

# Bayesian Methods for Multiaspect Target Tracking in Image Sequences

Marcelo G. S. Bruno, *Member, IEEE*

**Abstract**—In this paper, we introduce new algorithms for automatic tracking of multiaspect targets in cluttered image sequences. We depart from the conventional correlation filter/Kalman filter association approach to target tracking and propose instead a nonlinear Bayesian methodology that enables direct tracking from the image sequence incorporating the statistical models for the background clutter, target motion, and target aspect change. Proposed algorithms include 1) a batch hidden Markov model (HMM) smoother and a sequential HMM filter for joint multiframe target detection and tracking and 2) two mixed-state sequential importance sampling trackers based on the sampling/importance resampling (SIR) and the auxiliary particle filtering (APF) techniques. Performance studies show that the proposed algorithms outperform the association of a bank of template correlators and a Kalman filter in adverse scenarios of low target-to-clutter ratio and uncertainty in the true target aspect.

**Index Terms**—Bayesian estimation, hidden Markov models, multiaspect target tracking, noncausal Gauss–Markov random fields, particle filters.

## I. INTRODUCTION

REMOTE imaging sensors scan a certain surveillance region recording returns from both targets of interest and spurious scatterers in the background. The raw sensor measurements are sampled and processed to form a sequence of two-dimensional (2-D) cluttered images. In typical scenarios of practical interest, designated targets are heavily obscured by structured clutter, becoming barely visible to a human observer. In addition, the aspect of the clutter-free image of a target of interest may itself change randomly over time as a result of rotational motion and/or variations in the conditions of observation of the target. Finally, targets of interest may be moving randomly in the background, adding further uncertainty on their true spatial location.

Previous literature on multiaspect target detection, e.g., [1]–[3], is concerned mostly with stationary targets and focuses on designing correlation filters that are robust to distortions of the target's template. The literature on moving target tracking with imaging sensors (see, e.g., [4] and [5]) is in turn based on a suboptimal decoupling of the detection and tracking tasks. Typically, a preliminary single frame detection stage uses image processing and/or pattern recognition techniques such as

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 with 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). This suboptimal association of correlation detectors and KBF trackers has been shown to perform poorly for low target-to-clutter ratios, even in simpler test scenarios where the target aspect and the clutter statistics are known; see [6].

We propose to overcome the limitations of the conventional correlation filter/KBF association using a Bayesian methodology that enables direct target tracking from the image sequence, eliminating any preliminary single frame image correlation or pattern recognition step and integrating detection and tracking. We design recursive, multiframe algorithms that fully incorporate the (stochastic) dynamic models for target motion and aspect and take full advantage of the knowledge of the statistical model for the spatially correlated clutter background. The proposed algorithms are *clutter adaptive*, learning the parameters of the clutter correlation model from the available test data with no need for training.

Previous work [7]–[10] considered the problem of classification/identification of stationary, multiaspect targets using hidden Markov models (HMMs) [11] to represent the aspect-dependent electromagnetic [7] or acoustical [9], [10] scattering characteristics of the targets. We use a similar strategy to handle the uncertainty in the target aspect. However, instead of processing the target's scattered waveforms, we have as input data a preprocessed sequence of digital images in which we use the HMM formalism to model the rotations and/or deformations (scaling or shearing) of the target template from frame to frame. The target's aspect state is incorporated into our model as a dynamic unknown variable that is jointly estimated with the target's kinematic state. To model the evolution in time of the kinematic state, we consider, on the other hand, two strategies, using, respectively, discrete-valued and continuous-valued state variables.

*Grid-Based Detector/Trackers:* In the discrete-space approach, we take advantage of the sensor's finite resolution to model the target's centroid motion as a 2-D discrete-valued HMM defined directly on the sensor's image grid. Detection and tracking are then easily integrated by adding a dummy absent target state to the motion HMM. Using the discrete-valued models for target motion and aspect, we derive a batch HMM smoother and an online HMM filter for joint multiframe, multiaspect detection and tracking. The proposed HMM smoother

Manuscript received August 28, 2002; revised August 28, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Arnab K. Shaw.

The author is with the Divisão de Engenharia Eletrônica, Instituto Tecnológico de Aeronáutica, São José dos Campos SP 12228-900, Brazil (e-mail: bruno@ele.ita.br).

Digital Object Identifier 10.1109/TSP.2004.828903

is a generalization of the forward-backward Baum–Welch recursions [12], [13] and provides an alternative to Viterbi-based trackers; see, e.g., [14]. The online HMM filter in this paper generalizes, in turn to multiaspect targets, the 2-D sequential HMM detector/tracker briefly introduced in [6].

*Particle Filter Trackers:* Despite their good performance, the grid-based solutions described in the previous paragraph suffer from two shortcomings: 1) They are computationally intensive, and 2) a discrete-valued model for target motion may be too simplistic to represent real-world target dynamics. To circumvent these problems, we also introduce, in this paper, a second class of multiaspect Bayesian trackers that assume continuous-valued motion models. The dynamic target aspect model remains, however, discrete-valued. To approximate the optimal sequential Bayesian estimator for this mixed discrete-valued/continuous-valued state model, we resort to sequential importance sampling [15], which is also known as particle filtering [16], Monte Carlo filtering [17], or, in the computer vision literature, the condensation algorithm [18], [19]. We present two *mixed-state* particle filter algorithms for direct *multiaspect* target tracking from image sequences. The proposed trackers are based, respectively, on the sampling/importance resampling (SIR) method [20], [21] and on the alternative auxiliary particle filter (APF) [22] technique. An earlier, simpler version of those trackers for *single aspect* targets was introduced in [23].

*Observation Model:* An important tenet in our approach to target detection/tracking, both in the grid-based and in the sequential importance sampling solutions, is that, unlike in previous work reported in the literature, we do not use validated measurements or contacts as input data for the tracker, but rather treat the image frames themselves as input data. In order to do that, we propose a nonlinear observation model that, when a target is present, maps a target centroid location into a spatial distribution of target pixels with a certain shape and intensity. The associated likelihood function fully incorporates the models for target shape and intensity and the model for the background clutter.

*Clutter Adaptation:* Previous work on multiframe detection—see, e.g., [24]—neglects the spatial correlation of the clutter treating the background as white Gaussian noise. We propose instead to capture the clutter spatial correlation using a 2-D noncausal Gauss–Markov random field (GMrf) model [25]–[27]. We use, in this paper, a suboptimal approximate maximum likelihood (AML) parameter estimation algorithm [26], [28] that fits a GMrf model to the clutter data at each frame in the image sequence.

*Outline of the Paper:* The paper is divided into seven sections. Section I is this introduction. In Section II, we introduce the models for target and clutter that underly our algorithms, and we review the proposed model for the likelihood function. Section III derives the proposed batch and online versions of the grid-based multiframe HMM detector/tracker. In Section IV, we introduce the mixed-state sequential Monte Carlo SIR and APF trackers. In Section V, we discuss clutter adaptation. Section VI examines the tracking performance of the various proposed algorithms using simulated multiaspect target sequences generated from real infrared airborne radar (IRAR) [29] data. Finally,

Section VII summarizes the main contributions and conclusions of our work.

## II. PROBLEM SETUP

Our primary goal is to estimate the 2-D spatial location of designated targets in an image sequence. For simplicity, we restrict our discussion to the situation when there is at most one target of interest present in the imaged scene. We formalize next the target motion, target signature, target aspect, and background clutter models that underly the derivation of the algorithms in this paper.

### A. Target Motion Model

In this subsection, we discuss two strategies for target translational motion modeling using, respectively, continuous-valued and discrete-valued state variables. Throughout 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) of (absolutely) continuous random variables/vectors and uppercase  $P$  to denote probability mass functions (pmfs) of discrete random variables/vectors.

