathbf{S}^*]$, where $\mathbf{x}_0^*, \ldots, \mathbf{x}_k^*$ correspond to a reduction or amplification of the position and velocity parts of the states $\mathbf{x}_0, \ldots, \mathbf{x}_k$ by the same factor λ, the orientation part remaining unchanged. As for the parameters, $\boldsymbol{\Theta}_\mathbf{S}^* = \lambda \cdot \boldsymbol{\Theta}_\mathbf{S}$, i.e. the moved shape is the homothetic of the original shape by the factor λ (the proportions remain unchanged). As a matter of fact, the move is nothing else than envisaging that the shape-particle is bigger or smaller, but respectively further or closer in such a way that the expected measurements would be unchanged.

In order to construct a Markov chain transition kernel K_λ that keeps the joint posterior distribution $p(x_{0:k}, \theta | y_{1:k})$ invariant, a Metropolis-Hastings method is chosen with a proposal $q(\boldsymbol{\Lambda}_k^* | \boldsymbol{\Lambda}_k)$. The acceptance probability of the move is:

$$\alpha(\boldsymbol{\Lambda}_k, \boldsymbol{\Lambda}_k^*) = \min\left\{1, \frac{p(\boldsymbol{\Lambda}_k^* | y_{1:k})}{p(\boldsymbol{\Lambda}_k | y_{1:k})} \cdot \frac{q(\boldsymbol{\Lambda}_k | \boldsymbol{\Lambda}_k^*)}{q(\boldsymbol{\Lambda}_k^* | \boldsymbol{\Lambda}_k)}\right\} \qquad (3)$$

Using the Markovian property of the dynamic process and the conditional independence of the observations given the process, the ratio $\rho\left(\mathbf{\Lambda}_k, \mathbf{\Lambda}_k^*\right) = \frac{p\left(\mathbf{\Lambda}_k^* \middle| y_{1:k}\right)}{p\left(\mathbf{\Lambda}_k \middle| y_{1:k}\right)}$ can be developed:

$$\rho\left(\mathbf{\Lambda}_k, \mathbf{\Lambda}_k^*\right) = \frac{\left[\prod_{i=1}^{k} p\left(y_i \middle| \mathbf{x}_i^*, \mathbf{\Theta}_\mathbf{S}^*\right)\right] \cdot \left[\prod_{i=0}^{k-1} p\left(\mathbf{x}_{i+1}^* \middle| \mathbf{x}_i^*\right)\right] \cdot p\left(\mathbf{x}_0^*, \mathbf{\Theta}_\mathbf{S}^*\right)}{\left[\prod_{i=1}^{k} p\left(y_i \middle| \mathbf{x}_i, \mathbf{\Theta}_\mathbf{S}\right)\right] \cdot \left[\prod_{i=0}^{k-1} p\left(\mathbf{x}_{i+1} \middle| \mathbf{x}_i\right)\right] \cdot p\left(\mathbf{x}_0, \mathbf{\Theta}_\mathbf{S}\right)} \quad (4)$$

Since the successive likelihood remains unchanged after the move, the previous ratio is equal to:

$$\rho\left(\Lambda_k, \Lambda_k^*\right) = \frac{\left[\prod_{i=0}^{k-1} p\left(\mathbf{x}_{i+1}^* \middle| \mathbf{x}_i^*\right)\right] \cdot p\left(\mathbf{x}_0^*, \mathbf{\Theta}_\mathbf{S}^*\right)}{\left[\prod_{i=0}^{k-1} p\left(\mathbf{x}_{i+1} \middle| \mathbf{x}_i\right)\right] \cdot p\left(\mathbf{x}_0, \mathbf{\Theta}_\mathbf{S}\right)} \quad (5)$$

As the dynamic model gives:

$$p\left(\mathbf{x}_{i+1} \middle| \mathbf{x}_i\right) = \frac{1}{\left(\sqrt{2\pi}\sigma_\upsilon\right)^3} e^{-\frac{\gamma_i}{2\sigma_\upsilon{}^2}} \text{ with:}$$

$$\gamma_i = \left(\frac{\dot{x}_{i+1} - \dot{x}_i}{dT}\right)^2 + \left(\frac{\dot{y}_{i+1} - \dot{y}_i}{dT}\right)^2 + \left(\frac{\dot{z}_{i+1} - \dot{z}_i}{dT}\right)^2 \quad (6)$$

