Sequential Monte Carlo Filtering for Multi-Aspect Detection/Tracking

Marcelo G. S. Bruno, Rafael V. de Araújo and Anton G. Pavlov
Instituto Tecnológico de Aeronáutica
São José dos Campos, SP 12228-900, Brazil
e-mail: {bruno,vidal,anton}@ele.ita.br, Phone:+55-12-3947-6906.

Abstract—We propose in this paper a mixed-state sequential Monte Carlo (SMC) filter for joint multiframe detection and tracking of multiaspect targets in cluttered image sequences. The proposed detector/tracker is a sampling/importance resampling (SIR) particle filter that uses resampling according to the weights to combat particle degeneracy and also includes an additional Metropolis-Hastings (MH) move step to avoid particle impoverishment. The dynamic models for target motion and target aspect and the statistical model for the spatially correlated background clutter are assumed as prior knowledge in the design of the filter. The performance of the algorithm is investigated using simulated image sequences generated from real infrared airbone radar (IRAR) data.

Keywords—Mixed-state particle filters, Markov Chain Monte Carlo, Gauss-Markov random fields, multiframe detection, multiaspect target tracking.

TABLE OF CONTENTS

1. INTRODUCTION

We present in this paper an algorithm for joint multiframe detection and tracking of multiaspect targets in two-dimensional (2D) image sequences. We consider the situation when the target of interest is obscured by structured clutter and the clutter-free target aspect changes randomly from frame to frame as a result of rotational motion and/or variations in the conditions of observation of the target.

The conventional contact/association approach to target tracking, see e.g. [1], is based on a suboptimal decoupling of the detection and tracking tasks. Typically, a preliminary single frame detection stage uses segmentation, clustering and correlation filtering to generate initial estimates of the true position of a potential target of interest. These preliminary estimates are subsequently associated to a multiframe tracker or, alternatively, are discarded as false measurements originating from clutter. The sequential estimator used for tracking from validated measurements is usually a linearized Kalman-Bucy filter (KBf). Although efficient for highly visible targets, this suboptimal association of correlation detectors and KBf trackers has been shown to perform poorly in scenarios of low target-to-clutter ratios [2]. In this paper, we propose an alternative Bayesian approach that eliminates the preliminary single frame correlation detector and enables integrated, multiframe detection and tracking taking full advantage of the knowledge of the dynamic models for target motion and target aspect and also incorporating the statistical model for the spatially-correlated clutter background.

In previous work [3], we proposed a grid-based joint detector/tracker assuming discrete-valued hidden Markov models (HMMs) for both the target's kinematic state (e.g. position and velocity) and the target's aspect state. The HMM-filter solution in [3] had the disadvantages, however, of great computational cost and lack of flexibility to capture more complex target motion. To circumvent these limitations, we allow the hidden aspect state of the target in this paper to take values only on a finite set $\mathcal{I}$, but assume that the target's position and velocity at each frame are continuous random variables.

Each symbol in the finite set of values assumed by the target's aspect state may be interpreted as a pointer to one possible target template obtained by an affine transformation of the target's base image. In order to integrate detection and tracking, we extend further the finite alphabet $\mathcal{I}$ to include an additional dummy absent target state. The evolution over time of the kinematic and aspect states is described then by a coupled dynamical model where the probability of transition from one aspect state to another between frame $n-1$ and frame n is dependent on the kinematic state of the target at frame $n-1$. Conversely, the probability density function of the kinematic state at frame n conditioned on the kinematic state at the previous frame is also dependent on the target's aspect changes between frames $n-1$ and n. The observation model consists of a nonlinear function that maps a given target centroid position at frame n onto a spatial distribution of pixels centered in the centroid and with intensity that is dependent on the current target aspect state. Finally, the time-varying, aspect-changing target image is superimposed to a structured clut-

0-7803-8870-4/05/$20.00
IEEEAC paper # XXX

ter background whose spatial correlation is captured using a noncausal Gauss-Markov random field (GMrf) model [4], [5]. The parameters of the GMrf clutter model are estimated online from the available data using an approximate maximum likelihood algorithm [6].

With the framework described in the previous paragraph, the detection/tracking problem reduces, from a Bayesian perspective, to computing in a recursive fashion the joint posterior distribution of the target's discrete-valued aspect state and the target's continuous-valued kinematic state at frame n conditioned on the observed frames from instant 0 up to instant n. Due to the nonlinear nature of the observation and possibly motion models, it is however impossible to compute such posterior distribution analytically. We resort then to a sequential Monte Carlo approach known as particle filtering [7], [8] also referred to in the computer vision literature as the condensation algorithm [9].

In summary, the basic idea in particle filtering is to represent the posterior distribution of interest by a set of properly weighted samples , also known as particles, such that, at each instant n, the weighted average of the particles converges (in some statistical sense) to the minimum square error (MMSE) estimate of the hidden state vector given the observed frames. Using a standard mixed-state importance sampling (IS) strategy [10], we sample the particle population sequentially from the (coupled) prior model for target motion and target aspect change and then update the weights of the sampled particles using the likelihood function that incorporates the target signature and background clutter models. A particle resampling step with replacement according to the sample weights [11], [12] is added to control the increase in the variance of the weights. Finally, we also add a Markov Chain Monte Carlo (MCMC) [13] move step [14] after resampling to rejuvenate the particle population and prevent particle impoverishment.

This paper is divided into 5 sections. Section 1 is this Introduction. Section 2 reviews the models for target motion and target aspect that underly our derivations. We also introduce in Section 2 the observation model and the associated likelihood function. In Section 3, we describe the joint detector/tracker [1]. In Section 4, we evaluate the performance of the proposed filter using simulated infrared airbone radar (IRAR) [15] data. Finally, we summarize in Section 5 the contributions of our work.

2. The Model

We introduce in this section the dynamic models for target motion and target aspect that underly the derivation of the joint detector/tracker in this paper. For simplicity, we restrict our discussion to the situation when there is at most one target of interest present in the imaged scene. Throughout the paper, we use lowercase letters to to denote both random variables/vectors and realizations (samples) of random variables/vectors; the proper interpretation is implied in context. We use lowercase p to denote probability density functions (pdfs) and uppercase P to denote the probability mass function of a discrete random variable. The symbol $Pr(A)$ is used to denote the probability of an event A in the σ-algebra of the sample space.