*Continuous-State Model:* Assuming Cartesian coordinates, we use the indices  $i = 1$  and  $i = 2$  to refer, respectively, to each of the two dimensions of the plane. Let

$$\mathbf{x}_i(t) = \left[ x_i(t) \quad \frac{d}{dt}x_i(t) \right]^T, \quad t \in \mathfrak{R} \quad (1)$$

be a continuous-time state vector that collects the position and velocity of the target centroid at instant  $t$  in dimension  $i$ . Now, let

$$\mathbf{x}_{n,i} = \mathbf{x}_i(n\Delta), \quad n \in \mathcal{Z}, \quad i = 1, 2 \quad (2)$$

be the corresponding discrete-time state vector with  $\Delta$  denoting the sampling period in time. We build the four-dimensional target state vector

$$\mathbf{x}_n = \left[ \mathbf{x}_{n,1}^T \quad \mathbf{x}_{n,2}^T \right]^T \quad (3)$$

where  $\mathbf{x}_{n,i}$ ,  $i = 1, 2$ , are defined as in (2).

We assume that  $\{\mathbf{x}_n\}$ ,  $n \geq 0$  is a first-order Markov random sequence specified by the conditional probability density function (pdf)  $p(\mathbf{x}_{n+1}|\mathbf{x}_n)$  and by the pdf of the initial state  $p(\mathbf{x}_0)$ . In practice, we may assume further that the sequence of kinematic state vectors in dimension 1  $\{\mathbf{x}_{n,1}\}$  is statistically independent of the sequence of kinematic state vectors in dimension 2  $\{\mathbf{x}_{n,2}\}$  for  $n \geq 0$ . It follows then that

$$p(\mathbf{x}_{n+1}|\mathbf{x}_n) = p(\mathbf{x}_{n+1,1}|\mathbf{x}_{n,1})p(\mathbf{x}_{n+1,2}|\mathbf{x}_{n,2}). \quad (4)$$

In particular, in many real-world problems, we model the sequence of random vectors  $\{\mathbf{x}_{n,i}\}$ ,  $i = 1, 2$  by the (generally nonlinear) dynamic system

$$\mathbf{x}_{n+1,i} = \mathbf{f}_i(\mathbf{x}_{n,i}; n) + \mathbf{g}_i(\mathbf{x}_{n,i}; n)\mathbf{u}_{n,i}, \quad n \geq 0, \quad i = 1, 2 \quad (5)$$

where  $\{\mathbf{u}_{n,i}\}$ ,  $n \geq 0$  is assumed to be a sequence of zero-mean, independent, identically distributed (i.i.d.) Gaussian random vectors. Equation (5) is initialized with a stochastic initial condition  $\mathbf{x}_{0,i}$ , independent of the sequence  $\{\mathbf{u}_{n,i}\}$ ,  $n \geq 0$ , and with a pdf  $p(\mathbf{x}_{0,i})$ , which may be, in general, non-Gaussian.

*Discrete-State Model:* An alternative to the continuous-state motion model discussed before is to restrict the values of the unknown target centroid positions in dimensions  $i = 1$  and  $i = 2$  to pixel locations in the image grid  $\mathcal{L} = \{(r, j) | 1 \leq r \leq L, 1 \leq j \leq M\}$ , where  $L \times M$  is the size in pixels of each frame in the image sequence. We then track the centroid pixel locations directly, as opposed to tracking real-valued centroid coordinates. Let

$$\mathbf{z}_n = (r_n, j_n) \quad (r_n, j_n) \in \mathcal{L} \quad (6)$$

denote the unknown pixel location of the target centroid. The dynamic state evolution of the hidden vector  $\mathbf{z}_n$  can be modeled by a 2-D discrete-valued hidden Markov chain specified by the transition probability masses  $P(\mathbf{z}_n | \mathbf{z}_{n-1})$  and by the initial probability masses  $P(\mathbf{z}_0)$ ,  $\mathbf{z}_0 \in \mathcal{L}$ .

### B. Observation Model

The raw sensor measurements at instant  $n$  are sampled and processed to form a 2-D digital sensor image, which is referred to as a *frame*. Let  $H_0$  denote the hypothesis that the designated target is absent from the scene at frame  $n$  and denote by  $H_1$  the hypothesis that the target of interest is present at frame  $n$ . Assuming a *continuous-state* motion model, we represent the  $n$ th frame by the  $L \times M$  matrix

$$H_1 : \mathbf{Y}_n = \mathbf{H}(\mathbf{x}_n^*, s_n) + \mathbf{V}_n \quad (7)$$

$$H_0 : \mathbf{Y}_n = \mathbf{V}_n \quad (8)$$

where matrix  $\mathbf{V}_n$  represents the background clutter, and matrix  $\mathbf{H}(\mathbf{x}_n^*, s_n)$  is the clutter-free target image model, which is a function of the 2-D pixel location of the target centroid  $\mathbf{x}_n^*$  and the target aspect state  $s_n$ , both of which are unknown and random and must be sequentially estimated from the observed data  $\{\mathbf{Y}_n\}$ . The 2-D hidden random vector  $\mathbf{x}_n^*$  takes values on the finite sensor grid  $\mathcal{L} = \{(r, j) | 1 \leq r \leq L, 1 \leq j \leq M\}$  (see Section II-A) and is obtained from the 4-D continuous-valued state vector  $\mathbf{x}_n$  in (3) by making

$$x_n^*(1) = \text{round} \left( \frac{x_{n,1}(1)}{\xi_1} \right) \quad (9)$$

$$x_n^*(2) = \text{round} \left( \frac{x_{n,2}(1)}{\xi_2} \right) \quad (10)$$

where  $\xi_1$  and  $\xi_2$  are the image resolutions, respectively, in dimensions  $i = 1$  and  $i = 2$ . We are reminded [see (2)] that  $x_{n,i}(1)$ ,  $i = 1, 2$  denote the position coordinates of the target centroid at instant  $t = n\Delta$ , respectively, in dimensions  $i = 1$  and  $i = 2$ .

Otherwise, assuming a *discrete-valued* target position state  $\mathbf{z}_n$  (see Section II-A), we can write the observation model for frame  $n$  under hypothesis  $H_1$  (target present) directly as

$$H_1 : \mathbf{Y}_n = \mathbf{H}(\mathbf{z}_n, s_n) + \mathbf{V}_n. \quad (11)$$

*Target Aspect Model:* We assume that the target aspect state  $s_n$  is defined on a finite, discrete-valued set  $\mathcal{I} = \{1, 2, 3, \dots, K\}$ , where each index  $k \in \mathcal{I}$  is a pointer to one possible template model in a target aspect library such that the several entries in the library represent possible rotation, scaling, and/or shearing of the target's base template. The random sequence of aspect states  $\{s_n\}$ ,  $n \geq 0$  is modeled as a first-order HMM defined on the set  $\mathcal{I}$  and specified by the vector of initial probabilities  $\underline{\pi}^s$  such that  $\pi^s(k) = P(s_0 = k)$ ,  $k \in \mathcal{I}$ , and by the matrix of transition probabilities  $\mathbf{T}_2$  such that

$$T_2(k, l) = P(s_n = k | s_{n-1} = l) \quad (k, l) \in \mathcal{I} \times \mathcal{I}. \quad (12)$$

*Clutter-Free Target Model:* We assume that at any given frame for any aspect state  $s_n$ , 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.

For each pixel centroid position  $\mathbf{x}_n^* = (r_n, j_n) \in \mathcal{L}$  in the continuous-valued motion model, or  $\mathbf{z}_n = (r_n, j_n) \in \mathcal{L}$  in the discrete-valued motion model, the nonlinear function  $\mathbf{H}(\cdot, \cdot)$  in (7) and (11) returns a spatial distribution of (real-valued) pixel intensities  $\{a_{k,l}(s_n)\}$ ,  $-r_i \leq k \leq r_s$ ,  $-l_i \leq l \leq l_s$ , centered at  $(r_n, j_n)$ , and dependent on the aspect state  $s_n$ . Formally, we write

$$\mathbf{H}(r_n, j_n, s_n) = \sum_{k=-r_i}^{r_s} \sum_{l=-l_i}^{l_s} a_{k,l}(s_n) \mathbf{E}_{r_n+k, j_n+l} \quad (13)$$

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  $s_n = \tilde{s} \in \mathcal{I}$ , the coefficients  $\{a_{k,l}(\tilde{s})\}$  in (13) are referred to as the target *signature parameters* corresponding to that particular template. The signature coefficients are the product of a binary parameter  $b_{k,l}(s_n) \in \mathcal{B} = \{0, 1\}$  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  $s_n$ .