It is then possible to obtain:

$$\log\left(\prod_{i=0}^{k-1} p\left(\mathbf{x}_{i+1} \middle| \mathbf{x}_i\right)\right) = -\frac{1}{2\sigma_W{}^2} \sum_{i=0}^{k-1} \gamma_i + k \cdot C \quad (7)$$

Denoting $\bar{\gamma}_k$ the mean of the quadratic acceleration of the target and developing the prior, it eventually leads to the simple expression:

$$\rho\left(\Lambda_k, \Lambda_k^*\right) = \underbrace{\frac{p\left(r_0^*\right) \cdot p\left(v_0^*\right) \cdot p\left(\mathbf{\Theta}_\mathbf{S}^*\right)}{p\left(r_0\right) \cdot p\left(v_0\right) \cdot p\left(\mathbf{\Theta}_\mathbf{S}\right)}}_{\text{prior}} \cdot \underbrace{e^{\frac{k \cdot \bar{\gamma}_k \cdot \left(1-\lambda^2\right)}{2\sigma_\upsilon{}^2}}}_{\text{trajectory}} \quad (8)$$

The sufficient statistic $\left(r_0, v_0, \mathbf{\Theta}_\mathbf{S}, \bar{\gamma}_k\right)$ sums up all the information needed to evaluate the re-scaling move K_λ. Thus, it only requires to add the quadratic acceleration mean $\bar{\gamma}_k$ to the state and to update it recursively, which is straightforward, at each time k. Notice also that the associated re-scaling move is then: $\bar{\gamma}_k^* = \lambda^2 \cdot \bar{\gamma}_k$. Although there is an unavoidable degeneracy in the past of the particles, it has appeared that in the video tracking application the distribution of $\bar{\gamma}_k$, for a given trajectory scale, is quite narrow. Thus, the committed error is quite negligible in proportion to a proposed re-scaling move λ.

As in [6], an uniform symmetric proposal density $q\left(\mathbf{\Lambda}_k^* \middle| \mathbf{\Lambda}_k\right) = q\left(\mathbf{\Lambda}_k \middle| \mathbf{\Lambda}_k^*\right)$ is chosen for the re-scaling move K_λ, i.e. $\lambda \sim U\left[\lambda_1 \quad 1/\lambda_1\right]$; the value of the chosen constant λ_1 is tuned and decreases in time in order to assure more or less a rate of acceptance move around 30%.

The re-scaling move K_λ is added to the transition MCMC move K_T of the SAPF recursion described in Section 5.1, resulting in a cycle hybrid Kernel $K_T \cdot K_\lambda$.

Remarks concerning (8):

- If $\bar{\gamma}_k = 0$, there is no apparent motion and (8) indicates that all the re-scaling moves are possible, meaning for example that the target could be just as well further and bigger, which is exactly the uncertainty of the information provided by only one measurement.
- Low values of λ correspond to high values of the trajectory part of the ratio $\rho\left(\Lambda_k, \Lambda_k^*\right)$; that means that closer trajectories are favoured as long as the prior allows it. According only to the dynamic model, closer trajectories are more likely since they undergo less important acceleration and deceleration.
- The higher the dynamic noise σ_ϑ is, the more likely the further trajectories are.
- While k increases, the possible moves become smaller and smaller.

3.2 Deformation artificial diffusion

To complement the re-scaling move K_λ and to be able to reach any point of the parameter space, a complementary deformation moves K_D has to be performed, consisting of: $\mathbf{\Lambda}_k = \left[\mathbf{x}_0 \ \ldots \ \mathbf{x}_k \ \ \mathbf{\Theta}_\mathbf{S}\right] \xrightarrow{K_D} \mathbf{\Lambda}_k^* = \left[\mathbf{x}_0 \ \ldots \ \mathbf{x}_k \ \ \mathbf{\Theta}_\mathbf{S}^*\right]$. The trajectory of the shape-particle is unaltered; only its dimensions are changed now without trying to keep the proportions.