State Variables Let n be a non-negative integer number and let superscript T denote the transpose of a vector or matrix. We define the target's kinematic state at frame n as the four-dimensional vector $\mathbf{s}_n = [x_n \ \dot{x}_n \ y_n \ \dot{y}_n]^T$, that collects the positions, x_n and y_n, and the velocities, $\dot{x}_n$ and $\dot{y}_n$, of the target's centroid in a system of 2D Cartesian coordinates (x, y). We model $\mathbf{s}_n$ as a continuous random vector taking values in $\Re^4$. Conversely, the target's aspect state at frame n, denoted z_n, is modeled as a discrete random variable taking values in the finite set $\mathcal{I} = \{0, 1, 2, 3, \ldots, K\}$ where the event $\{z_n = 0\}$ indicates that no target is present in the imaged scene at frame n. Each "present target"state i, $i = 1, 2, \ldots, K$, is, on the other hand, a pointer to one possible template model in a target aspect library accounting for rotation, scaling and/or shearing of the target's mother template.

Observation Model

The raw sensor measurements at instant n are sampled and processed to form a 2D digital sensor image, referred to as a *frame*. We represent the nth frame then by the $L \times M$ matrix

$$\mathbf{Y}_n = \mathbf{H}(\mathbf{s}_n^*, z_n) + \mathbf{V}_n \qquad (1)$$

where the matrix $\mathbf{V}_n$ represents the background clutter, and the matrix $\mathbf{H}(\mathbf{s}_n^*, z_n)$ is the clutter-free target image model, which is a function of the 2D pixel location of the target centroid, $\mathbf{s}_n^*$ and the target aspect state, z_n. The two-dimensional random vector $\mathbf{s}_n^*$ takes values on the finite sensor grid $\mathcal{L} = \{(r, j) \mid 1 \leq r \leq L, \, 1 \leq j \leq M\}$ and is obtained from the four-dimensional continuous-valued state vector $\mathbf{s}_n$ by making

$$s_n^*(1) = \text{round}(\frac{s_{n,1}(1)}{\xi_1}) \qquad (2)$$

$$s_n^*(2) = \text{round}(\frac{s_{n,2}(1)}{\xi_2}) \qquad (3)$$

where ξ_1 and ξ_2 are the image resolutions respectively in the coordinates x and y.

Clutter-Free Target Model We assume that, any given frame, for any aspect state z_n, the clutter-free image of a target that is present is contained in a bounded rectangular region of size $(r_i + r_s + 1) \times (l_i + l_s + 1)$. In this notation, r_i and r_s denote the maximum vertical pixel distances in the target image when we move away, respectively up and down, from the target centroid. Analogously, l_i and l_s are the maximum horizontal pixel distances in the target image when we move away, respectively left and right, from the target centroid.

[1] A simpler version of the algorithm in Section 3, assuming the target is always present and suitable for tracking-only, was introduced in reference [3]

For each pixel centroid position $\mathbf{s}_n^* = (r_n,\, j_n) \in \mathcal{L}$, the nonlinear function $\mathbf{H}(:,:)$ in (1) returns a spatial distribution of (real-valued) pixel intensities $\{a_{k,\,l}(z_n)\}$, $-r_i \leq k \leq r_s$, $-l_i \leq l \leq l_s$, centered at $(r_n,\, j_n)$ and dependent on the aspect state z_n. Formally, we write

$$\mathbf{H}(r_n,\, j_n,\, z_n) = \sum_{k=-r_i}^{r_s} \sum_{l=-l_i}^{l_s} a_{k,\,l}(z_n) \mathbf{E}_{r_n+k, j_n+l} \quad (4)$$

where $\mathbf{E}_{g,\,t}$ is an $L \times M$ matrix whose entries are all equal to zero, except for the element $(g,\, t)$ which is equal to 1.

For a given fixed template model $z_n = i \in \mathcal{I}$, the coefficients $\{a_{k,l}(i)\}$ in (4) are referred to as the target *signature parameters* corresponding to that particular template. The signature coefficients are the product of a binary parameter $b_{k,\,l}(z_n) \in \mathcal{B} = \{0,\, 1\}$, that defines the target shape for each aspect state, and a real coefficient $\phi_{k,\,l}(z_n) \in \Re$, that specifies the pixel intensities of the target, again for the various states in the alphabet $\mathcal{I}$. For simplicity, we assume that the pixel intensities and shapes are deterministic and known at each frame for a given value of z_n. In particular if z_n takes the value 0 denoting absence of target, then the function $\mathbf{H}(:,:)$ in (1) reduces to the identically zero matrix $\mathbf{0}_{L\times M}$, meaning the observations consist only of clutter.

Remark To write (4), we assumed that the target is sufficiently far from the borders of the image grid so that we do not have to worry about boundary conditions. Boundary effects can be easily taken into account by changing the summation limits accordingly in (4) for centroid locations near the borders.

Clutter Model We describe the 2D spatial correlation of the background clutter using a noncausal, spatially homogeneous Gauss-Markov random field (GMrf) model [4]. The random clutter returns at frame n, $V_n(r,\, j)$, $1 \leq r \leq L$, $1 \leq j \leq M$, are described then by the 2D finite difference equation

$$\begin{aligned} V_n(r,\, j) &= \beta_{v,n}^c \left[V_n(r-1,\, j) + V_n(r+1,\, j)\right] \\ &+ \quad \beta_{h,n}^c \left[V_n(r,\, j-1) + V_n(r,\, j+1)\right] + \varepsilon_n(r,\, j) \end{aligned} \quad (5)$$

where $E\left[V_n(r,\, j)\, \varepsilon_n(k,\, l)\right] = \sigma_{c,n}^2\, \delta_{r-k,\, j-l}$. We use subscript n in the notation for the parameters β_v^c, β_h^c and σ_c to emphasize that the clutter parameters may be time-variant and change from frame to frame. The assumption of zero-mean clutter implies a pre-processing of the data that subtracts the mean of the background.