*Remark:* To write (13), 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 (13) for centroid locations near the borders.

*Clutter Model:* We capture the 2-D spatial correlation of the background clutter using a noncausal, spatially homogeneous Gauss-Markov random field (GMrf) model [25]. The random clutter returns at frame  $n$ ,  $V_n(r, j)$ ,  $1 \leq r \leq L$ ,  $1 \leq j \leq M$  are then described by the 2-D finite difference equation

$$V_n(r, j) = \beta_{v,n}^c [V_n(r-1, j) + V_n(r+1, j)] + \beta_{h,n}^c [V_n(r, j-1) + V_n(r, j+1)] + \varepsilon_n(r, j) \quad (14)$$

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

Let  $\mathbf{v}_n$  be a 1-D equivalent representation of the matrix  $\mathbf{V}_n$  in (14) that is obtained by scanning the matrix row by row and sequentially stacking the scanned rows in a long vector. Writing (14) in compact matrix notation, it can be shown [25], [27] by the application of the principle of orthogonality that the inverse of the covariance matrix  $\Sigma_v = E[\mathbf{v}_n \mathbf{v}_n^T]$  has a *block-tridiagonal* structure of the form

$$\sigma_{c,n}^2 \Sigma_v^{-1} = \mathbf{I}_L \otimes (\mathbf{I}_M - \beta_{h,n}^c \mathbf{K}_M) + \mathbf{K}_L \otimes (-\beta_{v,n}^c \mathbf{I}_M) \quad (15)$$

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

### C. Likelihood Function

Let  $\mathbf{y}_n$  and  $\mathbf{h}(\mathbf{x}_n^*, s_n)$  be the 1-D long vector representations, respectively, of the image frame  $\mathbf{Y}_n$  and the clutter-free target image  $\mathbf{H}(\mathbf{x}_n^*, s_n)$  in (7), assuming *continuous-valued* kinematic target states as in (3). For a GMrf background as in (14), the likelihood function under hypothesis  $H_1$  for a fixed template state  $s_n = \tilde{s}$  may be written as

$$p(\mathbf{y}_n | \mathbf{x}_{n,1}, \mathbf{x}_{n,2}, \tilde{s}) \propto \exp \left[ \frac{2\lambda(\mathbf{x}_{n,1}, \mathbf{x}_{n,2}, \tilde{s}) - \rho(\tilde{s})}{2\sigma_{c,n}^2} \right] \quad (16)$$

where the symbol  $\propto$  denotes ‘‘proportional to,’’  $\rho(\tilde{s})$  is a target energy term given by

$$\rho(\tilde{s}) = \mathbf{h}^T(\mathbf{x}_n^*, \tilde{s}) \left( \sigma_{c,n}^2 \Sigma_v^{-1} \right) \mathbf{h}(\mathbf{x}_n^*, \tilde{s}) \quad (17)$$

that does not vary with  $\mathbf{x}_n$  away from the image borders (see [6] for further details), and  $\lambda(\mathbf{x}_n, \tilde{s})$  is a data-dependent term such that

$$\lambda(\mathbf{x}_n, \tilde{s}) = \mathbf{y}_n^T \left( \sigma_{c,n}^2 \Sigma_v^{-1} \right) \mathbf{h}(\mathbf{x}_n^*, \tilde{s}). \quad (18)$$

Using the structure of the inverse of the covariance matrix in (15), it has been shown in [30] that  $\lambda(\mathbf{x}_n, \tilde{s})$  in (18) is computed by the expression

$$\lambda(\mathbf{x}_{n,1}, \mathbf{x}_{n,2}, \tilde{s}) = \sum_{k=-r_i}^{r_s} \sum_{l=-l_i}^{l_s} a_{k,l}(\tilde{s}) d(x_n^*(1) + k, x_n^*(2) + l) \quad (19)$$

where  $x_n^*(i)$ ,  $i = 1, 2$  are obtained, respectively, from (9) and (10), and  $d(r, j)$  is the output of the differential operator

$$d(r, j) = Y_n(r, j) - \beta_{h,n}^c [Y_n(r, j-1) + Y_n(r, j+1)] - \beta_{v,n}^c [Y_n(r-1, j) + Y_n(r+1, j)] \quad (20)$$

with Dirichlet (identically zero) boundary conditions. Equation (19) is valid for  $r_i + 1 \leq x_n^*(1) \leq L - r_s$  and  $l_i + 1 \leq x_n^*(2) \leq$

$M - l_s$ . For centroid positions close to the image borders, the summation limits in (19) must be varied accordingly, as shown in [6].

*Remark:* For a *discrete-valued* target state  $\mathbf{z}_n$  as in (6), we may rewrite the likelihood function equations by replacing  $x_n^*(1)$  and  $x_n^*(2)$  directly with the components of the vector  $\mathbf{z}_n$ .

## III. GRID-BASED HMM DETECTION/TRACKING

In this section, we derive grid-based Bayesian algorithms for joint multiframe target detection and tracking in cluttered image sequences using the models from Section II. We build an integrated framework for detection and tracking by conveniently extending the image grid  $\mathcal{L}$  to account for both boundary effects and absence of target.

### A. Lattice Extension and Integration of Detection and Tracking

To model situations when targets move in and out of the sensor image, we introduce the augmented *centroid lattice*  $\hat{\mathcal{L}} = \{(r, j) : -r_s + 1 \leq r \leq L + r_i, -l_s + 1 \leq j \leq M + l_i\}$ . The centroid lattice collects all possible values of the target centroid position for which at least one target pixel may lie inside the sensor's image.

Let  $\tilde{\mathcal{L}}$  be an equivalent 1-D representation of the centroid lattice  $\hat{\mathcal{L}}$  obtained by row lexicographic ordering. We build an integrated framework for detection and tracking by augmenting  $\tilde{\mathcal{L}}$  with an additional dummy state that represents the absence of the target. For convenience, we assign to the absent state the index  $L_1 + 1$ , where  $L_1 = (L + r_i + r_s)(M + l_i + l_s)$ . The final 1-D *extended lattice* is

$$\tilde{\mathcal{L}} = \{l : 1 \leq l \leq (L + r_i + r_s)(M + l_i + l_s) + 1\}. \quad (21)$$

*Modified Discrete-Valued Motion Model:* Now, let  $z_n \in \tilde{\mathcal{L}}$  be the new unknown (hidden) random variable at frame  $n$ . If the target is present at frame  $n$  at a centroid location  $\mathbf{z}_n = (r_n, j_n)$ ,  $(r_n, j_n) \in \hat{\mathcal{L}}$ , the scalar variable  $z_n$  takes, by definition, the value

$$z_n = (r_n + r_s - 1)(M + l_i + l_s) + j_n + l_s.$$

Otherwise, if the target is absent at frame  $n$ , then, as explained before,  $z_n = L_1 + 1$ . Like in Section II-A, the dynamic evolution of the sequence of random variables  $\{z_n\}$ ,  $n \geq 0$  may be described by a discrete-valued Markov chain, which is now defined on the extended lattice  $\tilde{\mathcal{L}}$  that includes the absent target state. This Markov chain is specified by the matrix of transition probabilities  $\mathbf{T}_1$  such that

$$T_1(l, q) = P(z_n = l | z_{n-1} = q) \quad (l, q) \in \tilde{\mathcal{L}} \times \tilde{\mathcal{L}}. \quad (22)$$