Choosing again a MH method, the ratio of the acceptance probability of the move is:

$$\rho\left(\mathbf{\Theta}_\mathbf{S}, \mathbf{\Theta}_\mathbf{S}^*\right) = \frac{p\left(\mathbf{\Lambda}_k^* \middle| y_{1:k}\right)}{p\left(\mathbf{\Lambda}_k \middle| y_{1:k}\right)} = \frac{\left[\prod_{i=1}^{k} p\left(y_i \middle| \mathbf{x}_i, \mathbf{\Theta}_\mathbf{S}^*\right)\right] \cdot p\left(\mathbf{\Theta}_\mathbf{S}^*\right)}{\left[\prod_{i=1}^{k} p\left(y_i \middle| \mathbf{x}_i, \mathbf{\Theta}_\mathbf{S}\right)\right] \cdot p\left(\mathbf{\Theta}_\mathbf{S}\right)} \quad (9)$$

As it was previously mentioned in Section 4, such a move would necessitate the evaluation of all the likelihood terms corresponding to all the past measurements, leading to computational and memory requirements increasing in time. Furthermore, the evaluation would be biased since the joint distribution $p\left(\mathbf{x}_{0:k}, \mathbf{\Theta}_\mathbf{S}^* \middle| y_i\right)$ is not really estimated

by the particle filter and the past of the particles is highly degenerated. Each evaluation of $p\left(y_i \middle| \mathbf{x}_i, \mathbf{\Theta}_{\mathbf{S}}^*\right)$ would be distorted, as $\mathbf{x}_i$ is a very specific point of the state space, not really more likely than other points that have been forgotten. This is illustrated in the extreme situation of Figure 3 where a parallelepiped appears in front of the camera; two completely different likely orientations are represented, belonging to two separated modes I and II of the posterior distribution space. A few measurements later, as one of the two modes is kept from this past measurement, moves in the length dimension are possible (Mode I) or impossible (Mode II). The posterior distribution on the length will undoubtedly suffer from the arbitrariness of the selected mode.

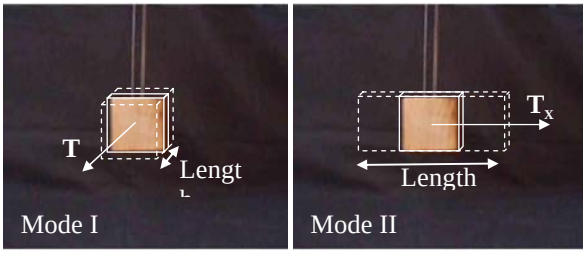


Figure 3. Arbitrary of the move evaluation

And yet, the deformation move K_D can be easily evaluated at the time of the 1st measurement where:

$$\rho\left(\mathbf{\Theta}_{\mathbf{S}}, \mathbf{\Theta}_{\mathbf{S}}^*\right) = \frac{p\left(y_1 \middle| \mathbf{x}_1, \mathbf{\Theta}_{\mathbf{S}}^*\right) \cdot p\left(\mathbf{\Theta}_{\mathbf{S}}^*\right)}{p\left(y_1 \middle| \mathbf{x}_1, \mathbf{\Theta}_{\mathbf{S}}\right) \cdot p\left(\mathbf{\Theta}_{\mathbf{S}}\right)} \qquad (10)$$

More, it can be fully integrated with the transition move K_T of the SAPF recursion so that the final probability ratio, corresponding to the joint MCMC move $K_{T \oplus D}$, is:

$$\rho\left(\mathbf{\Lambda}_1, \mathbf{\Lambda}_1^*\right) = \frac{p\left(\mathbf{x}_1^* \middle| \mathbf{x}_0\right) \cdot p\left(y_1 \middle| \mathbf{x}_1^*, \mathbf{\Theta}_{\mathbf{S}}^*\right) \cdot p\left(\mathbf{\Theta}_{\mathbf{S}}^*\right)}{p\left(\mathbf{x}_1 \middle| \mathbf{x}_0\right) \cdot p\left(y_1 \middle| \mathbf{x}_1, \mathbf{\Theta}_{\mathbf{S}}\right) \cdot p\left(\mathbf{\Theta}_{\mathbf{S}}\right)} \qquad (11)$$

The proposal distribution on the parameter moves is chosen in the same way as the one of the transition move described previously. Notice that, for the 1st measurement, the SAPF designed for the video tracking application strictly behaves as a simulated annealing importance sampling based on MCMC moves.