Target Motion and Aspect Models

The random sequence $\{(\mathbf{s}_n,\, z_n)\}$, $n \geq 0$, is modeled as a first-order hidden Markov model (HMM) specified by a mixed initial probability density function (pdf)

$$p(\mathbf{s}_0,\, z_0) = p(\mathbf{s}_0)\, P(z_0) \quad (6)$$

and by the mixed transition pdf

$$\begin{aligned} p(\mathbf{s}_n, z_n \mid \mathbf{s}_{n-1}, z_{n-1}) &= p(\mathbf{s}_n \mid z_n,\, \mathbf{s}_{n-1},\, z_{n-1}) \\ &\times \quad P(z_n \mid z_{n-1}, \mathbf{s}_{n-1}) \end{aligned} \quad (7)$$

for $n \geq 1$. Note that the mixed densities on the left-hand side of equations (6) and (7) are defined as actual probability density functions in the continuous-valued variables and probability mass functions in the discrete random variables, i.e., if s is a continuous random variable taking values in $\Re$ and z is a discrete random variable taking values in the finite set $\mathcal{I}$, then

$$p(s', z') = \frac{\partial}{\partial s'} Pr(\{s \leq s'\} \cap \{z = z'\}).$$

Aspect Model Based on the observation model, we introduce the *augmented centroid lattice* $\widehat{\mathcal{L}} = \{(r, j)\colon -r_s + 1 \leq r \leq L + r_i,\ -l_s + 1 \leq j \leq M + l_i\}$ that collects all possible values of the target centroid position for which at least one target pixel may still lie inside the sensor's image. In the sequel, let $\mathbf{T}$ be a $K \times K$ transition probability matrix such that $T(i, j) \geq 0$ for any $i, j = 1, 2, \ldots, K$ and

$$\sum_{i=1}^{K} T(i, j) = 1 \qquad \forall j = 1, \ldots, K.$$

The probability of a change in the target's aspect between frames $n-1$ and n from state j to state i, $Pr(\{z_n = i\} \mid \{z_{n-1} = j\}\,,\, \mathbf{s}_{n-1})$, is modeled in this paper as

$$\left\{ \begin{array}{cc} T(i,j)\, Pr(\left\{\mathbf{s}_n^* \in \widehat{\mathcal{L}}\right\} \mid \mathbf{s}_{n-1}, \{z_{n-1} = j\}) & i, j = 1, \ldots, K \\ 1 - Pr(\left\{\mathbf{s}_n^* \in \widehat{\mathcal{L}}\right\} \mid \mathbf{s}_{n-1}, \{z_{n-1} = j\}) & i = 0; j \neq 0 \\ \frac{p_a}{K} & i \neq 0; j = 0 \\ 1 - p_a & i = 0; j = 0\ . \end{array} \right. \quad (8)$$

In (8), the parameter p_a denotes the probability of a new target appearing in the scene once the previous target becomes absent. By setting $Pr(\{z_n = i\} \mid \{z_{n-1} = 0\}\,,\, \mathbf{s}_{n-1}) = pa/K$, $i = 1, \ldots, K$, we assume that, given that a new target became present at frame n, there is a uniform probability of that new target assuming any of the K possible aspect states. On the other hand, by making

$$\begin{aligned} &Pr(\{z_n = 0\} \mid \{z_{n-1} = j\}\,,\, \mathbf{s}_{n-1}) = \\ &1 - Pr(\left\{\mathbf{s}_n^* \in \widehat{\mathcal{L}}\right\} \mid \mathbf{s}_{n-1}, \{z_{n-1} = j\}), \end{aligned} \quad (9)$$

we are implicitly assuming that a target may become absent only if it leaves the image grid.

Motion Model For simplicity, we make an additional assumption that, unless the aspect state changes from zero (absent target state) to a nonzero (present target) state, the distribution of the current kinematic state $\mathbf{s}_n$ conditioned on the previous kinematic state $\mathbf{s}_{n-1}$ is independent of the current and previous aspect states, z_{n-1} and z_n. Specifically, let $f_s(\mathbf{s}_n \mid \mathbf{s}_{n-1})$ be a valid conditional pdf (not necessarily Gaussian) and let $f_0(\mathbf{s}_n)$ be a valid marginal pdf. We make $p(\mathbf{s}_n \mid z_n,\, \mathbf{s}_{n-1}, z_{n-1})$ equal to

$$\left\{ \begin{array}{cc} f_0(\mathbf{s}_n) & \text{if } z_n \neq 0 \text{ and } z_{n-1} = 0 \\ f_s(\mathbf{s}_n \mid \mathbf{s}_{n-1}) & \text{otherwise.} \end{array} \right. \quad (10)$$

Usually, we make $f_0(\mathbf{s}_n)$ a non-informative prior (e.g., a continuous uniform pdf). With the assumption of independence between the kinematic and aspect states for present targets, we also write, for any $j = 1, \ldots, K$,

$$Pr(\left\{\mathbf{s}_n^* \in \widehat{\mathcal{L}}\right\} \mid \mathbf{s}_{n-1}, \{z_{n-1} = j\}) = \int_{\left\{\mathbf{s}_n \mid \mathbf{s}_n^* \in \widehat{\mathcal{L}}\right\}} f_s(\mathbf{s}_n \mid \mathbf{s}_{n-1})\, d\mathbf{s}_n \ . \tag{11}$$

Let $\mathcal{N}(\mathbf{s} - \mathbf{a}, \mathbf{P})$ denote the multivariable normal function of argument $\mathbf{s}$, mean $\mathbf{a}$ and covariance matrix $\mathbf{P}$. Without loss of generality, we assume in this paper that $f_s(\mathbf{s}_n \mid \mathbf{s}_{n-1})$ represents a linear, white-Gaussian-noise acceleration motion [1] on the 2D plane. i.e.,

$$f_s(\mathbf{s}_n \mid \mathbf{s}_{n-1}) = \mathcal{N}(\mathbf{s}_n - \tilde{\mathbf{F}}\,\mathbf{s}_{n-1}, \tilde{\mathbf{Q}}) \tag{12}$$

where

$$\tilde{\mathbf{F}} = \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{F} \end{bmatrix}, \qquad \tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix} . \tag{13}$$