Realizations of this Markov chain are initialized with samples from the distribution of prior probabilities specified by the vector  $\underline{\pi}$  such that

$$\pi(l) = P(z_0 = l) \quad l \in \tilde{\mathcal{L}}. \quad (23)$$

### B. HMM Smoother

Let  $\mathbf{y}_n$  be, as before, an equivalent long vector representation of the  $n$ th sensor frame  $\mathbf{Y}_n$ , and let  $\mathbf{Y}_0^N = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_N\}$  be a collection of  $N+1$  observed frames. Introduce the matrices  $\underline{\alpha}_n$ ,  $\underline{\beta}_n$ , and  $\mathbf{S}_n$  such that

$$\alpha_n(i, j) = p(z_n = i, s_n = j, \mathbf{Y}_0^n) \quad (24)$$

$$\beta_n(i, j) = p(\mathbf{Y}_{n+1}^N | z_n = i, s_n = j) \quad (25)$$

$$\mathbf{S}_n(i, j) = p(\mathbf{y}_n | z_n = i, s_n = j) \quad (26)$$

for  $i \in \tilde{\mathcal{L}}$  and  $j \in \mathcal{I}$ . Note that the term on the right-hand side of (24) is a mixed probability density function, which is defined as an actual density function in the continuous-valued variables and a probability mass function in the discrete-valued variables. We use lowercase  $p$  on the right-hand side of (24) as an abuse of notation.

In order to obtain the optimal Bayesian estimates of the unknown target position and aspect at frame  $n$ , we must compute the joint posterior statistics of the hidden variables  $z_n$  and  $s_n$  at each frame  $n$  conditioned on the observed data  $\mathbf{Y}_0^N$ . We then derive, in the sequel, an algorithm for the recursive computation of the matrix  $\underline{\gamma}_n$  such that

$$\underline{\gamma}_n(i, j) = P(z_n = i, s_n = j | \mathbf{Y}_0^N) \quad (i, j) \in \tilde{\mathcal{L}} \times \mathcal{I}. \quad (27)$$

We make the following assumptions in the derivation.

- The sequence of clutter frames  $\{\mathbf{V}_k\}$ ,  $k \geq 0$ , after a pre-processing of the data that subtracts the spatially variant local background mean, is i.i.d. as well as statistically independent of the sequences of target centroid positions  $\{z_k\}$  and target aspect states  $\{s_k\}$ ,  $k \geq 0$ .
- The sequences  $\{z_k\}$  and  $\{s_k\}$ ,  $k \geq 0$  are mutually independent first-order discrete Markov processes.

*Forward Recursion:* From the observation model and using the previous assumptions that the sequence  $\{\mathbf{V}_k\}$  is i.i.d., we note that

$$p(\mathbf{y}_{n+1} | z_{n+1} = i, s_{n+1} = j, \mathbf{Y}_0^n) = p(\mathbf{y}_{n+1} | z_{n+1} = i, s_{n+1} = j). \quad (28)$$

On the other hand, using the assumptions that the sequences  $\{z_k\}$  and  $\{s_k\}$  are two mutually independent first-order Markov sequences and are also statistically independent of the sequence of clutter frames  $\{\mathbf{V}_k\}$ , we conclude that

$$\begin{aligned} P(z_{n+1} = i, s_{n+1} = j | z_n = l, s_n = r, \mathbf{Y}_0^n) \\ = P(z_{n+1} = i, s_{n+1} = j | z_n = l, s_n = r) \\ = P(z_{n+1} = i | z_n = l) P(s_{n+1} = j | s_n = r). \end{aligned} \quad (29)$$

Using Bayes' law and (28), we write

$$\begin{aligned} \alpha_{n+1}(i, j) &= p(z_{n+1} = i, s_{n+1} = j, \mathbf{Y}_0^{n+1}) \\ &= \underbrace{p(\mathbf{y}_{n+1} | z_{n+1} = i, s_{n+1} = j)}_{S_{n+1}(i, j)} \\ &\quad \times p(z_{n+1} = i, s_{n+1} = j, \mathbf{Y}_0^n). \end{aligned}$$

Using now the Theorem of Total Probability and recalling (29), we get

$$\begin{aligned} p(z_{n+1} = i, s_{n+1} = j, \mathbf{Y}_0^n) \\ = \sum_l \sum_r p(z_{n+1} = i, s_{n+1} = j, z_n = l, s_n = r, \mathbf{Y}_0^n) \\ = \sum_r \sum_l P(z_{n+1} = i, s_{n+1} = j | z_n = l, s_n = r) \\ \quad \times p(z_n = l, s_n = r, \mathbf{Y}_0^n) \\ = \sum_l \underbrace{P(z_{n+1} = i | z_n = l)}_{T_1(i, l)} \sum_r \underbrace{p(z_n = l, s_n = r, \mathbf{Y}_0^n)}_{\alpha_n(l, r)} \\ \quad \times \underbrace{P(s_{n+1} = j | s_n = r)}_{T_2(j, r)} \end{aligned} \quad (30)$$

where  $\mathbf{T}_1$  and  $\mathbf{T}_2$  are the motion and aspect transition probability matrices defined, respectively, in (22) and (12). Using (30), we now write

$$\alpha_{n+1}(i, j) = S_{n+1}(i, j) \sum_l T_1(i, l) \left[ \sum_r \alpha_n(l, r) T_2^T(r, j) \right] \quad (31)$$

or, in compact matrix notation

$$\underline{\alpha}_{n+1} = \mathbf{S}_{n+1} \odot [\mathbf{T}_1 \underline{\alpha}_n \mathbf{T}_2^T] \quad (32)$$

where  $\odot$  denotes pointwise multiplication, and the superscript  $T$  stands for the transpose of a matrix.

*Backward Recursion:* From the Theorem of Total Probability

$$\begin{aligned} \beta_n(i, j) &= p(\mathbf{Y}_{n+1}^N | z_n = i, s_n = j) \\ &= \sum_l \sum_r p(\mathbf{Y}_{n+1}^N, z_{n+1} = l \\ &\quad s_{n+1} = r | z_n = i, s_n = j). \end{aligned} \quad (33)$$

From the previous assumptions on the statistics of  $\{z_k\}$ ,  $\{s_k\}$ , and  $\{\mathbf{V}_k\}$ , we note, however, that

$$p(\mathbf{y}_{n+1} | z_{n+1} = l, s_{n+1} = r, z_n = i, s_n = j, \mathbf{Y}_{n+2}^N) = p(\mathbf{y}_{n+1} | z_{n+1} = l, s_{n+1} = r). \quad (34)$$

Similarly, we also note that

$$p(\mathbf{Y}_{n+2}^N | z_{n+1} = l, s_{n+1} = r, z_n = i, s_n = j) = p(\mathbf{Y}_{n+2}^N | z_{n+1} = l, s_{n+1} = r). \quad (35)$$

Using Bayes' law to expand (33) and recalling (34) and (35), we get

$$\begin{aligned} \beta_n(i, j) &= \sum_l \sum_r \underbrace{p(\mathbf{y}_{n+1} | z_{n+1} = l, s_{n+1} = r)}_{S_{n+1}(l, r)} \\ &\quad \times \underbrace{p(\mathbf{Y}_{n+2}^N | z_{n+1} = l, s_{n+1} = r)}_{\beta_{n+1}(l, r)} \\ &\quad \times \underbrace{P(z_{n+1} = l | z_n = i)}_{T_1(l, i)} \underbrace{P(s_{n+1} = r | s_n = j)}_{T_2(r, j)} \\ &= \sum_l T_1^T(i, l) \sum_r S_{n+1}(l, r) \\ &\quad \times \beta_{n+1}(l, r) T_2(r, j) \end{aligned} \quad (36)$$

or, in compact matrix notation

$$\underline{\beta}_n = \mathbf{T}_1^T [\mathbf{S}_{n+1} \odot \underline{\beta}_{n+1}] \mathbf{T}_2. \quad (37)$$