For the next measurements, an artificial diffusion of deformation is used through the successive layers of the SAPF. It is derived from Gaussian Kernel smoothing and solves the degeneracy problem by introducing a slow random walk evolution of the target dimensions.

At time k, from the previous posterior parameter samples $\left\{\theta_{k-1}^i, w_{k-1}^i\right\}_{i=1}^{N_S}$ approximating $p\left(\theta \middle| y_{1:k-1}\right)$, Gaussian kernel smoothing would consist in:

$$p\left(\theta_k \middle| \theta_{k-1}^i\right) \sim N\left(\theta_k \middle| \theta_{k-1}^i, h^2 V_{k-1}\right) \qquad (12)$$

where h is the kernel bandwidth (h can be chosen according to the optimal bandwidth that assumes an underlying Gaussian density) and V_{k-1} the empirical covariance matrix of $\left\{\theta_{k-1}^i, w_{k-1}^i\right\}_{i=1}^{N_S}$.

An adaptation is required to use such a method through the simulated annealing recursion of the SAPF where intermediate densities and samples are handled. At the layer m, the parameter undergoes the random walk:

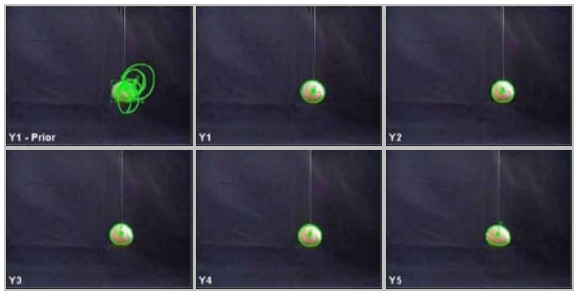
$$p\left(\theta_k^{(m)} \middle| \theta_{k-1}^{i\ (m-1)}\right) \sim N\left(\theta_k \middle| \tilde{\theta}_{k-1}^{i\ (m-1)}, h^2 V_{k-1}\right) \qquad (13)$$

where $\tilde{\theta}_{k-1}^{i\ (m-1)}$ is the nearest neighbour of $\theta_{k-1}^{i\ (m-1)}$ among the set of samples $\left\{\theta_{k-1}^i\right\}_{i=1}^{N_S}$. This empirical method allows a slow diffusion through the layers while still referencing from the samples at time $k-1$ and their related smooth Kernel density. In addition to (13), it is checked that the parameters remain in the support of the prior $p\left(\mathbf{\Theta}_{\mathbf{S}}^*\right)$.

4 Results

4.1 3D tracking

Figure 4 represents the tracking of a hemisphere during 75 measurements at the frame rate of 15 Hz. Here the radius of the hemisphere is known (20 cm). For a while, the target moves slowly and only shows its spherical face; after the 25th measurement, it moves faster while rotating. On the whole, the SAPF tracks the target well. Sometimes, slight imperfections occur when the extracted features of the image are mediocre, i.e. mainly when the edges are badly situated (measurements Y_{30}, Y_{50} and Y_{65}). Around Y_{35}, the tracker has not found the true orientation and has been satisfied with a local minimum. Subsequently, it manages to find again the true mode



