

Research Article

Sequential Monte Carlo Methods for Joint Detection and Tracking of Multiaspect Targets in Infrared Radar Images

Marcelo G. S. Bruno, Rafael V. Araújo, and Anton G. Pavlov

Instituto Tecnológico de Aeronáutica, São José dos Campos, SP 12228, Brazil

Correspondence should be addressed to Marcelo G. S. Bruno, bruno@ele.ita.br

Received 30 March 2007; Accepted 7 August 2007

Recommended by Yvo Boers

We present in this paper a sequential Monte Carlo methodology for joint detection and tracking of a multiaspect target in image sequences. Unlike the traditional contact/association approach found in the literature, the proposed methodology enables integrated, multiframe target detection and tracking incorporating the statistical models for target aspect, target motion, and background clutter. Two implementations of the proposed algorithm are discussed using, respectively, a resample-move (RS) particle filter and an auxiliary particle filter (APF). Our simulation results suggest that the APF configuration outperforms slightly the RS filter in scenarios of stealthy targets.

Copyright © 2008 Marcelo G. S. Bruno et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

This paper investigates the use of sequential Monte Carlo filters [1] for joint multiframe detection and tracking of randomly changing multiaspect targets in a sequence of heavily cluttered remote sensing images generated by an infrared airborne radar (IRAR) [2]. For simplicity, we restrict the discussion primarily to a single target scenario and indicate briefly how the proposed algorithms could be modified for multi-object tracking.

Most conventional approaches to target tracking in images [3] are based on suboptimal decoupling of the detection and tracking tasks. Given a reference target template, a two-dimensional (2D) spatial matched filter is applied to a single-frame of the image sequence. The pixel locations where the output of the matched filter exceeds a pre-specified threshold are treated then as initial estimates of the true position of detected targets. Those preliminary position estimates are subsequently assimilated into a multiframe tracking algorithm, usually a linearized Kalman filter, or alternatively discarded as false alarms originating from clutter.

Depending on its level of sophistication, the spatial matched filter design might or might not take into account the spatial correlation of the background clutter and random distortions of the true target aspect compared to the refer-

ence template. In any case, however, in a scenario with dim targets in heavily cluttered environments, the suboptimal association of a single-frame matched filter detector and a multiframe linearized tracking filter is bound to perform poorly [4].

As an alternative to the conventional approaches, we introduced in [5, 6] a Bayesian algorithm for joint multiframe detection and tracking of known targets, fully incorporating the statistical models for target motion and background clutter and overcoming the limitations of the usual association of single-frame correlation detectors and Kalman filter trackers in scenarios of stealthy targets. An improved version of the algorithm in [5, 6] was later introduced in [7] to enable joint detection and tracking of targets with unknown and randomly changing aspect. The algorithms in [5–7] were however limited by the need to use discrete-valued stochastic models for both target motion and target aspect changes, with the “absent target” hypothesis treated as an additional dummy aspect state. A conventional hidden Markov model (HMM) filter was used then to perform joint minimum probability of error multiframe detection and maximum a posteriori (MAP) tracking for targets that were declared present in each frame. A smoothing version of the joint multiframe HMM detector/tracker, based essentially on a 2D version of the forward-backward (Baum-Welch)

algorithm, was later proposed in [4]. Furthermore, we also proposed in [4] an alternative tracker based on particle filtering [1, 8] which, contrary to the original HMM tracker in [7], assumed a continuous-valued kinematic (position and velocity) state and a discrete-valued target aspect state. However, the particle filter algorithm in [4] enabled tracking only (assuming that the target was always present in all frames) and used decoupled statistically independent models for target motion and target aspect.

To better capture target motion, we drop in this paper the previous constraint in [5–7] and, as in the later sections of [4], allow the unknown 2D position and velocity of the target to be continuous-valued random variables. The unknown target aspect is still modeled however as a discrete random variable defined on a finite set \mathcal{L} , where each symbol is a pointer to a possibly rotated, scaled, and/or sheared version of the target's reference template. In order to integrate detection and tracking, building on our previous HMM work in [7], we extend the set \mathcal{L} to include an additional dummy state that represents the absence of a target of interest in the scene. The evolution over time of the target's kinematic and aspect states is described then by a *coupled* stochastic dynamic model where the sequences of target positions, velocities, and aspects are *mutually dependent*.

Contrary to alternative feature-based trackers in the literature, the proposed algorithm in this paper detects and tracks the target directly from the raw sensor images, processing pixel intensities only. The clutter-free target image is modeled by a nonlinear function that maps a given target centroid position into a spatial distribution of pixels centered around the (quantized) centroid position, with shape and intensity being dependent on the current target aspect. Finally, the target is superimposed to a structured background whose spatial correlation is captured by a noncausal Gauss-Markov random field (GMRF) model [9–11]. The GMRF model parameters are adaptively estimated from the observed data using an approximate maximum likelihood (AML) algorithm [12].

Given the problem setup described in the previous paragraph, the optimal solution to the integrated detection/tracking problem requires the recursive computation at each frame n of the joint posterior distribution of the target's kinematic and aspect states conditioned on all observed frames from instant 0 up to instant n . Given, however, the inherent nonlinearity of the observation and (possibly) motion models, the exact computation of that posterior distribution is generally not possible. We resort then to mixed-state particle filtering [13] to represent the joint posterior by a set of weighted samples (or particles) such that, as the number of particles goes to infinity, their weighted average converges (in some statistical sense) to the desired minimum mean-square error (MMSE) estimate of the hidden states. Following a sequential importance sampling (SIS) [14] approach, the particles may be drawn recursively from the coupled prior statistical model for target motion and aspect, while their respective weights may be updated recursively using a likelihood function that takes into account the models for the target's signature and for the background clutter.

We propose two different implementations for the mixed-state particle filter detector/tracker. The first imple-

mentation, which was previously discussed in a conference paper (see [15]) is a resample-move (RS) filter [16] that uses particle resampling [17] followed by a Metropolis-Hastings move step [18] to combat both particle degeneracy and particle impoverishment (see [8]). The second implementation, which was not included in [15], is an auxiliary particle filter (APF) [19] that uses the current observed frame at instant n to preselect those particles at instant $n - 1$ which, when propagated through the prior dynamic model, are more likely to generate new samples with high likelihood. Both algorithms are original with respect to the previous particle filtering-based tracking algorithm that we proposed in [4], where the problem of joint detection and tracking with coupled motion and aspect models was not considered.

Related work and different approaches in the literature

Following the seminal work by Isard and Blake [20], particle filters have been extensively applied to the solution of visual tracking problems. In [21], a sequential Monte Carlo algorithm is proposed to track an object in video subject to model uncertainty. The target's aspect, although unknown, is assumed, however, to be fixed in [21], with no dynamic aspect change. On the other hand, in [22], an adaptive appearance model is used to specify a time-varying likelihood function expressed as a Gaussian mixture whose parameters are updated using the EM [23] algorithm. As in our work, the algorithm in [22] also processes image intensities directly, but, unlike our problem setup, the observation model in [22] does not incorporate any information about spatial correlation of image pixels, treating instead each pixel as independent observations. A different Bayesian algorithm for tracking nonrigid (randomly deformable) objects in three-dimensional images using multiple conditionally independent cues is presented in [24]. Dynamic object appearance changes are captured by a mixed-state shape model [13] consisting of a discrete-valued cluster membership parameter and a continuous-valued weight parameter. A separate kinematic model is used in turn to describe the temporal evolution of the object's position and velocity. Unlike our work, the kinematic model in [24] is assumed statistically independent of the aspect to model.

Rather than investigating solutions to the problem of multiaspect tracking of a single target, several recent references, for example, [25, 26], use mixture particle filters to tackle the different but related problem of detecting and tracking an unknown number of multiple objects with different but fixed appearance. The number of terms in the nonparametric mixture model, that represents the posterior of the unknowns, is adaptively changed as new objects are detected in the scene and initialized with a new associated observation model. Likewise, the mixture weights are also recursively updated from frame to frame in the image sequence.

Organization of the paper

The paper is divided into 6 sections. Section 1 is this introduction. In Section 2, we present the coupled model

for target aspect and motion and review the observation and clutter models focusing on the GMRf representation of the background and the derivation of the associated likelihood function for the observed (target + clutter) image. In Section 3, we detail the proposed detector/tracker in the RS and APF configurations. The performance of the two filters is discussed in Section 4 using simulated infrared airborne radar (IRAR) data. A preliminary discussion on multitarget tracking is found in Section 5, followed by an illustrative example with two targets. Finally, we present in Section 6 the conclusions of our work.