*Initialization:* In the sequel, we discuss how to initialize the forward and backward recursions. In the forward case, we have

$$\begin{aligned} \alpha_0(i, j) &= p(z_0 = i, s_0 = j, \mathbf{y}_0) \\ &= \underbrace{p(\mathbf{y}_0 | z_0 = i, s_0 = j)}_{S_0(i, j)} \underbrace{P(z_0 = i)}_{\pi_i} \underbrace{P(s_0 = j)}_{\pi_j^s} \end{aligned} \quad (38)$$

where  $\pi_i = P(z_0 = i)$  and  $\pi_j^s = P(s_0 = j)$  are the prior probability mass functions, respectively, for the initial centroid position and aspect of the target. Similarly, for the backward recursion, we write

$$\begin{aligned} \beta_{N-1}(i, j) &= p(\mathbf{y}_N | z_{N-1} = i, s_{N-1} = j) \\ &= \sum_k \sum_l p(\mathbf{y}_N | z_N = k, s_N = l) \\ &\quad \times P(z_N = k | z_{N-1} = i) P(s_N = l | s_{N-1} = j) \\ &= \sum_k \sum_l \mathbf{T}_1^T(i, k) \\ &\quad \times [S_N(k, l) \beta_N(k, l)] \mathbf{T}_2(l, j) \end{aligned} \quad (39)$$

where, by definition

$$\beta_N(k, l) = 1 \quad \forall (k, l). \quad (40)$$

*Smoothing Posterior:* From the modeling assumptions

$$\begin{aligned} \gamma_n(i, j) &= P(z_n = i, s_n = j | \mathbf{Y}_0^N) \\ &= C_N p(\mathbf{Y}_{n+1}^N | z_n = i, s_n = j) p(z_n = i, s_n = j, \mathbf{Y}_0^n) \\ &= C_N \alpha_n(i, j) \beta_n(i, j) \end{aligned} \quad (41)$$

where  $C_N = 1/p(\mathbf{Y}_0^N)$  is a normalization constant. Equation (41) may be rewritten in compact matrix notation as

$$\underline{\gamma}_n = C_N \underline{\alpha}_n \odot \underline{\beta}_n. \quad (42)$$

From  $\underline{\gamma}_n$ , we obtain the marginal posterior probability mass function for the hidden centroid position  $z_n$

$$\tilde{\gamma}_n(i) = P(z_n = i | \mathbf{Y}_0^N) = \sum_j \gamma_n(i, j). \quad (43)$$

*Multiframe Target Detection:* Denoting by  $H_0$  and  $H_1$  the hypotheses that the target is, respectively, absent from and present in the scene at frame  $n$ , the minimum probability of error multiframe detector at frame  $n$  is given by the test

$$\tilde{\gamma}_n(L_1 + 1) \underset{H_1}{\overset{H_0}{\gtrless}} 1 - \tilde{\gamma}_n(L_1 + 1) \quad (44)$$

where  $z_n = L_1 + 1$  is the absent target state; see Section III-A.

*Multiframe Target Tracking:* If hypothesis  $H_1$  is declared true at instant  $n$ , we find then the multiframe maximum *a posteriori* (MAP) estimate of the target's centroid position  $z_n$  at frame  $n$  using the expression

$$\hat{z}_{n|N+1} = \arg \max_{l \in \bar{\mathcal{L}}} \tilde{\gamma}_n(l) \quad (45)$$

where  $\bar{\mathcal{L}}$  is the centroid lattice; see Section III-A.

*Summary of the Algorithm:* Table I summarizes the main steps in the proposed multiaspect, multiframe detector/tracker.

Before we leave this subsection, we make a few brief remarks concerning the actual implementation of the algorithm in Table I.

*Remark 1:* The forward and backward recursions in (32) and (37) may be interpreted as a generalization of the Baum–Welch smoother [12], [13]. We perform simultaneous smoothing of two hidden Markov chains and process the observed sensor images directly, incorporating the 2-D models for target signature and clutter. For the present target states, the corresponding entries in the matrix  $\mathbf{S}_n$  are computed up to a proportionality constant using the likelihood function model in Section II-C. For the absent target state, the conditional density of the observation reduces to the clutter probability density function. We omit algorithmic details on the normalization of  $\mathbf{S}_n$  for the sake of conciseness.

*Remark 2:* As reported in the HMM literature (see, e.g., [11]), Baum–Welch smoothers may suffer from numerical instability if  $N$  is large. To avoid these problems, practical implementations of our algorithm use additional normalizations of the matrices  $\underline{\alpha}$  and  $\underline{\beta}$  for numerical error control.

*Remark 3:* Most discrete-valued motion models are based on a principle of locality that leads to a highly sparse and structured transition matrix  $\mathbf{T}_1$ . We take advantage of this structure to speed up computations and save memory storage space. Further details are omitted here for lack of space.

### C. HMM Filter

The multiframe detector/tracker of Section III-B is a batch algorithm that uses forward and backward recursions to compute the joint posterior probability mass function of the hidden variables at instant  $n$  conditioned on all past, present, and future frames in the data volume. In this section, we derive an alternative *online* algorithm that is based on the recursive computation of the joint posterior *filtering* probability mass function

$$f_{n|n}(i, j) = P(z_n = i, s_n = j | \mathbf{Y}_0^n) \quad (46)$$

conditioned only on present and past data. Note, however, that

$$f_{n|n}(i, j) = \frac{p(z_n = i, s_n = j, \mathbf{Y}_0^n)}{p(\mathbf{Y}_0^n)} = C_n \alpha_n(i, j). \quad (47)$$

From (47), we conclude that, barring a normalization constant, the recursion for the computation of  $f_{n|n}$  coincides with the forward recursion of the HMM smoother, i.e.,

$$\mathbf{f}_{n+1|n+1} = K_{n+1} \mathbf{S}_{n+1} \odot [\mathbf{T}_1 \mathbf{f}_{n|n} \mathbf{T}_2^T] \quad (48)$$

where  $K_{n+1} = 1/p(\mathbf{y}_{n+1} | \mathbf{Y}_0^n)$  is computed such that  $f_{n+1|n+1}(i, j)$  is summable to 1. From  $\mathbf{f}_{n|n}$ , we compute the marginal filtering distribution

$$\tilde{f}_{n|n}(i) = \sum_j f_{n|n}(i, j). \quad (49)$$

The multiframe detection and tracking steps are now identical to the detector and estimator in (44) and (45) but replacing  $\tilde{\gamma}_n$  with  $\tilde{f}_{n|n}$ .

*Remark:* In practical situations, assuming *a priori* knowledge of the class of objects to be tracked, the discrete-valued

TABLE I  
ALGORITHM I: GRID-BASED MULTIFRAME, MULTIASPECT  
DETECTOR/TRACKER USING BAYESIAN SMOOTHING

---

1. **Initialization**

- $\alpha_0(i, j) = \pi_i S_0(i, j) \pi_j^s \quad i \in \tilde{\mathcal{L}}, j \in \mathcal{I}$ .
- $\beta_N(k, l) = 1 \quad k \in \tilde{\mathcal{L}}, l \in \mathcal{I}$ .

2. **Forward Recursion**

For  $n = 0$  to  $N - 1$

- $\underline{\alpha}_{n+1} = \mathbf{S}_{n+1} \odot [\mathbf{T}_1 \underline{\alpha}_n \mathbf{T}_2^T]$ .

End of forward recursion.

3. **Backward Recursion**

For  $n = N - 1$  to  $0$

- $\underline{\beta}_n = \mathbf{T}_1^T [\mathbf{S}_{n+1} \odot \underline{\beta}_{n+1}] \mathbf{T}_2$ .

End of backward recursion.