Let q be a positive real number and denote by Δ the sampling period. Matrices $\mathbf{F}$ and $\mathbf{Q}$ in (13) are given then by [1]

$$\mathbf{F} = \begin{bmatrix} 1 & \Delta \\ 0 & 1 \end{bmatrix}, \qquad \mathbf{Q} = q \begin{bmatrix} \frac{\Delta^3}{3} & \frac{\Delta^2}{2} \\ \frac{\Delta^2}{2} & \Delta \end{bmatrix} . \tag{14}$$

Likelihood Function

Let $\mathbf{y}_n$ and $\mathbf{h}(\mathbf{s}_n^*, z_n)$ be the 1D representations respectively of the image frame $\mathbf{Y}_n$ and the clutter free target image $\mathbf{H}(\mathbf{s}_n^*, z_n)$ in (1), obtained by scanning the matrices row by row and sequentially stacking the scanned rows in a long vector. Similarly, let $\mathbf{v}_n$ be the long vector representation of the matrix $\mathbf{V}_n$ in (5) and let $\mathbf{\Sigma}_v = E\left[\mathbf{v}_n \mathbf{v}_n^T\right]$ denote its associated covariance matrix. For a GMrf background as in (5), the likelihood function for a fixed template state $z_n = \tilde{z}$, $\tilde{z} \in \{1, 2, 3, \ldots, K\}$ is given by

$$p(\mathbf{y}_n \mid \mathbf{s}_n,\, \tilde{z}) = p(\mathbf{y}_n \mid \mathbf{s}_n,\, 0) \exp\left[\frac{2\lambda(\mathbf{s}_n, \tilde{z}) - \rho(\tilde{z})}{2\sigma_{c,n}^2}\right] . \tag{15}$$

where $\rho(\tilde{z})$ is a target energy term given by

$$\rho(\tilde{z}) = \mathbf{h}^T(\mathbf{s}_n^*, \tilde{z})(\sigma_{c,n}^2 \mathbf{\Sigma}_v^{-1})\mathbf{h}(\mathbf{s}_n^*, \tilde{z}) \tag{16}$$

and $\lambda(\mathbf{s}_n,\, \tilde{z})$ is a data-dependent term such that

$$\lambda(\mathbf{s}_n,\, \tilde{z}) = \mathbf{y}_n^T(\sigma_{c,n}^2 \mathbf{\Sigma}_v^{-1})\mathbf{h}(\mathbf{s}_n^*, \tilde{z}) \ . \tag{17}$$

Finally, $p(\mathbf{y}_n \mid \mathbf{s}_n,\, 0)$ in (15) is the likelihood of the absent target state which reduces to

$$p(\mathbf{y}_n \mid \mathbf{s}_n,\, 0) = \frac{1}{(2\pi)^{\frac{LM}{2}}\ \left[\det(\mathbf{\Sigma}_v)\right]^{1/2}} \exp(-\frac{1}{2}\mathbf{y}_n^T \mathbf{\Sigma}_v^{-1} \mathbf{y}_n)$$

for all $\mathbf{s}_n \in \Re^4$.

Writing the difference equation (5) in compact matrix notation, it can be shown [4], [5], [6] by the application of the principle of orthogonality that $\mathbf{\Sigma}_v^{-1}$ has a *block-tridiagonal* structure of the form

$$\sigma_{c,n}^2 \mathbf{\Sigma}_v^{-1} = \mathbf{I}_L \otimes (\mathbf{I}_M - \beta_{h,n}^c \mathbf{K}_M) + \mathbf{K}_L \otimes (-\beta_{v,n}^c \mathbf{I}_M) \tag{18}$$

where $\otimes$ denotes the Kronecker product, $\mathbf{I}_J$ is $J \times J$ identity matrix, and $\mathbf{K}_J$ is a $J \times J$ matrix whose entries $K_J(k,\, l) = 1$ if $\mid k - l \mid = 1$ and are equal to zero otherwise.

Using the block-banded structure of $\mathbf{\Sigma}_v^{-1}$ in (18), it can be shown that $\lambda(\mathbf{s}_n,\, \tilde{z})$ in (17) may be computed by the expression

$$\lambda(\mathbf{s}_n, \tilde{z}) = \sum_{k=-r_i}^{r_s} \sum_{l=-l_i}^{l_s} a_{k,l}(\tilde{z}) d(s_n^*(1) + k,\, s_n^*(2) + l) \tag{19}$$

where $s_n^*(i)$, $i = 1, 2$, are obtained respectively from equations (2) and (3), and $d(r, j)$ is the output of the differential operator

$$\begin{aligned} d(r,j) &= Y_n(r,j) - \beta_{h,n}^c \left[Y_n(r, j-1) + Y_n(r, j+1)\right] \\ &- \beta_{v,n}^c \left[Y_n(r-1, j) + Y_n(r+1, j)\right] \end{aligned} \tag{20}$$

with Dirichlet (identically zero) boundary conditions. Equation (19) is valid for $r_i + 1 \leq s_n^*(1) \leq L - r_s$ and $l_i + 1 \leq s_n^*(2) \leq M - l_s$. For centroid positions close to the image borders, the summation limits in (19) must be varied accordingly, as shown in [2].

<u>Remark: Estimation of Clutter Parameters</u> We assume that the time-varying clutter parameters $\beta_{h,n}$, $\beta_{v,n}$ and $\sigma_{c,n}$ are *unknown* to the tracking filter and must be adaptively estimated from the image sequence. Ideally, the clutter parameters should be jointly estimated with the hidden state variables $\mathbf{s}_n$ and z_n in a Bayesian framework. For computational simplicity though, we use in this paper a *suboptimal* approach to clutter adaptation where the unknown GMrf parameters corresponding to each available sensor frame $\mathbf{Y}_n$ are assumed deterministic and are independently estimated from frame to frame using a single frame variation of the *approximate maximum likelihood* (AML) parameter estimation algorithm introduced in [6]. For a description of the clutter parameter estimation algorithm, we refer the readers to section V in the reference [3].