2. THE MODEL

In the sequel, we present the target and clutter models that are used in this paper. We use lowercase letters 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 functions (pmfs) of discrete random variables. 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 nonnegative integer number and let superscript T denote the transpose of a vector or matrix. The kinematic state of the target at frame n is defined as the four-dimensional continuous (real-valued) random 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) . On the other hand, the target's aspect state at frame n , denoted by z_n , is assumed to be a discrete random variable that takes values in the finite set $\mathcal{L} = \{0, 1, 2, 3, \dots, K\}$, where the symbol "0" is a dummy state that denotes that the target is absent at frame n , and each symbol $i, i = 1, \dots, K$, is in turn a pointer to one possibly rotated, scaled, and/or sheared version of the target's reference template.

2.1. Target motion and aspect models

The random sequence $\{\mathbf{s}_n, z_n\}, n \geq 0$, is modeled as first-order Markov process specified by the pdf of the initial kinematic state $p(\mathbf{s}_0)$, the transition pdf $p(\mathbf{s}_n | z_n, \mathbf{s}_{n-1}, z_{n-1})$, the transition probabilities $\Pr(\{z_n = i\} | \{z_{n-1} = j\}, \mathbf{s}_{n-1})$, $(i, j) \in \mathcal{L} \times \mathcal{L}$, and the initial probabilities $\Pr(\{z_0 = i\}), i \in \mathcal{L}$.

Aspect change model

Assume that, at any given frame, for any aspect state z_n , the clutter-free target image lies within a bounded rectangle of size $(r_i + r_s + 1) \times (l_i + l_s + 1)$. In this notation, r_i and r_s denote the maximum 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.

Assume also that each image frame has the size of $L \times M$ pixels. We introduce next the *extended grid* $\hat{\mathcal{L}} = \{(r, j) : -r_s + 1 \leq r \leq L + r_i, -l_s + 1 \leq j \leq M + l_i\}$ that contains all possible target centroid locations for which at least one target pixel still lies in the sensor image. Next, let \mathbf{G} be a matrix of size $K \times K$ such that $G(i, j) \geq 0$ for any $i, j = 1, 2, \dots, K$ and

$$\sum_{i=1}^K G(i, j) = 1 \quad \forall j = 1, \dots, K. \quad (1)$$

Assuming that a transition from a "present target" state to the "absent target" state can only occur when the target moves out of the image, we model the probability of a change in the target's aspect from the state j to the state i , $\Pr(\{z_n = i\} | \{z_{n-1} = j\}, \mathbf{s}_{n-1})$, as

$$\begin{aligned} G(i, j) \Pr(\{\mathbf{s}_n^* \in \hat{\mathcal{L}}\} | \mathbf{s}_{n-1}, \{z_{n-1} = j\}), \quad i, j = 1, \dots, K, \\ 1 - \Pr(\{\mathbf{s}_n^* \in \hat{\mathcal{L}}\} | \mathbf{s}_{n-1}, \{z_{n-1} = j\}), \quad i = 0, j \neq 0, \\ \frac{p_a}{K}, \quad i \neq 0, \quad j = 0, \\ 1 - p_a, \quad i = 0, \quad j = 0, \end{aligned} \quad (2)$$

where the two-dimensional vector $\mathbf{s}_n^* = (x_n^*, y_n^*)$ denotes the quantized target centroid position defined on the extended image grid and obtained from the four-dimensional continuous kinematic state \mathbf{s}_n by making

$$\begin{aligned} x_n^* &= \text{round}\left(\frac{x_n}{\zeta_1}\right), \\ y_n^* &= \text{round}\left(\frac{y_n}{\zeta_2}\right), \end{aligned} \quad (3)$$

where ζ_1 and ζ_2 are the spatial resolutions of the image, respectively, in the directions x and y . The parameter p_a in (2) denotes in turn the probability of a new target entering the image once the previous target became absent. For simplicity, we restrict the discussion in this paper to the situation where there is at most one single target of interest present in the scene at each image frame. The specification $\Pr(\{z_n = i\} | \{z_{n-1} = 0\}, \mathbf{s}_{n-1}) = p_a/K, i = 1, \dots, K$, corresponds to assuming the worst-case scenario where, given that a new target entered the scene, there is a uniform probability that the target will take any of the K possible aspect states. Finally, the term $1 - \Pr(\{\mathbf{s}_n^* \in \hat{\mathcal{L}}\} | \mathbf{s}_{n-1}, \{z_{n-1} = j\})$ in (2) is the probability of a target moving out of the image at frame n given its kinematic and aspect states at frame $n - 1$.

Motion model

For simplicity, we assume that, except in the situation where there is a transition from the "absent target" state to the "present target" state, the conditional pdf $p(\mathbf{s}_n | z_n, \mathbf{s}_{n-1}, z_{n-1})$ is independent of the current and previous aspect states, respectively, z_n and z_{n-1} . In other words, unless $z_{n-1} = 0$ and $z_n \neq 0$, we make

$$p(\mathbf{s}_n | z_n, \mathbf{s}_{n-1}, z_{n-1}) = f_s(\mathbf{s}_n | \mathbf{s}_{n-1}), \quad (4)$$

where $f_s(\mathbf{s}_n | \mathbf{s}_{n-1})$ is an arbitrary pdf (not necessarily Gaussian) that models the target motion. Otherwise, if $z_{n-1} = 0$ and $z_n \neq 0$, we reset the target's position and make

$$p(\mathbf{s}_n | z_n, \mathbf{s}_{n-1}, z_{n-1}) = f_0(\mathbf{s}_n), \quad (5)$$

where $f_0(\mathbf{s}_n)$ is typically a noninformative (e.g., uniform) prior pdf defined in a certain region (e.g., upper-left corner) of the image grid. Given the independence assumption in (4), it follows that, for any $j = 1, \dots, K$,

$$\Pr(\{\mathbf{s}_n^* \in \hat{\mathcal{L}}\} | \mathbf{s}_{n-1}, \{z_{n-1} = j\}) = \int_{\{\mathbf{s}_n | \mathbf{s}_n^* \in \hat{\mathcal{L}}\}} f_s(\mathbf{s}_n | \mathbf{s}_{n-1}) d\mathbf{s}_n. \quad (6)$$

2.2. Observation model and likelihood function

Next, we discuss the target observation model. Previous references mentioned in Section 1, for example, [21, 22, 24–26], are concerned mostly with video surveillance of near objects (e.g., pedestrian or vehicle tracking), or other similar applications (e.g., face tracking in video). For that class of applications, effects such as object occlusion are important and must be explicitly incorporated into the target observation model. In this paper by contrast, the emphasis is on a different application, namely, detection and tracking of small, quasipoint targets that are observed by remote sensors (usually mid-to high-altitude airborne platforms) and move in highly structured, generally smooth backgrounds (e.g., deserts, snow-covered fields, or other forms of terrain). Rather than modeling occlusion, our emphasis is instead on additive natural clutter.

Image frame model

Assuming a single target scenario, the n th frame in the image sequence is modeled as the $L \times M$ matrix:

$$\mathbf{Y}_n = \mathbf{H}(\mathbf{s}_n^*, z_n) + \mathbf{V}_n, \quad (7)$$

where the matrix \mathbf{V}_n represents the background clutter and $\mathbf{H}(\mathbf{s}_n^*, z_n)$ is a nonlinear function that maps the quantized target centroid position, $\mathbf{s}_n^* = (x_n^*, y_n^*)$, (see (3)) into a spatial distribution of pixels centered at \mathbf{s}_n^* and specified by a set of deterministic and known *target signature coefficients* dependent on the aspect state z_n . Specifically, we make [4]

$$\mathbf{H}(x_n^*, y_n^*, z_n) = \sum_{k=-r_s}^{r_s} \sum_{l=-l_s}^{l_s} a_{k,l}(z_n) \mathbf{E}_{x_n^*+k, y_n^*+l}, \quad (8)$$

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 (8) are the target signature coefficients responding 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 $\phi_{k,l}(s_n) \in \mathfrak{R}$, 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 each possible value of z_n . In particular, if z_n takes the value 0 denoting absence of target, then the function $\mathbf{H}(\cdot, \cdot)$ in (7) reduces to the identically zero matrix, indicating that sensor observations consist of clutter only.

Remark 1. Equation (8) assumes that the target's template is entirely located within the sensor image grid. Otherwise, for targets that are close to the image borders, the summation limits in (8) must be changed accordingly to take into account portions of the target that are no longer visible.

Clutter model

In order to describe the spatial correlation of the background clutter, we assume that, after suitable preprocessing to remove the local means, the random field $V_n(r, j)$, $1 \leq r \leq L$, $1 \leq j \leq M$, is modeled as a first-order noncausal Gauss-Markov random field (GMrf) described by the finite difference equation [9]

$$V_n(r, j) = \beta_{v,n}^c [V_n(r-1, j) + V_n(r+1, j)] + \beta_{h,n}^c [V_n(r, j-1) + V_n(r, j+1)] + \varepsilon_n(r, j), \quad (9)$$

where $E\{V_n(r, j)\varepsilon_n(k, l)\} = \sigma_{c,n}^2 \delta_{r-k, j-l}$, with $\delta_{i,j} = 1$ if $i = j$ and zero otherwise. The symbol $E\{\cdot\}$ denotes here the expectation (or expected value) of a random variable/vector.