4. **Multiframe Detection/Tracking**

For  $n = 0$  to  $N$

- $\underline{\gamma}_n \propto \underline{\alpha}_n \odot \underline{\beta}_n, \quad \sum_i \sum_j \gamma_n(i, j) = 1$ .
- for  $i = 1$  to  $L_1 + 1$
- $\tilde{\gamma}_n(i) = \sum_j \gamma_n(i, j)$ .
- end

$$\begin{matrix} H_0 \\ \gtrsim \\ H_1 \end{matrix}$$

- **Detection Test:**  $\tilde{\gamma}_n(L_1 + 1) \gtrsim 1 - \tilde{\gamma}_n(L_1 + 1)$ .
- If  $H_1$  decided,  $\hat{z}_{nN+1} = \arg \max_l \tilde{\gamma}_n(l)$ .

End

---

target motion HMM may be derived by discretizing in space the (known) real-valued dynamic model of the target of interest. On the other hand, assuming knowledge of the motion parameters and assuming that a volume of training data is available, the unknown HMM target aspect parameters may be recursively estimated using the expectation–maximization (EM) [31] algorithm. An efficient implementation of the EM recursion for HMM training is obtained using the Baum–Welch re-estimation algorithm [12], [13]. We omit this discussion here for conciseness.

#### IV. MIXED-STATE PARTICLE FILTER TRACKERS

The grid-based trackers introduced in Section III outperform conventional contact/association trackers in scenarios of low target-to-clutter ratio. However, they suffer from two limitations. First, a discrete-valued target motion model may be not flexible enough to capture the structure of real-world motion. Second, grid-based filters are computationally intensive. A less computationally intensive and more flexible alternative to the grid-based HMM trackers is to retain a (possibly nonlinear) continuous-valued target motion model, as described in Section II-A, and resort to sampling/importance resampling (SIR) filters, which are also known as particle filters [15]–[17], [20]. SIR filters are sequential Monte Carlo methods for dynamic state estimation that have been successfully applied to several practical problems of interest including mobile robot localization [32], computer vision [33], [34], terrain navigation [35], and many others. In the sequel, we propose a mixed-state SIR filter for multiaspect target tracking in images. Similar mixed-state particle filters assuming different dynamic and observation models have been used in the context of computer vision in [19]. For independent related work, see also [36].

TABLE II  
ALGORITHM II: MIXED-STATE SIR FILTER FOR MULTIASPECT TARGET  
TRACKING IN CLUTTERED IMAGE SEQUENCES

---

1. **Initialization** For  $j = 1, \dots, N_p$

- Draw  $\mathbf{x}_{0,1}^{(j)} \sim p(\mathbf{x}_{0,1}), \mathbf{x}_{0,2}^{(j)} \sim p(\mathbf{x}_{0,2})$ , and  $s_0^{(j)} \sim P(s_0)$ .
- Make  $w_0^{(j)} = 1/N_p$  and set  $n = 1$ .

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

- Draw  $\tilde{\mathbf{x}}_{n,1}^{(j)} \sim p(\mathbf{x}_{n,1} | \mathbf{x}_{n-1,1}^{(j)})$  and  $\tilde{\mathbf{x}}_{n,2}^{(j)} \sim p(\mathbf{x}_{n,2} | \mathbf{x}_{n-1,2}^{(j)})$ .
- Draw  $\tilde{s}_n^{(j)} \sim P(s_n | s_{n-1}^{(j)})$ .
- Compute the importance weights  $\tilde{w}_n^{(j)} \propto w_{n-1}^{(j)} p(\mathbf{y}_n | \tilde{\mathbf{x}}_{n,1}^{(j)}, \tilde{\mathbf{x}}_{n,2}^{(j)}, \tilde{s}_n^{(j)}) \quad \sum_{j=1}^{N_p} \tilde{w}_n^{(j)} = 1$  using equations (16), (19), and (20).

3. **Selection**

- Generate a new set of samples  $\{\mathbf{x}_{n,1}^{(j)}, \mathbf{x}_{n,2}^{(j)}, s_n^{(j)}\} \quad 1 \leq j \leq N_p$  such that  $P([\mathbf{x}_n^{(j)}, s_n^{(j)}] = [\tilde{\mathbf{x}}_n^{(k)}, \tilde{s}_n^{(k)}]) = \tilde{w}_n^{(k)}$ .
- Make  $w_n^{(j)} = 1/N_p, 1 \leq j \leq N_p$ .
- Set  $n = n + 1$  and go back to step 2.

---

TABLE III  
ALGORITHM III: MIXED-STATE AUXILIARY PARTICLE FILTER FOR  
MULTIASPECT TARGET TRACKING IN IMAGE SEQUENCES

---

For  $n=1,2,\dots$

**Importance Sampling:** for  $j = 1, \dots, N_p$

- Draw  $\underline{\mu}_{n,1}^{(j)} \sim p(\mathbf{x}_{n,1} | \mathbf{x}_{n-1,1}^{(j)})$ ,  $\underline{\mu}_{n,2}^{(j)} \sim p(\mathbf{x}_{n,2} | \mathbf{x}_{n-1,2}^{(j)})$  and  $\mu_{n,a}^{(j)} \sim P(s_n | s_{n-1}^{(j)})$ .
- Compute the first-stage weights  $\Lambda_n^{(j)} \propto w_{n-1}^{(j)} p(\mathbf{y}_n | \underline{\mu}_{n,1}^{(j)}, \underline{\mu}_{n,2}^{(j)}, \mu_{n,a}^{(j)}) \quad \sum_{j=1}^{N_p} \Lambda_n^{(j)} = 1$  using equations (16), (19), and (20).
- Draw  $k^{(j)} \sim \{1, 2, \dots, N_p\}$  with  $\{P(k^{(j)} = i) = \Lambda_n^{(i)}\}, i = 1, \dots, N_p$ .
- Draw  $\mathbf{x}_{n,1}^{(j)} \sim p(\mathbf{x}_{n,1} | \mathbf{x}_{n-1,1}^{(k^{(j)})})$ ,  $\mathbf{x}_{n,2}^{(j)} \sim p(\mathbf{x}_{n,2} | \mathbf{x}_{n-1,2}^{(k^{(j)})})$  and  $s_n^{(j)} \sim P(s_n | s_{n-1}^{(k^{(j)})})$ .
- Compute the second-stage weights  $w_n^{(j)} \propto \frac{p(\mathbf{y}_n | \mathbf{x}_{n,1}^{(j)}, \mathbf{x}_{n,2}^{(j)}, s_n^{(j)})}{p(\mathbf{y}_n | \underline{\mu}_{n,1}^{(k^{(j)})}, \underline{\mu}_{n,2}^{(k^{(j)})}, \mu_{n,a}^{(k^{(j)})})} \quad \sum_{j=1}^{N_p} w_n^{(j)} = 1$  using equations (16), (19), and (20).

---

##### A. Mixed-State SIR Tracker

In a simpler tracking-only problem (i.e. assuming presence of target), we treat the continuous-state kinematic target vector  $\mathbf{x}_n$  in (3) (see Section II-A) and the discrete-valued target aspect state  $s_n$  (see Section II-B) as the hidden variables at frame  $n$ . Discarding the initial frame  $\mathbf{y}_0$ , our goal is to represent the mixed posterior  $p(\mathbf{x}_n, s_n | \mathbf{Y}_1^n)$  at step  $n$  by a properly weighted set of particles  $\{\mathbf{x}_n^{(j)}, s_n^{(j)}\}, 1 \leq j \leq N_p$ , with associated weights  $\{w_n^{(j)}\}$  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, i.e.,

$$\sum_{j=1}^{N_p} w_n^{(j)} \mathbf{x}_n^{(j)} \xrightarrow{N_p \rightarrow \infty} \int \mathbf{x}_n p(\mathbf{x}_n | \mathbf{Y}_1^n) d\mathbf{x}_n \quad (50)$$

$$\sum_{j=1}^{N_p} w_n^{(j)} s_n^{(j)} \xrightarrow{N_p \rightarrow \infty} \sum_{s_n \in \mathcal{I}} s_n P(s_n | \mathbf{Y}_1^n). \quad (51)$$

For all possible random realizations of the particle filter method, convergence in (50) and (51) is usually interpreted (see [37]) in the mean square sense.