3. JOINT DETECTOR/TRACKER

We describe in this section the proposed SMC detector/tracker. We begin the section with a description of the importance sampling step. Next, we detail the resampling and MCMC move steps. Finally, we present the multiframe detection test and the kinematic state estimation algorithm.

Sequential Importance Sampling Step

Given a sequence of observed lexicographed frames $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$, our goal is to represent the mixed posterior

$p(\mathbf{s}_n,\, z_n \mid \mathbf{y}_{1:n})$ at step n by a properly weighted set of particles $\left\{\mathbf{s}_n^{(j)},\, z_n^{(j)}\right\}$, $1 \le j \le N_p$, with associated weights $\left\{w_n^{(j)}\right\}$ such that, as N_p goes to infinity, the weighted averages of the particles converge to the minimum mean-square error (MMSE) estimates of the hidden states under the true posterior distributions. Using $p(\mathbf{s}_n, z_n \mid \mathbf{s}_{n-1}, z_{n-1})$ in (7) as importance function, a mixed-state sequential importance sampling (SIS) algorithm for the recursive generation of the desired properly weighted set is [8], [10]

1) Initialization For $j = 1, \ldots, N_p$

- Draw $\mathbf{s}_0^{(j)} \sim p(\mathbf{s}_0)$, and $z_0^{(j)} \sim P(z_0)$.
- Make $w_0^{(j)} = 1/N_p$ and set $n = 1$.

2) Importance Sampling For $j = 1, \ldots, N_p$

- Draw $\tilde{z}_n^{(j)} \sim P(z_n \mid z_{n-1}^{(j)}, \mathbf{s}_{n-1}^{(j)})$ according to (8).
- Draw $\tilde{\mathbf{s}}_n^{(j)} \sim p(\mathbf{s}_n \mid \tilde{z}_n^{(j)}, \mathbf{s}_{n-1}^{(j)},\, z_{n-1}^{(j)})$ according to (10).
- Compute the importance weights

$$\tilde{w}_n^{(j)} \propto w_{n-1}^{(j)}\; p(\mathbf{y}_n \mid \tilde{\mathbf{s}}_n^{(j)},\, \tilde{z}_n^{(j)}) \qquad \sum_{j=1}^{N_p} \tilde{w}_n^{(j)} = 1$$

using the likelihood function model in Section 2.

End-for

Remark From a computational point of view, the main advantage of the importance sampling solution above over the grid-based filter described in [3] is that the grid-based tracker evaluates the likelihood function in the entire discretized state space, whereas the importance sampling algorithm requires the evaluation of the likelihood function only for each sample in the current set of particles. Since the number of particles is usually smaller than the total number of pixels per frame times the total number of possible target templates, the computational savings may be significant.

Resampling and Move Steps

The major drawback of raw sequential importance sampling as described in the previous subsection is the increase over time in the variance of the weights $\tilde{w}_n^{(j)}$ leading in the limit to the phenomenon known as “particle degeneracy” [8], i.e., after a few steps, only a small number of particles will have normalized weights close to one whereas the majority of the particles will have negligible weight. In order to mitigate particle degeneracy, we follow the approcah in [11], [12] and resample from the existing particle population with replacement according to the particle weights so that high-weight particles are multiplied while low-weight particles are discarded. Concretely, in the particular application studied in this paper, we draw at instant n a set of indices

$$i^{(j)} \sim \{1,\, 2,\, \ldots,\, N_p\} \quad \text{with} \quad Pr(\left\{i^{(j)} = l\right\}) = \tilde{w}_n^{(l)} \quad (21)$$

and build a new particle set $\left\{\bar{\mathbf{s}}_n^{(j)},\, \bar{z}_n^{(j)}\right\}$, $j = 1, \ldots, N_p$ such that

$$(\bar{\mathbf{s}}_n^{(j)},\, \bar{z}_n^{(j)}) = (\tilde{\mathbf{s}}_n^{(i^{(j)})},\, \tilde{z}_n^{(i^{(j)})}).$$

Move Step The resampling step described before helps to avoid particle degeneracy, but also leads to an undesirable loss of particle diversity as resampling may result in multiple copies of only a few or, in the limit, only one particle. In order to rejuvenate the particle population without altering its statistics, we add a Markov Chain Monte Carlo (MCMC) move step, see [14], after the resampling routine. Specifically, for $j = 1, \ldots, N_p$, let

$$(\bar{\mathbf{s}}_{0:n}^{(j)}, \bar{z}_{0:n}^{(j)}) = (\mathbf{s}_{0:n-1}^{(i^{(j)})},\, \tilde{\mathbf{s}}_n^{(i^{(j)})}, z_{0:n-1}^{(i^{(j)})}, \tilde{z}_n^{(i^{(j)})}) \qquad (22)$$

be the resampled particle trajectories for the kinematic and aspect states from instant 0 up to instant n. The resulting trajectories $(\bar{\mathbf{s}}_{0:n}^{(j)}, \bar{z}_{0:n}^{(j)})$ are approximately distributed, see [16], according to the mixed posterior $p(\mathbf{s}_{0:n}, z_{0:n}|\mathbf{y}_{1:n})$. A possible strategy for sequential regeneration of the particle population is to apply, for $j = 1, \ldots, N_p$, a Markov chain transition kernel

$$k(\hat{\mathbf{s}}_n^{(j)}, \hat{z}_n^{(j)}|\bar{\mathbf{s}}_n^{(j)}, \bar{z}_n^{(j)}) \qquad (23)$$

which is invariant to the conditional mixture pdf $p\left(\mathbf{s}_n, z_n|\bar{\mathbf{s}}_{0:n-1}^{(j)},\, \bar{z}_{0:n-1}^{(j)}, \mathbf{y}_{1:n}\right)$. Provided that such invariance condition is satisfied, the moved sample trajectories

$$(\mathbf{s}_{0:n}^{(j)}, z_{0:n}^{(j)}) = \left(\bar{\mathbf{s}}_{0:n-1}^{(j)}, \hat{\mathbf{s}}_n^{(j)}, \bar{z}_{0:n-1}^{(j)},\, \hat{z}_n^{(j)}\right)$$

remain (approximately) distributed according to $p(\mathbf{s}_{0:n}, z_{0:n} \mid \mathbf{y}_{1:n})$. A possible Metropolis-Hastings (MH) [13] strategy to build a Markov Chain with the desired stationary distribution is as follows:

For $j = 1, \ldots, N_p$

- Draw $\hat{z}_n^{(j)} \sim P(z_n \mid \bar{z}_{n-1}^{(j)}, \bar{\mathbf{s}}_{n-1}^{(j)})$ according to (8).
- Draw $\hat{\mathbf{s}}_n^{(j)} \sim p(\mathbf{s}_n \mid \hat{z}_n^{(j)}, \bar{\mathbf{s}}_{n-1}^{(j)},\, \bar{z}_{n-1}^{(j)})$ according to (10).
- Draw $u \sim \mathcal{U}([0,1])$.

If $u \le \min\left\{1,\, \frac{p(\mathbf{y}_n \mid \hat{\mathbf{s}}_n^{(j)},\, \hat{z}_n^{(j)})}{p(\mathbf{y}_n \mid \bar{\mathbf{s}}_n^{(j)},\, \bar{z}_n^{(j)})}\right\}$, then $(\mathbf{s}_n^{(j)}, z_n^{(j)}) = (\hat{\mathbf{s}}_n^{(j)}, \hat{z}_n^{(j)})$.

Else $(\mathbf{s}_n^{(j)}, z_n^{(j)}) = (\bar{\mathbf{s}}_n^{(j)}, \bar{z}_n^{(j)})$.

- Reset $w_n^{(j)} = \frac{1}{N_p}$.

End-for

Multiframe Detection and Tracking

Let H_1 denote the hypothesis that the target is present at frame n and, conversely, let H_0 denote the hypothesis that the target is absent. Given the set particles $(\mathbf{s}_n^{(j)}, z_n^{(j)})$ approximately distributed according to $p(\mathbf{s}_n, z_n \mid \mathbf{y}_{1:n})$ at instant n, we compute the Monte Carlo estimate $\widehat{Pr}(\{z_n = 0\} \mid \mathbf{y}_{1:n})$ of $Pr(\{z_n = 0\} \mid \mathbf{y}_{1:n})$ and apply the minimum probability of error [17] multiframe detection test

$$\widehat{Pr}(\{z_n = 0\} \mid \mathbf{y}_{1:n}) \underset{H_1}{\overset{H_0}{\gtrless}} 1 - \widehat{Pr}(\{z_n = 0\} \mid \mathbf{y}_{1:n}) \quad (24)$$

to decide whether the target is present or not at frame n. If the hypothesis H_1 is declared true, the estimate $\hat{s}_{n|n}$ of the target's kinematic state at frame n is computed then using the Monte Carlo approximation of $E\left[\mathbf{s}_n \mid \mathbf{y}_{1:n}, \{z_n \neq 0\}\right]$.

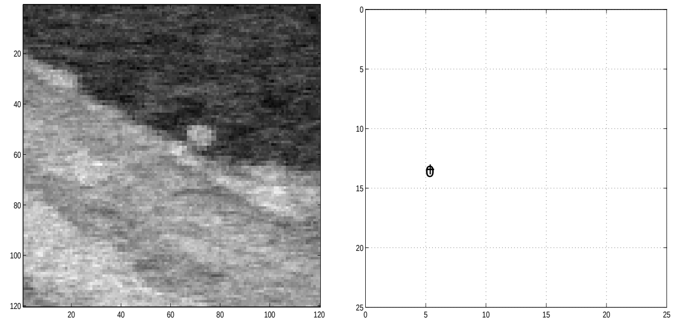
Particle Update

Finally, we make $n = n + 1$ and go back to the importance sampling step to draw a new set of particles $\left\{(\tilde{\mathbf{s}}_n^{(j)}, \tilde{z}_n^{(j)})\right\}$ with updated weights $\tilde{w}_n^{(j)}$. The recursion continues as long as there are image data available to be processed.

4. Simulation Results

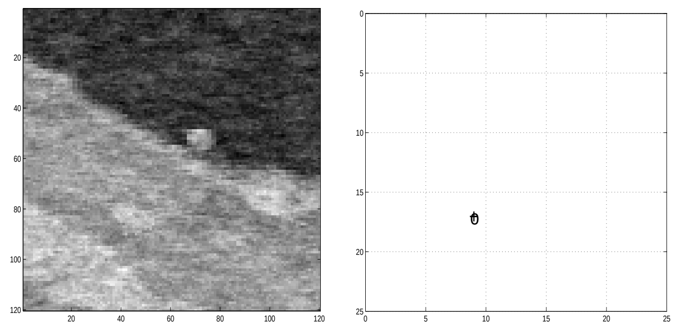
We study next the performance of the proposed SMC detector/tracker using simulated image sequences generated from real infrared airborne radar (IRAR) data. The IRAR intensity imagery is from the MIT Lincoln Laboratory's database and was obtained from the Center for Imaging Sciences at Johns Hopkins University. To simulate the target, we took an artificial template representing a military vehicle and generated a library of affine transformations of that template using composite operations of rotation, scaling and shearing. We then added the artificial target to the background sequence with the target centroid position changing from frame to frame according to the linear white noise acceleration model in (12), see Section 2. For the target simulations in this paper, we set $\Delta = 0.04$ and $q = 8$. The background clutter for the moving target sequence was simulated by adding a sequence of synthetic GMrf samples to a matrix of previously stored local means extracted from the database imagery. The GMrf samples were synthetized using correlation and prediction error variance parameters estimated from real data.