Likelihood function model

Let \mathbf{y}_n , $\mathbf{h}(\mathbf{s}_n^*, z_n)$, and \mathbf{v}_n be the one-dimensional equivalent representations, respectively, of \mathbf{Y}_n , $\mathbf{H}(\mathbf{s}_n^*, z_n)$ and \mathbf{V}_n in (7), obtained by row-lexicographic ordering. Let also $\Sigma_v = E[\mathbf{v}_n \mathbf{v}_n^T]$ denote the covariance matrix associated with the random vector \mathbf{v}_n , assumed to have zero mean after appropriate preprocessing. For a GMrf model as in (9), the corresponding likelihood function for a fixed aspect state $z_n = \tilde{z}$, $\tilde{z} \in \{1, 2, 3, \dots, K\}$, is given by [4]

$$p(\mathbf{y}_n | \mathbf{s}_n, \tilde{z}) = p(\mathbf{y}_n | \mathbf{s}_n, 0) \exp \left[\frac{2\lambda(\mathbf{s}_n, \tilde{z}) - \rho(\mathbf{s}_n, \tilde{z})}{2\sigma_{c,n}^2} \right] \quad (10)$$

where

$$\lambda(\mathbf{s}_n, \tilde{z}) = \mathbf{y}_n^T (\sigma_{c,n}^2 \Sigma_v^{-1}) \mathbf{h}(\mathbf{s}_n^*, \tilde{z}) \quad (11)$$

is referred to in our work as the *data term* and

$$\rho(\mathbf{s}_n, \tilde{z}) = \mathbf{h}^T(\mathbf{s}_n^*, \tilde{z}) (\sigma_{c,n}^2 \Sigma_v^{-1}) \mathbf{h}(\mathbf{s}_n^*, \tilde{z}) \quad (12)$$

is called the *energy term*. On the other hand, for $z_n = 0$, $p(\mathbf{y}_n | \mathbf{s}_n, z_n)$ reduces to the likelihood of the absent target state, which corresponds to the probability density function of \mathbf{y}_n assuming that the observation consists of clutter only, that is,

$$p(\mathbf{y}_n | \mathbf{s}_n, 0) = \frac{1}{(2\pi)^{LM/2} [\det(\Sigma_v)]^{1/2}} \exp \left(-\frac{1}{2} \mathbf{y}_n^T \Sigma_v^{-1} \mathbf{y}_n \right). \quad (13)$$

Writing the difference equation (9) in compact matrix notation, it can be shown [9–11] by the application of the principle of orthogonality that Σ_v^{-1} has a *block-tridiagonal* structure of the form

$$\sigma_{c,n}^2 \Sigma_v^{-1} = \mathbf{I}_L \otimes (\mathbf{I}_M - \beta_{h,n}^c \mathbf{B}_M) + \mathbf{B}_L \otimes (-\beta_{v,n}^c \mathbf{I}_M), \quad (14)$$

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

Using the block-banded structure of Σ_v^{-1} in (14), it can be further shown that $\lambda(\mathbf{s}_n, \tilde{\mathbf{z}})$ may be evaluated as the output of a modified 2D *spatial matched filter* using the expression

$$\lambda(\mathbf{s}_n, \tilde{\mathbf{z}}) = \sum_{k=-r_i}^{r_s} \sum_{l=-l_i}^{l_s} a_{k,l}(\tilde{\mathbf{z}}) d(s_n^*(1) + k, s_n^*(2) + l), \quad (15)$$

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

$$\begin{aligned} d(r, j) = & Y_n(r, j) - \beta_{h,n}^c [Y_n(r, j-1) + Y_n(r, j+1)] \\ & - \beta_{v,n}^c [Y_n(r-1, j) + Y_n(r+1, j)] \end{aligned} \quad (16)$$

with Dirichlet (identically zero) boundary conditions.

Similarly, the energy term $\rho(\mathbf{s}_n, \tilde{\mathbf{z}})$ can be also efficiently computed by exploring the block-banded structure of Σ_v^{-1} . The resulting expression is the difference between the autocorrelation of the signature coefficients $\{a_{k,l}\}$ and their lag-one cross-correlations weighted by the respective GMrf model parameters $\beta_{h,n}^c$ or $\beta_{v,n}^c$. Before we leave this section, we make two additional remarks.

Remark 2. As before, (15) 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 (15) must be varied accordingly (see [6] for details).

Remark 3. Within our framework, a crude non-Bayesian single frame maximum likelihood target detector could be built by simply evaluating the likelihood map $p(\mathbf{y}_n | \mathbf{s}_n, z_n)$ for each aspect state z_n and finding the maximum over the image grid of the sum of likelihood maps weighted by the a priori probability for each state z_n (usually assumed to be identical). A target would be considered present then if the weighted likelihood peak exceeded a certain threshold. In that case, the likelihood peak would also provide an estimate for the target location. The integrated joint detector/tracker presented in Section 3 outperforms, however, the decoupled single-frame detector discussed in this remark by fully incorporating the dynamic motion and aspect motion into the detection process and enabling multiframe detection within the context of a track-before-detect philosophy.

3. PARTICLE FILTER DETECTOR/TRACKER

3.1. Sequential importance sampling

Given a sequence of observed frames $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, our goal is to generate, at each instant n , a properly weighted set of

samples (or particles) $\{\mathbf{s}_n^{(j)}, z_n^{(j)}\}$, $j = 1, \dots, N_p$, with associated weights $\{w_n^{(j)}\}$ such that, according to some statistical criterion, as N_p goes to infinity,

$$\sum_{j=1}^{N_p} w_n^{(j)} \left[\left(\mathbf{s}_n^{(j)} \right)^T z_n^{(j)} \right]^T \rightarrow E \left\{ \left[\mathbf{s}_n^T z_n \right]^T | \mathbf{y}_{1:n} \right\}. \quad (17)$$

A possible mixed-state sequential importance sampling (SIS) strategy (see [4, 13]) for the recursive generation of the particles $\{\mathbf{s}_n^{(j)}, z_n^{(j)}\}$ and their proper weights is described in the algorithm below.

(1) *Initialization* For $j = 1, \dots, N_p$

(i) Draw $\mathbf{s}_0^{(j)} \sim p(\mathbf{s}_0)$, and $z_0^{(j)} \sim P(z_0)$.

(ii) Make $w_0^{(j)} = 1/N_p$ and $n = 1$.

(2) *Importance Sampling* For $j = 1, \dots, N_p$

(i) Draw $\hat{z}_n^{(j)} \sim P(z_n | z_{n-1}^{(j)}, \mathbf{s}_{n-1}^{(j)})$ according to (2).

(ii) Draw $\hat{\mathbf{s}}_n^{(j)} \sim p(\mathbf{s}_n | \hat{z}_n^{(j)}, \mathbf{s}_{n-1}^{(j)}, z_{n-1}^{(j)})$ according to (4) or (5).

(iii) Update the importance weights

$$\hat{w}_n^{(j)} \propto w_{n-1}^{(j)} p(\mathbf{y}_n | \hat{\mathbf{s}}_n^{(j)}, \hat{z}_n^{(j)}) \quad (18)$$

using the likelihood function in Section 2.2.

End-For

(i) Normalize the weights $\{\hat{w}_n^{(j)}\}$ such that $\sum_{j=1}^{N_p} \hat{w}_n^{(j)} = 1$.

(ii) For $j = 1, \dots, N_p$, make $\mathbf{s}_n^{(j)} = \hat{\mathbf{s}}_n^{(j)}$, $z_n^{(j)} = \hat{z}_n^{(j)}$, and $w_n^{(j)} = \hat{w}_n^{(j)}$.

(iii) Make $n = n + 1$ and go back to step 2.

3.2. Resample-move filter

The sequential importance sampling algorithm in Section 3.1 is guaranteed to converge asymptotically with probability one; see [27]. However, due to the increase in the variance of the importance weights, the raw SIS algorithm suffers from the ‘‘particle degeneracy’’ phenomenon [8, 14, 17]; that is, 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. As a result of particle degeneracy, the SIS algorithm is inefficient, requiring the use of a large number of particles to achieve adequate performance.