*Importance Sampling Step:* Let  $\mathbf{X}_n = [\mathbf{x}_n^T s_n]^T$  now be the mixed state vector at frame  $n$ . Using the mixed density version of Bayes' law, we can factor  $p(\mathbf{X}_n | \mathbf{X}_{n-1})$  as

$$p(\mathbf{X}_n | \mathbf{X}_{n-1}) = p(\mathbf{x}_n | s_n, \mathbf{x}_{n-1}, s_{n-1}) P(s_n | s_{n-1}, \mathbf{x}_{n-1}). \quad (52)$$

From (52), it follows that, using  $p(\mathbf{X}_n | \mathbf{X}_{n-1})$  as importance function, a possible sequential importance sampling methodology for recursive generation of the particle set  $\{\mathbf{x}_n^{(j)}, s_n^{(j)}\}$  and their proper weights at frame  $n$  is as follows.

For  $i = 1$  to  $N_p$

- Draw  $s_n^{(j)} \sim P(s_n | s_{n-1}^{(j)}, \mathbf{x}_{n-1}^{(j)})$ .
- Draw  $\mathbf{x}_n^{(j)} \sim p(\mathbf{x}_n | s_n^{(j)}, \mathbf{x}_{n-1}^{(j)}, s_{n-1}^{(j)})$ .
- Update the weights  $w_n^{(j)} \propto w_{n-1}^{(j)} p(\mathbf{y}_n | \mathbf{x}_n^{(j)}, s_n^{(j)})$ .
- Normalize the weights such that  $\sum_{j=1}^{N_p} w_n^{(j)} = 1$ .

Note that we dropped in (52) the requirement in Section III of statistical independence between the sequences  $\{\mathbf{x}_n\}$  and  $\{s_n\}$ . The probabilities of the different possible aspect changes between frames  $n-1$  and  $n$  are now dependent on the target kinematic state at frame  $n-1$ . Furthermore, instead of assuming a fixed motion transition kernel  $p(\mathbf{x}_n | \mathbf{x}_{n-1})$ , the problem setup now includes a collection of different motion models  $p_{rl}(\mathbf{x}_n | \mathbf{x}_{n-1})$  parameterized by the aspect state pairs  $(s_n = r, s_{n-1} = l)$ ,  $(r, l) \in \mathcal{I} \times \mathcal{I}$ . For computational simplicity and in order to compare the tracking performance of the proposed HMM and particle filters, we will assume, however, in the numerical simulations in this paper that without loss of generality, statistical independence between target motion and target aspect holds. Under that hypothesis and assuming further statistically decoupled centroid motion in dimensions 1 and 2, the sampling procedure is simplified to independently sampling  $s_n^{(j)} \sim P(s_n | s_{n-1}^{(j)})$ ,  $\mathbf{x}_{n,1}^{(j)} \sim p(\mathbf{x}_{n,1} | \mathbf{x}_{n-1,1}^{(j)})$ , and  $\mathbf{x}_{n,2}^{(j)} \sim p(\mathbf{x}_{n,2} | \mathbf{x}_{n-1,2}^{(j)})$ .

*Resampling Step:* A practical problem associated with the use of sequential importance sampling, as described before, is that only a small number of samples in the particle set may have significant importance weights. As time increases, the distribution of particle weights may get increasingly skewed, resulting in particle degeneracy. One of the early proposed solutions to circumvent this problem [20] was to add a particle selection step that consists of resampling from the original particle set with replacement according to the importance weights. The net result of this selection procedure is to discard low-weight particles, whereas high-weight particles are multiplied. After resampling, the importance weights of the new particle set are then made all equal to  $(1/N_p)$ .

TABLE IV  
AML PARAMETER ESTIMATION ALGORITHM FOR AN  $L \times M$  GMRF

- a) Unnormalized sample correlations:
  - $X_h = \sum_{i=1}^L \sum_{j=1}^{M-1} Y_n(i, j) Y_n(i, j+1)$ .
  - $X_v = \sum_{i=1}^{L-1} \sum_{j=1}^M Y_n(i, j) Y_n(i+1, j)$ .
- b) Unnormalized sample power:
  - $S_y = \sum_{i=1}^L \sum_{j=1}^M Y_n^2(i, j)$ .
- c) Make  $\delta = 10^{-3}$ , compute  $\epsilon = 0.5 - \delta$  and  $\alpha = \frac{(L-1)M}{L(M-1)}$ .
- d) Correlation coefficients estimates:
  - $\widehat{\beta}_{h,n}^c = \frac{\epsilon X_h}{|\chi_v| \cos(\frac{\pi}{L+1}) + \alpha |\chi_h| \cos(\frac{\pi}{M+1})}$ .
  - $\widehat{\beta}_{v,n}^c = \frac{\epsilon X_v}{|\chi_v| \cos(\frac{\pi}{L+1}) + \alpha |\chi_h| \cos(\frac{\pi}{M+1})}$ .
- e) Clutter power estimate:
  - $\widehat{\sigma}_{c,n}^2 = \frac{1}{LM} (S_y - 2 \widehat{\beta}_{h,n}^c * X_h - 2 \widehat{\beta}_{v,n}^c * X_v)$ .

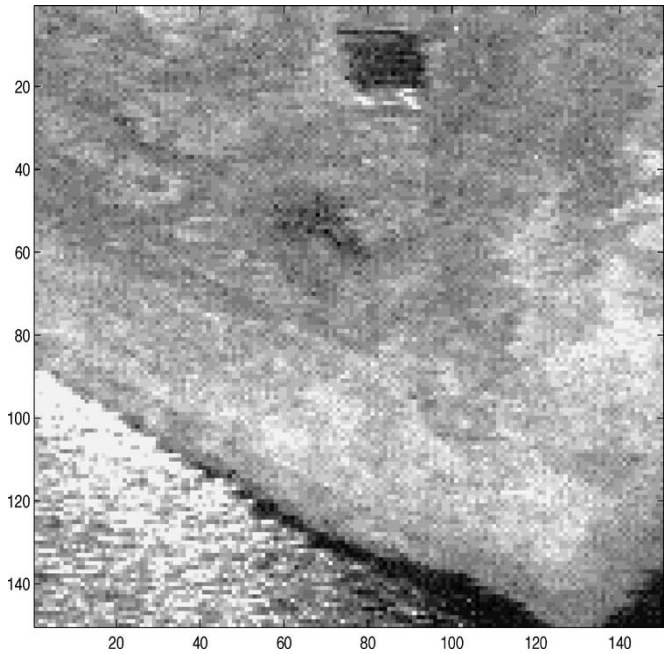


Fig. 1. IRAR intensity data from the Portage database (source: CIS, Johns Hopkins University, Baltimore, MD).

Monte Carlo filters that combine a sequential importance sampling step with a selection step, as described in the previous paragraph, are referred to in the literature as sampling/importance resampling (SIR) filters. Using the GMRF-based likelihood function model from Section II-C, we summarize in Table II the SIR filter for direct, multiaspect target tracking from image sequences in the special case when statistical independence between target motion and target aspect is assumed. The desired minimum mean-square error (MMSE) estimate of the unknown kinematic target state  $\mathbf{x}_n$  is obtained at each frame as the weighted average of the resampled Monte Carlo particles, i.e.,

$$\widehat{\mathbf{x}}_{n|n} \approx \frac{1}{N_p} \sum_{j=1}^{N_p} \mathbf{x}_n^{(j)}. \quad (53)$$

Alternatively, we could use the MAP estimate obtained from the histogram of the particles.