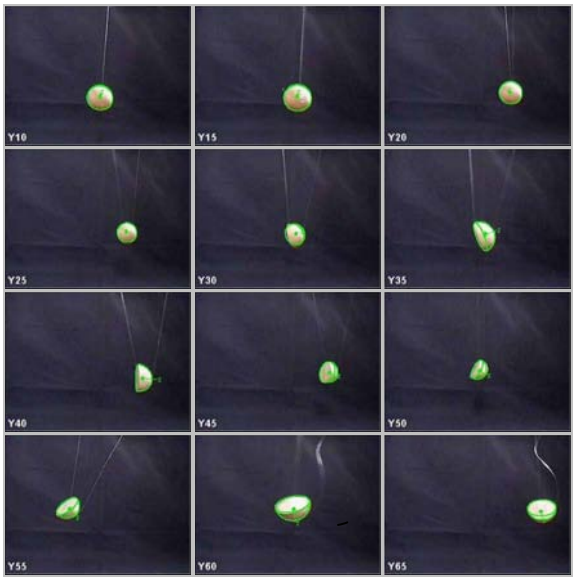
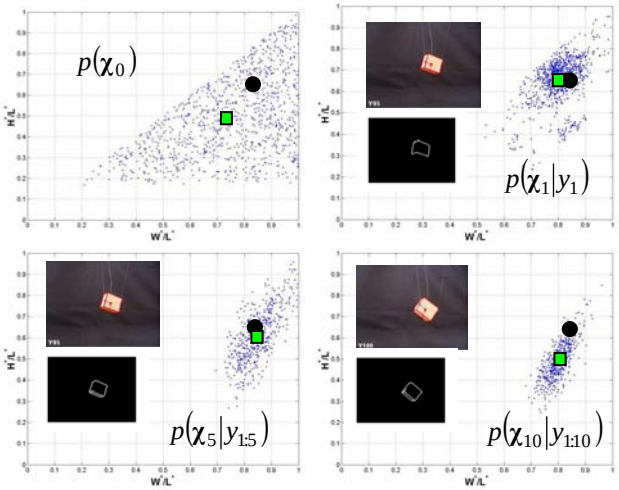


Figure 4. Hemisphere tracking.

4.2 3D hyper-parameter estimation

This example deals with the estimation of a 3D parameter, i.e. the length (*L*), the width (*W*) and the height (*H*) of a parallelepiped. The prior uncertainty on the 3 dimensions *L*, *W* and *H* is the same ~ U[10 cm 30 cm]. The proportions are represented in the upper left part of Figure 5, in the plane (W^*/L^*,H^*/L^*) where L^*, W^* and H^* are the sorted dimensions in a decreasing order. Black circles represent the true proportion while green squares represent the successive MMSE (Minimum Mean Square Error) estimator. After the 1st full MCMC based inference, the set of samples is tightening around the true point. The next hybrid inferences, based on MCMC and artificial diffusion, improves little by little the estimation quality. At time 20, the precision is around a few percents.



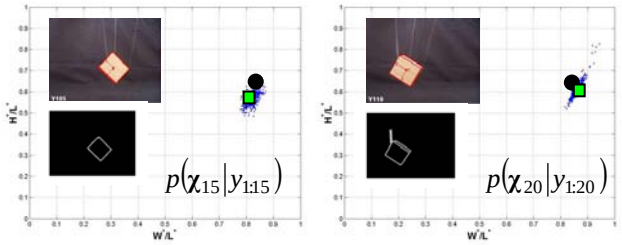


Figure 5. Parallelepiped tracking.

4.3 Joint tracking and identification

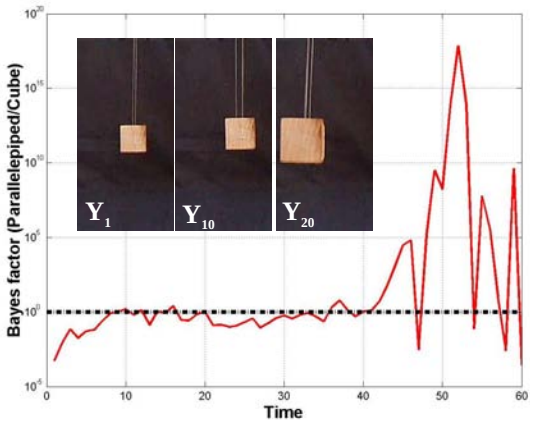


Figure 6. Bayes factor

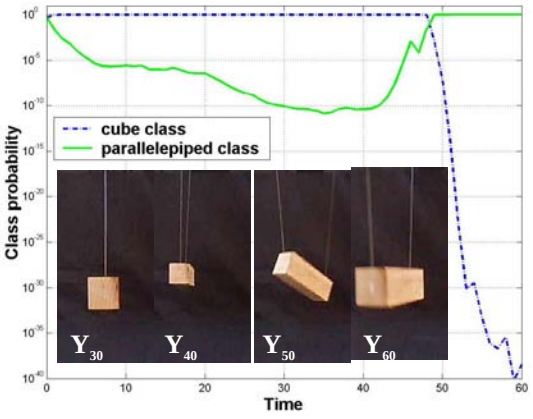


Figure 7. Class probability

In this second scenario, the real target is the long parallelepiped of Figure 1. For a while, it only shows is smallest square section (see measurements Y_1, Y_{10}, Y_{20}, and Y_{30} on Figure 6-7), then rotates at around time 40 (see Y_{40}) and turns its longest face towards the camera (see measurements Y_{50} and Y_{60}). There is now two equiprobable ($p_i = \frac{1}{2}$) classes: the cube class (r ~ U[10 cm 30 cm]) and the parallelepiped class (l = 39.5 cm - w, h ~ U[10 cm 30 cm]),