Resampling step

A possible approach to mitigate degeneracy is [17] to resample from the existing particle population with replacement according to the particle weights. Formally, after the normalization of importance weights $\{\hat{w}_n^{(j)}\}$, we draw $i^{(j)} \sim \{1, 2, \dots, N_p\}$ with $\Pr(\{i^{(j)} = l\}) = \hat{w}_n^{(l)}$, and build a new particle set $\{\mathbf{s}_n^{(j)}, z_n^{(j)}\}$, $j = 1, \dots, N_p$, such that $(\mathbf{s}_n^{(j)}, z_n^{(j)}) = (\hat{\mathbf{s}}_n^{(i^{(j)})}, \hat{z}_n^{(i^{(j)})})$. After the resampling step, the new selected trajectories $(\mathbf{s}_{0:n}^{(j)}, z_{0:n}^{(j)}) = (\mathbf{s}_{0:n-1}^{(i^{(j)})}, \hat{\mathbf{s}}_n^{(i^{(j)})}, z_{0:n-1}^{(i^{(j)})}, \hat{z}_n^{(i^{(j)})})$ are approximately distributed (see, e.g., [28]) according to the mixed

posterior pdf $p(\mathbf{s}_{0:n}, z_{0:n} | \mathbf{y}_{1:n})$ so that we can reset all particle weights to $1/N_p$.

Move step

Although particle resampling according to the weights reduces particle degeneracy, it also introduces an undesirable side effect, namely, loss of diversity in the particle population as the resampling processes generate multiple copies of a small number or, in the extreme case, only one high-weight particle. A possible solution, see [16], to restore sample diversity without altering the sample statistics is to move the current particles $\{\mathbf{s}_n^{(j)}, z_n^{(j)}\}$ to new locations $\{\tilde{\mathbf{s}}_n^{(j)}, \tilde{z}_n^{(j)}\}$ using a Markov chain transition kernel $k(\mathbf{s}_n^{(j)}, z_n^{(j)} | \tilde{\mathbf{s}}_n^{(j)}, \tilde{z}_n^{(j)})$, that is, invariant to the conditional mixture pdf $p(\mathbf{s}_n, z_n | \mathbf{s}_{0:n-1}^{(j)}, z_{0:n-1}^{(j)}, \mathbf{y}_{1:n})$. Provided that the invariance condition is satisfied, the new particle trajectories $(\tilde{\mathbf{s}}_{0:n}^{(j)}, \tilde{z}_{0:n}^{(j)}) = (\mathbf{s}_{0:n-1}^{(j)}, z_{0:n-1}^{(j)}, \tilde{\mathbf{s}}_n^{(j)}, \tilde{z}_n^{(j)})$ remain distributed according to $p(\mathbf{s}_{0:n}, z_{0:n} | \mathbf{y}_{1:n})$ and the associated particle weights may be kept equal to $1/N_p$. A Markov chain that satisfies the desired invariance condition can be built using the following Metropolis-Hastings strategy [15, 18].

For $j = 1, \dots, N_p$, the following algorithm holds.

- (i) Draw $\tilde{z}_n^{(j)} \sim P(z_n | z_{n-1}^{(j)}, \mathbf{s}_{n-1}^{(j)})$ according to (2).
- (ii) Draw $\tilde{\mathbf{s}}_n^{(j)} \sim p(\mathbf{s}_n | \tilde{z}_n^{(j)}, \mathbf{s}_{n-1}^{(j)}, z_{n-1}^{(j)})$ according to (4) or (5).
- (iii) Draw $u \sim \mathcal{U}([0, 1])$.
If

$$u \leq \min \left\{ 1, \frac{p(\mathbf{y}_n | \tilde{\mathbf{s}}_n^{(j)}, \tilde{z}_n^{(j)})}{p(\mathbf{y}_n | \mathbf{s}_n^{(j)}, z_n^{(j)})} \right\}, \quad (19)$$

then

$$(\mathbf{s}_n^{(j)}, z_n^{(j)}) = (\tilde{\mathbf{s}}_n^{(j)}, \tilde{z}_n^{(j)}). \quad (20)$$

Else,

$$(\mathbf{s}_n^{(j)}, z_n^{(j)}) = (\mathbf{s}_n^{(j)}, z_n^{(j)}). \quad (21)$$

- (iv) Reset $w_n^{(j)} = 1/N_p$.

End-For.

3.3. Auxiliary particle filter

An alternative to the resample-move filter in Section 3.2 is to use the current observation \mathbf{y}_n to preselect at instant $n-1$ a set of particles that, when propagated to instant n according to the system dynamics, is more likely to generate samples with high likelihood. That can be done using an auxiliary particle filter (APF) [19] which samples in two steps from a mixed importance function:

$$q(i, \mathbf{s}_n, z_n | \mathbf{y}_{1:n}) \propto w_{n-1}^{(i)} p(\mathbf{y}_n | \hat{\mathbf{s}}_n^{(i)}, \hat{z}_n^{(i)}) p(\mathbf{s}_n, z_n | \mathbf{s}_{n-1}^{(i)}, z_{n-1}^{(i)}), \quad (22)$$

where $\hat{z}_n^{(j)}$ and $\hat{\mathbf{s}}_n^{(j)}$ are drawn according to the mixed prior $p(\mathbf{s}_n, z_n | \mathbf{s}_{n-1}^{(j)}, z_{n-1}^{(j)})$. The proposed algorithm is summarized into the following steps.

(1) *Pre-sampling Selection Step* For $j = 1, \dots, N_p$

- (i) Draw $\hat{z}_n^{(j)} \sim P(z_n | z_{n-1}^{(j)}, \mathbf{s}_{n-1}^{(j)})$ according to (2).
- (ii) Draw $\hat{\mathbf{s}}_n^{(j)} \sim p(\mathbf{s}_n | \hat{z}_n^{(j)}, \mathbf{s}_{n-1}^{(j)}, z_{n-1}^{(j)})$ according to (4) or (5).
- (iii) Compute the first-stage importance weights

$$\lambda_n^{(j)} \propto w_{n-1}^{(j)} p(\mathbf{y}_n | \hat{\mathbf{s}}_n^{(j)}, \hat{z}_n^{(j)}), \quad \sum_{j=1}^{N_p} \lambda_n^{(j)} = 1, \quad (23)$$

using the likelihood function model in Section 2.2.

End-For

(2) *Importance Sampling with Auxiliary Particles* For $j = 1, \dots, N_p$

- (i) Sample $i^{(j)} \sim \{1, \dots, N_p\}$ with $\Pr(\{i^{(j)} = l\}) = \lambda_n^{(l)}$.
- (ii) Sample $\tilde{z}_n^{(j)} \sim P(z_n | z_{n-1}^{(i^{(j)})}, \mathbf{s}_{n-1}^{(i^{(j)})})$ according to (2).
- (iii) Sample $\tilde{\mathbf{s}}_n^{(j)} \sim p(\mathbf{s}_n | \tilde{z}_n^{(j)}, \mathbf{s}_{n-1}^{(i^{(j)})}, z_{n-1}^{(i^{(j)})})$ according to (4) or (5).
- (iv) Compute the second-stage importance weights

$$\tilde{w}_n^{(j)} \propto \frac{p(\mathbf{y}_n | \tilde{\mathbf{s}}_n^{(j)}, \tilde{z}_n^{(j)})}{p(\mathbf{y}_n | \hat{\mathbf{s}}_n^{(i^{(j)})}, \hat{z}_n^{(i^{(j)})})}. \quad (24)$$

End-For.

(v) Normalize the weights $\{\tilde{w}_n^{(j)}\}$ such that $\sum_{j=1}^{N_p} \tilde{w}_n^{(j)} = 1$.

(3) *Post-sampling Selection Step* For $j = 1, \dots, N_p$

- (i) Draw $k^{(j)} \sim \{1, \dots, N_p\}$ with $\Pr(\{k^{(j)} = l\}) = \tilde{w}_n^{(l)}$.
- (ii) Make $\mathbf{s}_n^{(j)} = \tilde{\mathbf{s}}_n^{(k^{(j)})}$, $z_n^{(j)} = \tilde{z}_n^{(k^{(j)})}$ and $w_n^{(j)} = 1/N_p$.

End-For.

(iii) Make $n = n + 1$ and go back to step 1.