In order to simulate the target's aspect dynamics, we initialized the target template state z_0 with a randomly-selected choice from the template library and then changed the aspect over time according to a first-order Markov chain. At any given frame, the true aspect of the target is unknown to the tracker. The target pixel intensity is on the other hand time-invariant and known and was set according to a desired low level of contrast between the template and the background. The initial position of the target is assumed uniformly distributed in a certain region of the image.



(a) (b)

Figure 1. (a) First frame of the cluttered target sequence, PTCR = 10.6 dB, (b) tracking results: actual target position ('+'), estimated target position ('o').

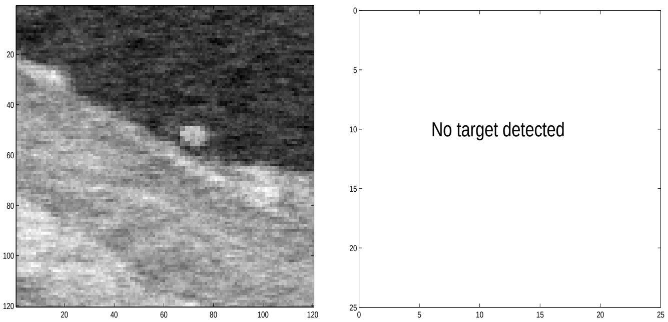


(a) (b)

Figure 2. (a) Tenth frame of the cluttered target sequence, PTCR = 10.6 dB, with target translation, rotation, scaling, and shearing; (b) tracking results: actual target position ('+'), estimated target position ('o').

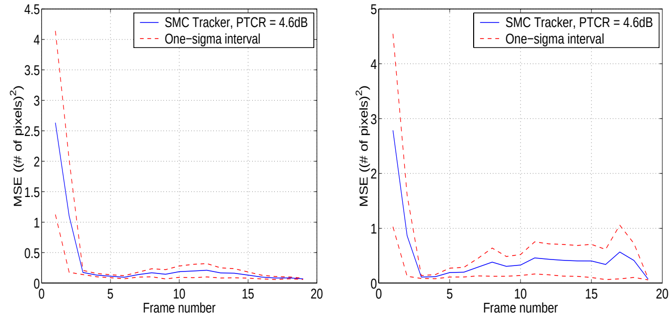
Using $N_p = 10,000$ particles, we tracked the simulated target over 50 consecutive frames to verify the capability of the algorithm : (1) to acquire the target and track it both at the center of the grid and near the image's boundaries where parts of the target are no longer present in the scene; (2) to detect when the target has completely left the image indicating no false detections at the frames where no target is present; and, (3) to detect when a new target enters the image, acquire it and track it sucessfully. A complete video demo of the operation of the algorithm over 50 frames can be seen at http://www.ele.ita.br/∼bruno. In the sequel, we show the detection/tracking results for a few selected frames.

Figure 1(a) shows the initial frame of the sequence with the target centered in the (quantized) coordinates $(65, 23)$ and superimposed to clutter. The simulated peak target-to-clutter ratio (PTCR) is 10.6 dB. The actual target position (indicated by a cross sign, '+') and the estimated position (indicated by a circle,'o') are shown in Figure 1(b). Next, Figure 2(a) shows the tenth frame in the image sequence. Note that the target from frame 1 has now undergone a random change in aspect in addition to translational motion. The tracking results corresponding to frame 10 are shown in Figure 2(b).



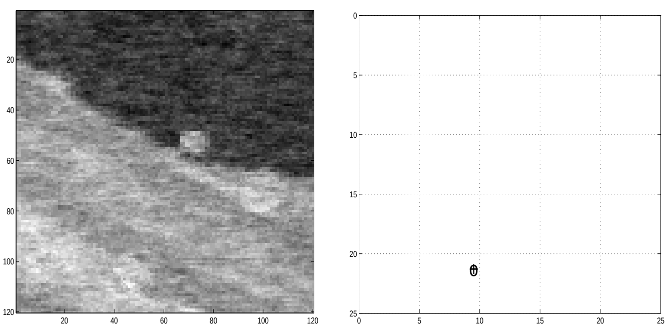
(a) (b)

Figure 3. (a) 36th frame of the cluttered target sequence with no target present, (b) Detection result indicating absence of target.



(a) (b)

Figure 5. Mean-square position estimation error in number of pixels squared, excluding divergent realizations: (a) x coordinate, (b) y coordinate; divergence rate = 5 % (out of 120 Monte Carlo trials), PTCR = 4.6 dB.



(a) (b)

Figure 4. (a) Fifty-first frame of the cluttered target sequence, PTCR = 10.6 dB, with a new target present in the scene; (b) tracking results: actual target position ('+'), estimated target position ('o').

In this particular simulation, the target leaves the scene at frame 31 and no target reappears until frame 37. The SMC tracker accurately detects the instant when the target disappears and shows no false alarms over the 6 absent target frames as illustrated in Figures 3(a) and (b) where we show respectively the clutter+background-only thirty-sixth frame and the corresponding tracking results indicating in this case that no target has been detected. Finally, when a new target re-appears, it is accurately acquired by the SMC algorithm. The final simulated frame with the new target at position $(104, 43)$ is shown for illustration purposes in Figure 4(a). Figure 4(b) shows the corresponding tracking results for the same frame.

In order to have a quantitative assessment of the tracking performance, we ran a Monte Carlo simulation with 120 independent trials. We used again 10,000 particles, but lowered the the peak target-to-clutter ratio to 4.6 dB to assess the performance of the algorithm in situations of heavily obscured (stealthy) targets. The proposed SMC tracker, operating over 20 frames, diverged in 6 out of 120 trials giving an approximate divergence rate of 5 %. The mean square error (MSE) for the position estimation errors in number of pixels squared *excluding the divergent tracks* are shown in Figures 5(a) and (b). The MSE values in number of pixels squared are superimposed to their respective one-sigma uncertainty intervals assuming 114 Monte Carlo trials (i.e., 120 minus the 6 divergent realizations). Note that, in the non-divergent realizations, the tracker shows an initial position estimation error that falls over time as more frames are processed. Note also that the uncertainty in the error is greater in the y coordinate.