Figures 6-7 represent respectively the evolution of the Bayes factor and the class probability of each class. Until

time 40, the class cube seems to be more probable, essentially due to the fact that rotating the long parallelepiped would have shown more quickly its other faces. After time 40, the classification process changes its mind and it becomes more and more evident that it is in reality a long parallelepiped. Notice the few jolts of the Bayes factor that come from momentary mediocre edge extractions and a resulting distorted likelihood.

4.4 Computational requirements

The shape video JTI system, developed in Matlab and C, was run on a PC Pentium II (350 MHz). Using 200 particles and 5-6 annealing layers on average, a typical sequence of 15 seconds (225 frames - 15 Hz – image resolution 320x240 pixels) takes approximately 2 hours. However, the code is not optimised; better tunings of the SAPF and improvements of the numerical likelihood evaluation routines could provide, among others, sizeable gains.

5 Conclusions

Sequential Monte Carlo methods are powerful and efficient tools to perform joint tracking and identification. For this video application, a class specific filters architecture has been chosen. For each class, a sophisticated particle filter, called SAPF, computes the posterior distribution through simulated annealing layers, integrating various MCMC moves.

More originally, the issue of the fixed hyper-parameter estimation, here the geometric shape dimensions, is solved by combining two different techniques. The first one consists of MCMC moves that rescale both the trajectory and the shape; it benefits from an efficient statistic which summaries the trajectory with regard to moves. The second one is an artificial deformation diffusion of the shape.

Acknowledgements

This research was sponsored by the French MOD - DGA/DSFP, contract n° 02.60.00.060.

References

[1] P. Minvielle, A. D. Marrs, S. Maskell, and A. Doucet, *Joint Target Tracking and Identification – Part I: Sequential Monte Carlo Model-Based Approaches*, 8th International Conference on Information Fusion, ISIF, 2005

[2] N. Gordon, S. Maskell, and T. Kirubarajan, *Efficient Particle Filters for Joint Tracking and Classification*, Signal and Data Processing of Small Targets 2002, SPIE Vol. 4728, 2002

[3] S. Godsill, and T. Clapp, *Improvement Strategies for Monte Carlo Particle Filters*, Sequential Monte Carlo Methods in Practice, Springer-Verlag, 2001

[4] T. Clapp, *Statistical Methods in the Processing of Communications Data*, Ph.D. thesis, Cambridge University Engineering Department, 2000

[5] R. M. Neal, *Annealed Importance Sampling*, Technical Report n° 9805, Department of Statistics, University of Toronto, 1998

[6] C. Berzuini, and W. Gilks, *Resample-Move Filtering with Cross-Model Jumps*, Sequential Monte Carlo Methods in Practice, Springer-Verlag, 2001

[7] J. Deutscher, A. Blake, and I. Reid, *Articulated Body Motion Capture by Annealed Particle Filtering*, IEEE Conference on Computer Vision and Pattern Recognition, 2000

[8] *Sequential Monte Carlo Methods in Practice*, Ed. A. Doucet, J. F. G. de Freitas, and N. Gordon, Springer-Verlag: New York, 2001

[9] A. Doucet, S. Godsill, and C. Andrieu, *On Sequential Monte Carlo Sampling Methods for Bayesian Filtering*, Statis. Comput., vol. 10, n° 3, pp. 197-208, 2000

[10] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, *A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking*, IEEE Transactions on Signal Processing, Vol. 50, n° 2, February 2002

[11] Y. Bar-Shalom, and X. R. Li, *Estimation and Tracking – Principles, Techniques, and Software*, Artech House, 1993

[12] P. Fearnhead, *MCMC, Sufficient Statistics and Particle Filters, Report*, Journal of Computational and Graphical Statistics, Vol. 11, n° 4, pp. 848-862, December 2002