3.4. Multiframe detector/tracker

The final result at instant n of either the RS algorithm in Section 3.2 or the APF algorithm in Section 3.3 is a set of equally weighted samples $\{\mathbf{s}_n^{(j)}, z_n^{(j)}\}$ that are approximately distributed according to the mixed posterior $p(\mathbf{s}_n, z_n | \mathbf{y}_{1:n})$. Next, let H_1 denote the hypothesis that the target of interest is present in the scene at frame n . Conversely, let H_0 denote the hypothesis that the target is absent. Given the equally weighted set $\{\mathbf{s}_n^{(j)}, z_n^{(j)}\}$, we compute then the *Monte Carlo estimate*, $\widehat{\Pr}(\{z_n = 0\} | \mathbf{y}_{1:n})$, of the posterior probability of target absence by dividing the number of particles for which $z_n^{(j)} = 0$ by the total number of particles N_p . The minimum probability of error test to decide between hypotheses H_1 and H_0 at frame n is approximated then by the decision rule

$$\widehat{\Pr}(\{z_n = 0\} | \mathbf{y}_{1:n}) \underset{H_1}{\overset{H_0}{\geq}} 1 - \widehat{\Pr}(\{z_n = 0\} | \mathbf{y}_{1:n}) \quad (25)$$

or, equivalently,

$$\widehat{\Pr}(\{z_n = 0\} \mid \mathbf{y}_{1:n}) \underset{H_1}{\overset{H_0}{\geq}} \frac{1}{2}. \quad (26)$$

Finally, if H_1 is accepted, the estimate $\widehat{\mathbf{s}}_{n|n}$ of the target’s kinematic state at instant n is obtained from the Monte Carlo approximation of $E[\mathbf{s}_n \mid \mathbf{y}_{1:n}, \{z_n \neq 0\}]$, which is computed by averaging out the particles $\mathbf{s}_n^{(j)}$ such that $z_n^{(j)} \neq 0$.

4. SIMULATION RESULTS

In this section, we quantify the performance of the proposed sequential Monte Carlo detector/tracker, both in the RS and APF configurations, using simulated infrared airborne radar (IRAR) data. The background clutter is simulated from real IRAR images from the MIT Lincoln Laboratory database, available at the CIS website, at Johns Hopkins University. An artificial target template representing a military vehicle is added to the simulated image sequence. The simulated target’s centroid moves in the image from frame to frame according to the simple white-noise acceleration model in [3, 4] with parameters $q = 6$ and $T = 0.04$ second. A total of four rotated, scaled, or sheared versions of the reference template is used in the simulation.

The target’s aspect changes from frame to frame following a known discrete-valued hidden Markov chain model where the probability of a transition to an adjacent aspect state is equal to 40%. In the notation of Section 2.1, that specification corresponds to setting $G(1, 1) = G(4, 4) = 0.6$, $G(2, 2) = G(3, 3) = 0.2$, $G(i, j) = 0.4$ if $|i - j| = 1$, and $G(i, j) = 0$ otherwise. All four templates are equally likely at frame zero, that is, $P(z_0) = 1/4$ for $z_0 = 1, 2, 3, 4$. The initial x and y positions of the target’s centroid at instant zero are assumed to be uniformly distributed, respectively, between pixels 50 and 70 in the x coordinate and pixels 10 and 20 in the y coordinate. The initial velocities v_x and v_y are in turn Gaussian-distributed with identical means (10 m/s or 2 pixels/frame) and a small standard deviation ($\sigma = 0.1$).

Finally, 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 synthesized using correlation and prediction error variance parameters estimated from real data using the algorithms developed in [11, 12] see [4] for a detailed pseudocode.

Two video demonstrations of the operation of the proposed detector/tracker are available for visualization by clicking on the links in [29]. The first video (peak target-to-clutter ratio, or PTCR ≈ 10 dB) illustrates the performance over 50 frames of an 8 000-particle RS detector/tracker implemented as in Section 3.2, whereas the second video (PTCR ≈ 6.5 dB) demonstrates the operation over 60 frames of a 5 000-particle APF detector/tracker implemented as in Section 3.3. Both video sequences show a target of interest that is tracked inside the image grid until it disappears from the scene; the algorithm then detects that the target is absent and correctly indicates that no target is present. Next, once a new target en-

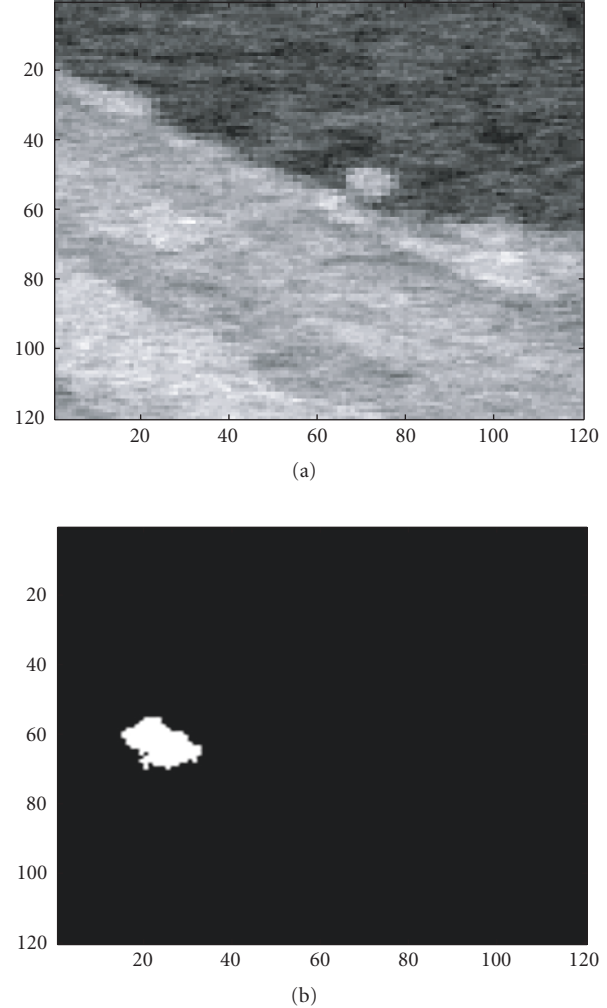


FIGURE 1: (a) First frame of the cluttered target sequence, PTCR = 10.6 dB; (b) target template and position in the first frame shown as a binary image.

ters the scene, that target is acquired and tracked accurately until, in the case of the APF demonstration, it leaves the scene and no target detection is once again correctly indicated.

Both video demos show the ability of the proposed algorithms to (1) detect and track a present target both inside the image grid and near its borders, (2) detect when a target leaves the image and indicate that there is no target present until a new target appears and (3), when a new target enters the scene, correctly detect that the target is present and track it accurately. In the sequel, for illustrative purposes only, we show in the paper the detection/tracking results for a few selected frames using the RS algorithm and a dataset that is different from the one shown in the video demos.

Figure 1(a) shows the initial frame of the sequence with the target centered in the (quantized) coordinates (65, 23) and superimposed on clutter. The clutter-free target template, centered at the same pixel location, is shown as a binary image in Figure 1(b). The simulated PTCR in Figure 1(b) is 10.6 dB.

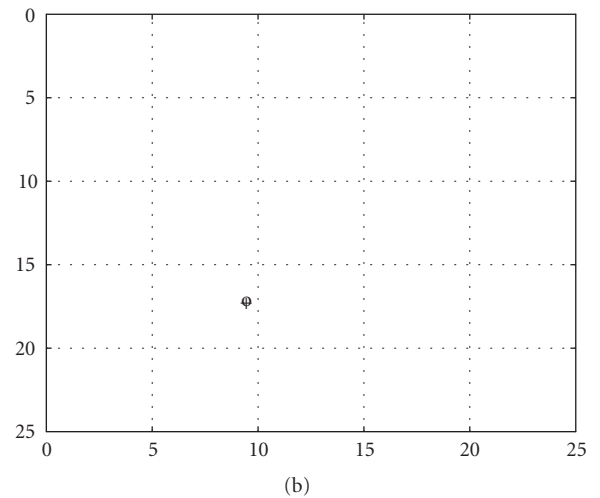
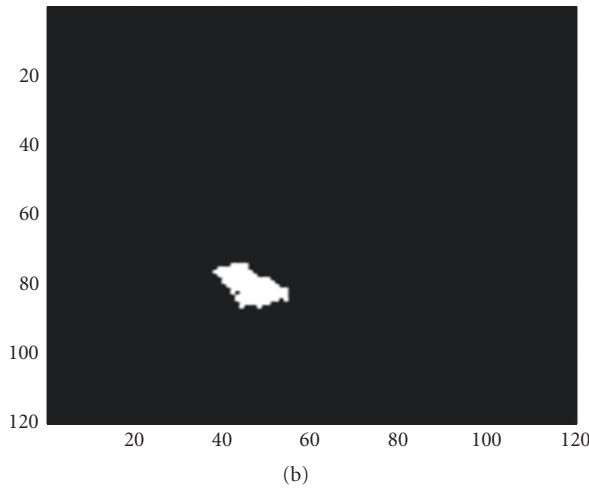
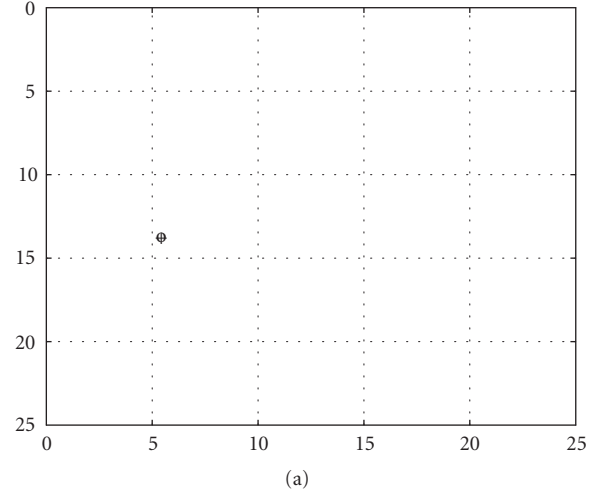
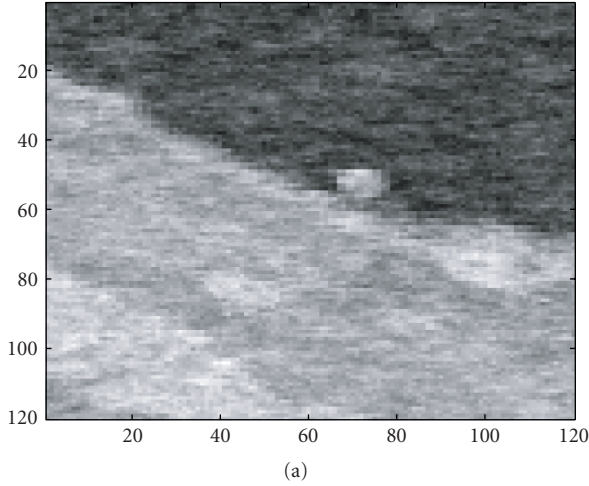