Remark In this particular paper, we quantified only the tracking performance of the proposed algorithm. In our simulations, the target was correctly declared present by the multi-frame detection test over all 20×120 frames. For a complete assessment of detection performance, one could run a continuous simulation with targets leaving and entering the scene and estimate a receiver operating characteristic (ROC) curve by introducing a detection threshold in the hypotheses test (24) and varying that threshold over a wide range of values. Due to the very low false alarm rates that were in practice obtained in our trials, an accurate estimation of the ROC curve would require very large-scale Monte Carlo simulations. We leave that for future work.

5. Summary

The conventional correlation filter/Kalman filter association approach to target detection/tracking in cluttered images has severe limitations in situations of low target-to-clutter ratio. In this paper, we proposed to overcome those limitations using an alternative sequential Monte Carlo (SMC) algorithm that enables joint multiframe detection and tracking and fully incorporates the statistical models for target motion, target aspect, and clutter. The target's aspect state at each frame was modeled as a discrete random variable that takes values in a finite set of target template models. Conversely, the target's position and velocity were modeled as continuous random variables. In order to integrate detection and tracking, we introduced an additional dummy aspect state that represented the absence of a target at a given frame. The spatial correlation of the background clutter was described using a noncausal Gauss-Markov random field (GMRf) model.

The proposed detector/tracker in this paper is a mixed-state sampling/importance resampling (SIR) filter enhanced with a Metropolis-Hastings move step to avoide sample impoverishment. Simulation results with synthetic images generated from real infrared airborne radar (IRAR) data show that, barring an approximately 5 % divergence rate, the mixed-state SIR filter operating with 10,000 particles shows good tracking performance in situations of heavily obscured targets and dense clutter.

6. ACKNOWLEDGEMENTS

The work of the second and third authors was supported by CAPES, Brazil. The work of the first author was partially supported by a previous grant from FAPESP, São Paulo, Brazil, under process number 98/13409-2.

REFERENCES

[1] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques.* YBS, Storrs, CT, 1995.

[2] M. G. S. Bruno and J. M. F. Moura, "Multiframe detection/tracking in clutter: optimal performance," *IEEE Transactions on Aerospace and Electronic Systems*, vol.37, n.3, pp. 925-946, July 2001.

[3] M. G. S. Bruno, "Bayesian methods for multiaspect target tracking in image sequences," *IEEE Transactions on Signal Processing*, vol. 52, n.7, pp. 1848-1861, July 2004.

[4] J. M. F. Moura and N. Balram,"Recursive Structure of Noncausal Gauss Markov Random Fields," *IEEE Trans. on Inf. Theory* IT-38(2), pp. 334-354, March 1992.

[5] J. M. F. Moura and M. G. S. Bruno, "DCT/DST and Gauss-Markov fields: conditions for equivalence," *IEEE Transactions on Signal Processing*, vol.46, n.9, pp 2571-2574, September 1998.

[6] J. M. F. Moura and N. Balram, "Noncausal Gauss-Markov random fields: parameter structure and estimation," *IEEE Transactions on Information Theory*, IT-39(4), pp. 1333-1355, July 1993.

[7] A. Doucet, J. F. G. Freitas, and N. J. Gordon, "An introduction to sequential Monte Carlo methods," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, J. F. G. Freitas, and N. J. Gordon, Editors. New York: Springer-Verlag, 2001.

[8] M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. on Signal Processing*, vol.50, n.2, pp. 174-188, February 2002.

[9] M. Isard and A. Blake, "Condensation - conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol.28, n.1, pp. 5-28, 1998.

[10] M. Isard and A. Blake, "A mixed-state condensation tracker with automatic model switching," *Proceedings of the 6th International Conference on Computer Vision* 1998, pp. 107-112.

[11] N. J. Gordon, D. J. Salmond, and C. Ewing, "Bayesian state estimation for tracking and guidance using the bootstrap filter", *Journal of Guidance, Control, and Dynamics*, vol.18, n.6, pp. 1434-1443, 1995.

[12] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputations," *Journal of the American Statistical Association*, 90(430), pp. 567-576, 1995.

[13] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer-Verlag New York, Inc., 1999.

[14] W. R. Gilks and C. Berzuini,"Following a moving target _ Monte Carlo inference for dynamic Bayesian models," *J. R. Statist. Soc. B*, vol.63, pp. 127-146, 2001.

[15] J. K. Bounds, *The Infrared Airbone Radar Sensor Suite*, RLE Technical Report No. 610, Cambridge, Massachusetts, December 1996.

[16] J. S. Liu, R. Chen, and T. Logvinenko,"A theoretical framework for sequential importance sampling with resampling," Chapter 11 in *Sequential Monte Carlo Methods in Practice*, A. Doucet, J. F. G. Freitas, e N. J. Gordon, Editores. New York: Springer-Verlag, 2001.

[17] L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, 1992.

Marcelo G. S. Bruno *received the Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University in 1998. Currently, he is an Assistant Professor at Instituto Tecnológico de Aeronáutica, São José dos Campos, Brazil. His research interests include hidden Markov Models, Bayesian networks, Markov random fields, sequential Monte Carlo methods, Markov Chain Monte Carlo, and their applications in target detection and tracking, communications, machine learning, image processing, computer vision, and robotics. Dr. Bruno is a member of several IEEE societies and currently serves as an Associate Editor for the IEEE Signal Processing Letters.*

Rafael Vidal de Araújo *received a B.S. degree in electrical engineering from Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, Brazil, in 2002 and is currently pursuing an M.S. degree in electrical engineering also at ITA. His research interests include target detection and tracking, digital image processing, adaptive filtering, and stochastic control.*

Anton G. Pavlov *received the B.S. degree and, shortly thereafter, the Specialist title in electrical engineering from Kharkov Military University, Kharkov, Ukraine, in 2001. In 2004, he received the M.S. degree in electrical engineering from Instituto Tecnológico de Aeronáutica, São José dos Campos, Brazil, where he is currently pursuing a Ph.D. degree, also in electrical engineering. His research interests are on probabilistic inference in Bayesian networks and sequential Monte Carlo methods. Mr. Pavlov is a student member of the IEEE.*