FIGURE 2: (a) Tenth frame of the cluttered target sequence, PTCR = 10.6 dB, with target translation, rotation, scaling, and shearing; (b) target template and position in the tenth frame shown as a binary image.

FIGURE 3: Tracking results: actual target position (+), estimated target position (o); (a) initial frame, (b) tenth frame.

Next, Figure 2(a) shows the tenth frame in the image sequence. Once again, we show in Figure 2(b) the corresponding clutter-free target image as a binary image. 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 frames 1 and 10 are shown, respectively, in Figures 3(a) and 3(b). The actual target positions are indicated by a cross sign ('+'), while the estimated positions are indicated by a circle ('o'). Note that the axes in Figures 1(a) and 1(b) and Figures 2(a) and 2(b) represent integer pixel locations, while the axes in Figures 3(a) and 3(b) represent real-valued x and y , coordinates assuming spatial resolutions of $\xi_1 = \xi_2 = 0.2$ meters/pixel such that the $[0, 120]$ pixel range in the axes of Figures 1(a) and 1(b) and Figures 2(a) and 2(b) corresponds to a $[0, 24]$ meter range in the axes of Figures 3(a) and 3(b).

In this particular example, 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 4(a) and 4(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 reappears, 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 5(a). Figure 5(b) shows the corresponding tracking results for the same frame.

In order to obtain a quantitative assessment of tracking performance, we ran 100 independent Monte Carlo simulations using, respectively, the 5000-particle APF detector/tracker and the 8000-particle RS detector/tracker. Both algorithms correctly detected the presence of the target over a sequence of 20 simulated frames in all 100 Monte Carlo runs. However, with PTCR = 6.5 dB, the 5000-particle APF tracker

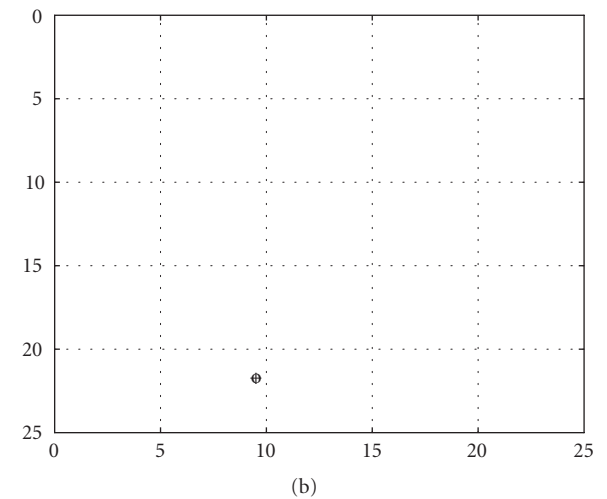
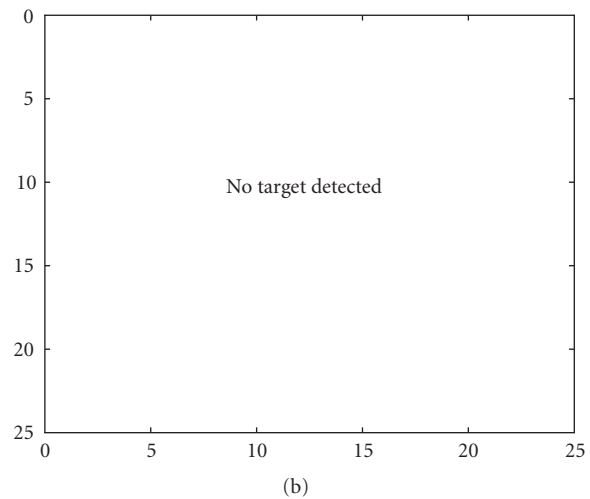
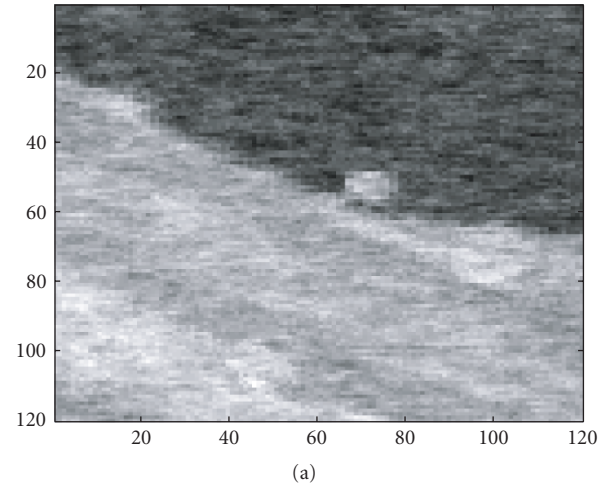
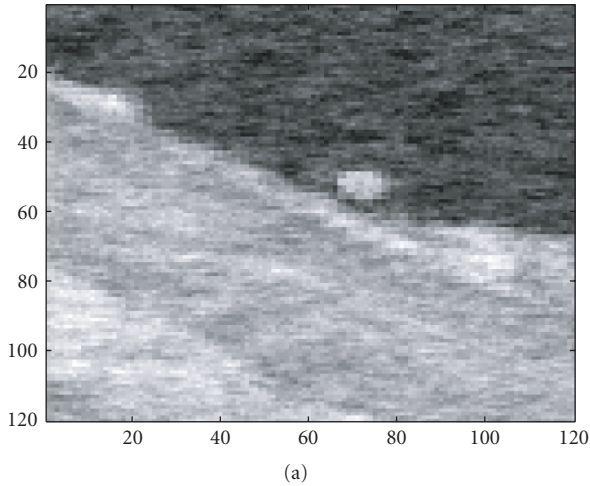


FIGURE 4: (a) Thirty-sixth frame of the cluttered target sequence with no target present; (b) detection result indicating absence of target.

FIGURE 5: (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).

diverged (i.e., failed to estimate the correct target trajectory) in 3 out of the 100 Monte Carlo trials, whereas the RS tracker diverged in 5 out of 100 runs. When we increased the PTCR to 8.1 dB, the divergence rates fell to 2 out of 100 for the APF, and 3 out of 100 for the RS filter. Figures 6(a) and 6(b) show, in the case of PTCR = 6.5 dB, the root mean square (RMS) error curves (in number of pixels) for the target’s position estimates, respectively, in coordinates x and y generated by both the APF and the RS trackers. The RMS error curves in Figure 6 were computed from the estimation errors recorded in each of the 100 Monte Carlo trials, *excluding* the divergent realizations. Our simulation results suggest that, despite the reduction in the number of particles from 8000 to 5000, the APF tracker still outperforms the RS tracker, showing similar RMS error performance with a slightly lower divergence rate. For both filters, in the nondivergent realizations, the estimation error is higher in the initial frames and decreases over time as the target is acquired and new images are processed.

5. PRELIMINARY DISCUSSION ON MULTITARGET TRACKING

We have considered so far a single target with uncertain aspect (e.g., random orientation or scale). In theory, however, the same modeling framework could be adapted to a scenario where we consider multiple targets with known (fixed) aspect. In that case, the discrete state z_n , rather than representing a possible target model, could denote instead a possible multitarget configuration hypothesis. For example, if we knew a priori that there is a maximum of N_T targets in the field of view of the sensor at each time instant, then z_n would take $K = 2^{N_T}$ possible values corresponding to the hypotheses ranging from “no target present” to “all targets present” in the image frame at instant n . The kinematic state s_n , on the other hand, would have variable dimension depending on the value assumed by z_n , as it would collect the centroid locations of all targets that are present in the image

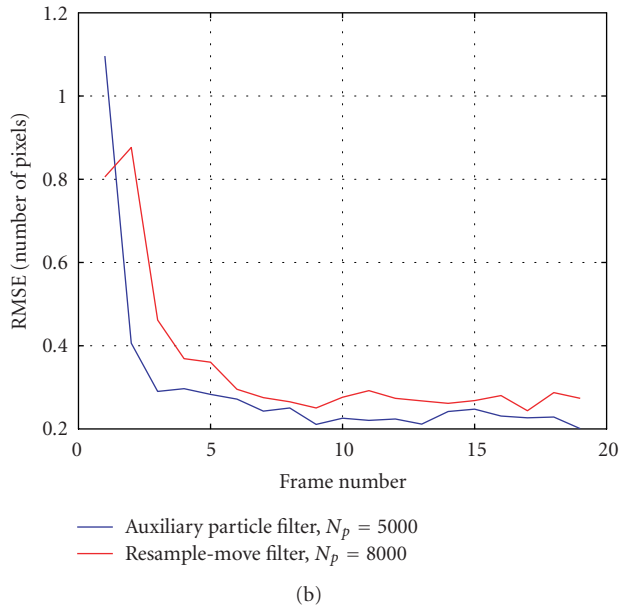
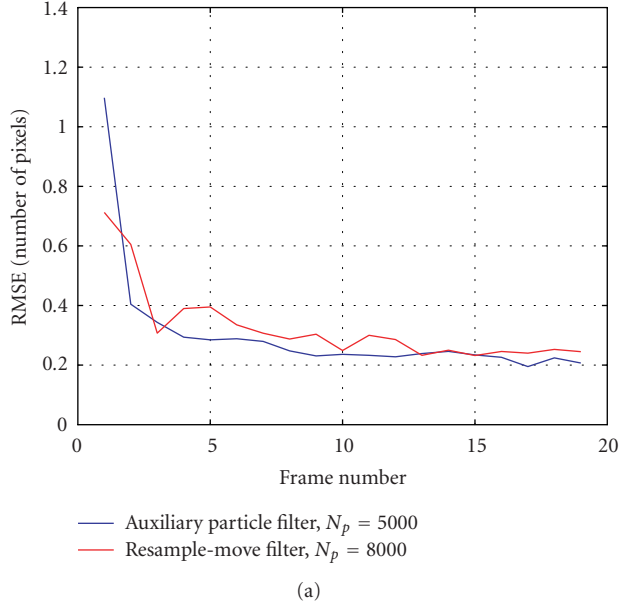


FIGURE 6: RMS error for the target's position estimate, respectively, for the APF (divergence rate, 3%) and resample-move (divergence rate, 5%) trackers, PTCR = 6.5 d; (a) x coordinate, (b) y coordinate.

given a certain target configuration hypothesis. Different targets could be assumed to move independently of each other when present and to disappear only when they move out of the target grid as discussed in Section 2. Likewise, a change in target configuration hypotheses would result in new targets appearing in uniformly random locations as in (5).

The main difficulty associated with the approach described in the previous paragraph is however that, as the number of targets increases, the corresponding growth in the dimension of the state space is likely to exacerbate particle depletion, thus causing the detection/tracking filters to di-

verge if the number of particles is kept constant. That may render the direct application of the joint detection/tracking algorithms in this paper unfeasible in a multitarget scenario. The basic tracking routines discussed in the paper may be still viable though when used in conjunction with more conventional algorithms for target detection/acquisition and data association. For a review of alternative approaches to multitarget tracking, mostly for video applications, we refer the reader to [30–33].

5.1. Likelihood function modification in a multitarget scenario

In the alternative scenario with multiple (at most N_T) targets, where z_n represents one of 2^{N_T} possible target configurations, the likelihood function model in (10) depends instead on a sum of data terms

$$\lambda_{n,i}(\mathbf{s}_n, z_n) = \mathbf{y}_n^T (\sigma_{c,n}^2 \boldsymbol{\Sigma}_v^{-1}) \mathbf{h}_i(\mathbf{s}_n, z_n), \quad 1 \leq i \leq 2^{N_T}, \quad (27)$$

and a sum of energy terms

$$\rho_{i,j}(\mathbf{s}_n, z_n) = \mathbf{h}_i^T(\mathbf{s}_n, z_n) (\sigma_{c,n}^2 \boldsymbol{\Sigma}_v^{-1}) \mathbf{h}_j(\mathbf{s}_n, z_n), \quad 1 \leq i, j \leq 2^{N_T}, \quad (28)$$

where $\mathbf{h}_i(\mathbf{s}_n, z_n)$ is the long-vector representation of the clutter-free image of the i th target under the target configuration hypothesis z_n , assumed to be identically zero for target configurations under which the i th target is not present. The sum of the data terms corresponds to the sum of the outputs of different correlation filters matched to each of the N_T possible (fixed) target templates taking into account the spatial correlation of the clutter background. The energy terms, $\rho_{i,j}(\mathbf{s}_n, z_n)$, are on the other hand constant with \mathbf{s}_n for most possible locations of targets i and j on the image grid, except when either one of the two targets or both are close to the image borders. Finally, for $i \neq j$, the energy terms are zero for present targets that are sufficiently apart from each other and, therefore, most of the time, they do not affect the computation of the likelihood function. The terms $\rho_{i,j}(\mathbf{s}_n, z_n)$ must be taken into account, however, for overlapping targets; in this case, they may be computed efficiently exploring the sparse structure of \mathbf{h}_i and $\boldsymbol{\Sigma}_v^{-1}$. For details, we refer the reader to future work.

5.2. Illustrative example with two targets

We conclude this preliminary discussion on multitarget tracking with an illustrative example where we track two simulated targets moving on the same real clutter background from Section 4 for 22 consecutive frames. This example differs, however, from the simulations in Section 4 in the sense that, rather than performing joint detection and tracking of the two targets, the algorithm assumes a priori that two targets are always present in the scene and performs target tracking only. The two targets are preacquired (detected) in the initial frame such that their initial positions are known up only to a small uncertainty. For this particular simulation, with PTCR ≈ 12.5 dB, that preliminary acquisition was

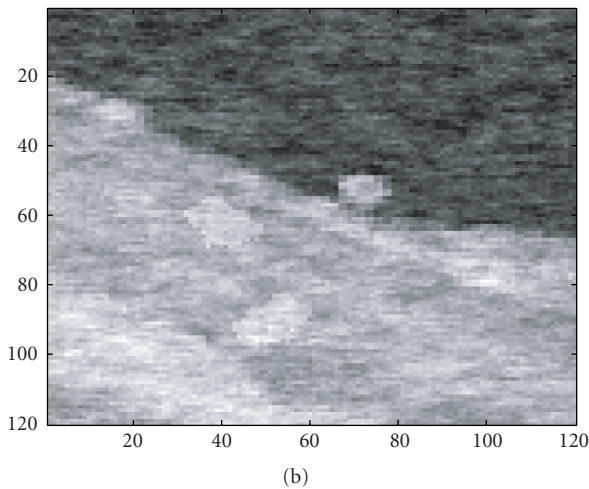
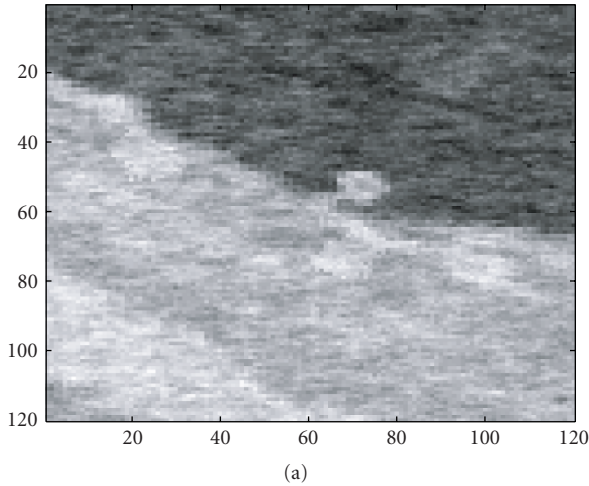


FIGURE 7: Simulated image sequence with two moving targets: (a) first frame, (b) tenth frame; PTCR \approx 12.5 dB.

done by applying the differential filter in (16) to the initial frame, and then applying the output of the differential filter to a bank of two spatial matched filters as in (15), designed according to the signature coefficients, respectively, for targets 1 and 2. The outputs of the two matched filters minus the corresponding energy terms for targets 1 and 2, respectively, are finally added together and thresholded to provide the initial estimates of the location of the two targets. Note that the cross-energy terms discussed in Section 5.1 may be ignored in this case since we are assuming that the two targets are initially sufficiently far apart. Frames 1 and 10 of the simulated cluttered sequence with the two targets are shown in Figures 7(a) and 7(b) for illustration purposes.

Once the two targets are initially acquired, we track them jointly from the raw sensor images (i.e., without any other conventional decoupled data association/tracking method) using an auxiliary particle filter that assumes the modified likelihood function of Section 5.1 and two independent motion models, respectively, for targets 1 and 2. The tracking filter uses $N_p = 8000$ particles. Figure 8 shows the actual

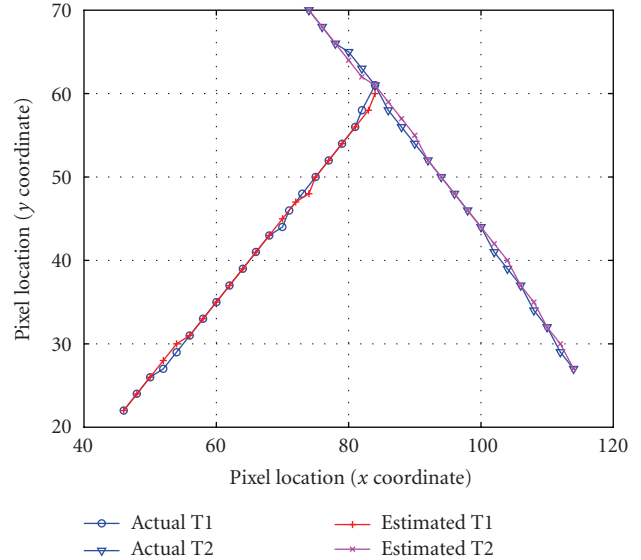


FIGURE 8: Illustrative example of two-target tracking using an auxiliary particle filter.

and estimated trajectories, respectively, for targets 1 and 2. As we can see from the plots, the experimental results are worse than those obtained in the single target, multispect case, but the filter was still capable of tracking the pixel locations of the centroids of the two targets fairly accurately with errors ranging from zero to two or three pixels at most and without using any ad hoc data association routine.

6. CONCLUSIONS AND FUTURE WORK

We discussed in this paper a methodology for joint detection and tracking of multispect targets in remote sensing image sequences using sequential Monte Carlo (SMC) filters. The proposed algorithm enables integrated, *multiframe* target detection and tracking incorporating the statistical models for target motion, target aspect, and spatial correlation of the background clutter. Due to the nature of the application, the emphasis is on detecting and tracking small, remote targets under additive clutter, as opposed to tracking nearby objects possibly subject to occlusion.

Two different implementations of the SMC detector/tracker were presented using, respectively, a resample-move (RS) particle filter and an auxiliary particle filter (APF). Simulation results show that, in scenarios with heavily obscured targets, the APF and RS configurations have similar tracking performance, but the APF algorithm has a slightly smaller percentage of divergent realizations. Both filters, on the other hand, were capable of correctly detecting the target in each frame, including accurately declaring absence of target when the target left the scene and, conversely, detecting a new target when it entered the image grid. The multiframe track-before-detect approach allowed for efficient detection of dim targets that may be near invisible in a single-frame but become detectable when seen across multiple frames.

The discussion in this paper was restricted to targets that assume only a finite number of possible aspect states defined on a library of target templates. As an alternative for future work, an appearance model similar to the one described in [24] could be used instead, allowing the discrete-valued aspect states s_n to denote different classes of continuous-valued target deformation models, as opposed to fixed target templates. Similarly, the framework in this paper could also be modified to allow for multiobject tracking as indicated in Section 5.

ACKNOWLEDGMENT

Part of the material in this paper was presented at the 2005 IEEE Aerospace Conference.

REFERENCES

- [1] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [2] J. K. Bounds, "The Infrared airborne radar sensor suite," RLE Tech. Rep. 610, Massachusetts Institute of Technology, Cambridge, Mass, USA, December 1996.
- [3] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, Conn, USA, 1995.
- [4] M. G. S. Bruno, "Bayesian methods for multiaspect target tracking in image sequences," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1848–1861, 2004.
- [5] M. G. S. Bruno and J. M. F. Moura, "Optimal multiframe detection and tracking in digital image sequences," in *Proceedings of the IEEE International Acoustics, Speech, and Signal Processing (ICASSP '00)*, vol. 5, pp. 3192–3195, Istanbul, Turkey, June 2000.
- [6] M. G. S. Bruno and J. M. F. Moura, "Multiframe detector/tracker: optimal performance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 3, pp. 925–945, 2001.
- [7] M. G. S. Bruno and J. M. F. Moura, "Multiframe Bayesian tracking of cluttered targets with random motion," in *Proceedings of the International Conference on Image Processing (ICIP '00)*, vol. 3, pp. 90–93, Vancouver, BC, Canada, September 2000.
- [8] 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, no. 2, pp. 174–188, 2002.
- [9] J. M. F. Moura and N. Balram, "Recursive structure of non-causal Gauss-Markov random fields," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 334–354, 1992.
- [10] 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, no. 9, pp. 2571–2574, 1998.
- [11] J. M. F. Moura and N. Balram, "Noncausal Gauss Markov random fields: parameter structure and estimation," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1333–1355, 1993.
- [12] S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: clutter adaptation in anomaly detection," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1855–1871, 2000.
- [13] M. Isard and A. Blake, "A mixed-state condensation tracker with automatic model-switching," in *Proceedings of the 6th International Conference on Computer Vision*, pp. 107–112, Bombay, India, January 1998.
- [14] 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, N. F. G. Freitas, and N. J. Gordon, Eds., Springer, New York, NY, USA, 2001.
- [15] M. G. S. Bruno, R. V. de Araújo, and A. G. Pavlov, "Sequential Monte Carlo filtering for multi-aspect detection/tracking," in *Proceedings of the IEEE Aerospace Conference*, pp. 2092–2100, Big Sky, Mont, USA, March 2005.
- [16] W. R. Gilks and C. Berzuini, "Following a moving target—Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 127–146, 2001.
- [17] N. Gordon, D. Salmond, and C. Ewing, "Bayesian state estimation for tracking and guidance using the bootstrap filter," *Journal of Guidance, Control, and Dynamics*, vol. 18, no. 6, pp. 1434–1443, 1995.
- [18] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer, New York, NY, USA, 1999.
- [19] M. K. Pitt and N. Shephard, "Filtering via simulation: auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.
- [20] M. Isard and A. Blake, "Condensation—conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.
- [21] B. Li and R. Chellappa, "A generic approach to simultaneous tracking and verification in video," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 530–544, 2002.
- [22] S. K. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1491–1506, 2004.
- [23] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 39, no. 1, pp. 1–38, 1977.
- [24] J. Giebel, D. M. Gavrila, and C. Schnörr, "A bayesian framework for multi-cue 3D object tracking," in *Proceedings of the 8th European Conference on Computer Vision (ECCV '04)*, vol. 4, pp. 241–252, Prague, Czech Republic, May 2004.
- [25] J. Vermaak, A. Doucet, and P. Pérez, "Maintaining multimodality through mixture tracking," in *Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV '03)*, vol. 2, pp. 1110–1116, Nice, France, October 2003.
- [26] K. Okuma, A. Taleghani, N. de Freitas, J. J. Little, and D. G. Lowe, "A boosted particle filter: multitarget detection and tracking," in *Proceedings of the 8th European Conference on Computer Vision (ECCV '04)*, vol. 3021, pp. 28–39, Prague, Czech Republic, May 2004.
- [27] J. Geweke, "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, vol. 57, no. 6, pp. 1317–1339, 1989.
- [28] J. S. Liu, R. Chen, and T. Logvinenko, "A theoretical framework for sequential importance sampling with resampling," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, J. F. G. Freitas, and N. J. Gordon, Eds., pp. 225–246, Springer, New York, NY, USA, 2001.
- [29] Video Demonstration 1 & 2, <http://www.ele.ita.br/~bruno>.

-
- [30] W. Ng, J. Li, S. Godsill, and J. Vermaak, "Tracking variable number of targets using sequential Monte Carlo methods," in *Proceedings of the IEEE/SP 13th Workshop on Statistical Signal Processing*, pp. 1286–1291, Bordeaux, France, July 2005.
 - [31] W. Ng, J. Li, S. Godsill, and J. Vermaak, "Multitarget tracking using a new soft-gating approach and sequential Monte Carlo methods," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 4, pp. 1049–1052, Philadelphia, Pa, USA, March 2005.
 - [32] C. Hue, J.-P. Le Cadre, and P. Pérez, "Tracking multiple objects with particle filtering," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 791–812, 2002.
 - [33] J. Czyz, B. Ristic, and B. Macq, "A color-based particle filter for joint detection and tracking of multiple objects," in *Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, pp. 217–220, Philadelphia, Pa, USA, March 2005.