*Computational Complexity:* From a computational point of view, the main advantage of a particle filter solution over the grid-based algorithms described in Section III is that, whereas the grid-based trackers evaluate the likelihood function in the entire discretized state space, the particle filter algorithm requires the evaluation of the likelihood function only for each sample in the current set of particles. For a fixed number of particles equal to  $N_p$  and using a GMrf-based likelihood function as in Section II-C, that translates into a computational cost in terms of required floating-point multiplications of order  $O(\nu N_p)$ , where  $\nu \approx 3 + (r_i + r_s + 1)(l_i + l_s + 1)$ . By contrast, the grid-based HMM algorithm as implemented in this paper has a computational cost, assuming an  $L \times L$  image grid, of order  $O(K\nu L^2)$ , where  $K$  (see Section II-B) is the number of target aspect states. If  $N_p \ll L^2$  and  $K$  is large, the computational savings may be considerable.

### B. Auxiliary Particle Filter Tracker

The SIR tracker presented previously samples blindly from the motion and aspect priors without taking into consideration the current available frame. That may result in a large number of particles with negligible weight, particularly if the likelihood function is highly peaked. A possible alternative to enhance performance is to use auxiliary particle filtering (APF) [22]. The intuitive idea in APF is to preselect a set of particles at step  $n - 1$ , taking advantage of the information conveyed by the current observation  $\mathbf{y}_n$ . Assuming decoupled motion in the two dimensions of the plane and statistical independence between target motion and target aspect that is formally accomplished in our problem by introducing a preselection step where we first draw samples  $\underline{\mu}_{n,1}^{(j)}$ ,  $\underline{\mu}_{n,2}^{(j)}$ , and  $\mu_{n,a}^{(j)}$ , respectively, from  $p(\mathbf{x}_{n,1} | \mathbf{x}_{n-1,1}^{(j)})$ ,  $p(\mathbf{x}_{n,2} | \mathbf{x}_{n-1,2}^{(j)})$ , and  $P(s_n | s_{n-1}^{(j)})$ . We then compute the first-stage importance weights

$$\Lambda_n^{(j)} \propto w_{n-1}^{(j)} p(\mathbf{y}_n | \underline{\mu}_{n,1}^{(j)}, \underline{\mu}_{n,2}^{(j)}, \mu_{n,a}^{(j)}) \quad (54)$$

and sample auxiliary indices  $k^{(j)}$  from the set  $\{1, 2, \dots, N_p\}$  with  $P(k^{(j)} = i) = \Lambda_n^{(i)}$ . Finally, we sample the next set of particles  $\mathbf{x}_{n,1}^{(j)}$ ,  $\mathbf{x}_{n,2}^{(j)}$ , and  $s_n^{(j)}$  from  $p(\mathbf{x}_{n,1} | \mathbf{x}_{n-1,1}^{(k^{(j)})})$ ,  $p(\mathbf{x}_{n,2} | \mathbf{x}_{n-1,2}^{(k^{(j)})})$  and  $P(s_n | s_{n-1}^{(k^{(j)})})$ . The second-stage importance weights are updated, taking into consideration both the preselection and the importance sampling steps by making

$$w_n^{(j)} \propto \frac{p(\mathbf{y}_n | \mathbf{x}_{n,1}^{(j)}, \mathbf{x}_{n,2}^{(j)}, s_n^{(j)})}{p(\mathbf{y}_n | \underline{\mu}_{n,1}^{(k^{(j)})}, \underline{\mu}_{n,2}^{(k^{(j)})}, \mu_{n,a}^{(k^{(j)})})}. \quad (55)$$

In (54) and (55), we use the GMrf-based likelihood function model from Section II-B [see (16), (19), and (20)] to compute the importance weights up to a proportionality constant.

Table III summarizes the APF algorithm used by our tracker. The initialization step is identical to that in Table II and is

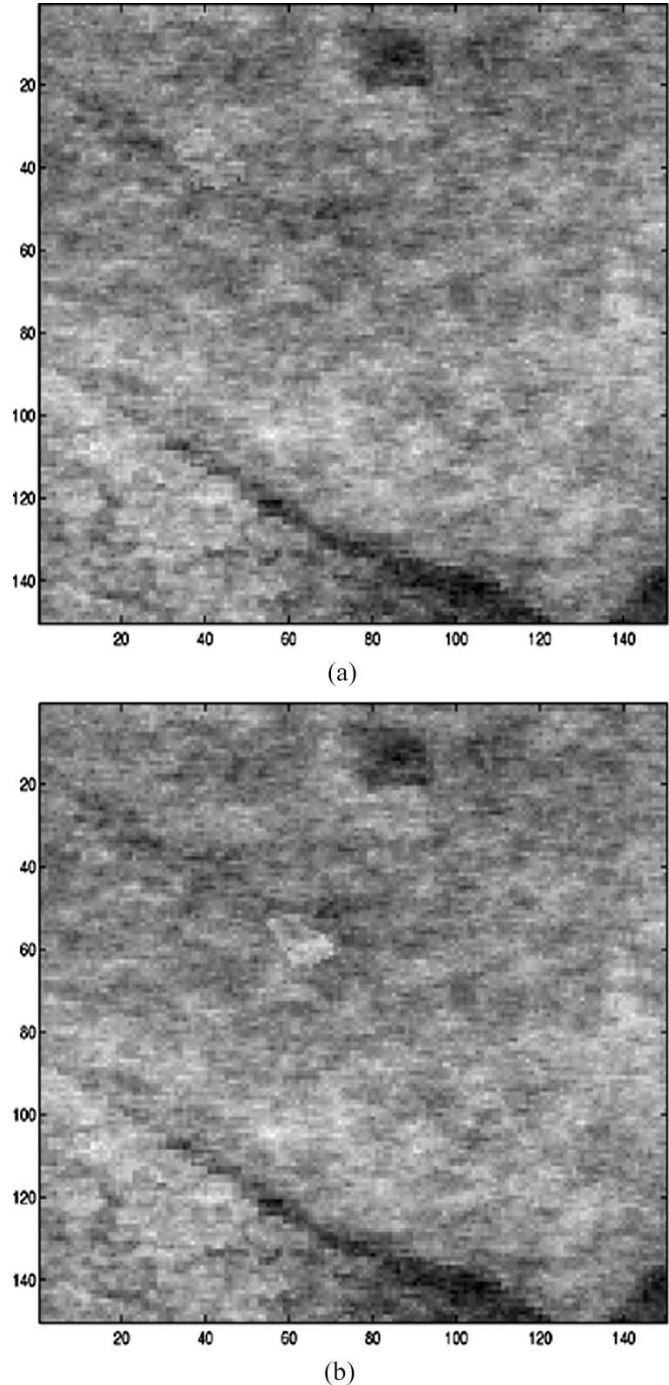


Fig. 2. Cluttered target sequence, PTCR = 3.6 dB. (a) First frame. (b) Tenth frame with random target translation, rotation, scaling, and shearing.

omitted accordingly for conciseness. The final Monte Carlo approximation of the MMSE estimate of the unknown kinematic state  $\mathbf{x}_n = [\mathbf{x}_{n,1}^T \ \mathbf{x}_{n,2}^T]^T$  is given then by

$$\hat{\mathbf{x}}_{n|n} \approx \sum_{j=1}^{N_p} w_n^{(j)} \mathbf{x}_n^{(j)} \quad (56)$$

with the weights  $w_n^{(j)}$  normalized to add up to 1. As noted before, if necessary, statistical dependence between target motion and target aspect could be incorporated into the importance sampling step in Table III.

## V. CLUTTER ADAPTIVE DETECTION/TRACKING

The algorithms in Sections III and IV assumed perfect knowledge of the parameters of the GMrf background clutter model. However, in practical scenarios with real clutter data, the clutter model parameters are either unknown *a priori* or vary from frame to frame and must be therefore adaptively learned from the sensor images. In this section, we discuss clutter parameter estimation.

### A. Clutter Parameter Adaptation

In practical situations, the residual clutter after the removal of the spatially variant local means in each frame is not perfectly modeled by a Gauss–Markov random field. Nevertheless, the highly parameterized and sparse nature of a GMrf's inverse covariance matrix [27] still makes this model attractive from a computational point of view by allowing us to avoid multiplications and inversions of large matrices in the computation of the likelihood function. In this paper, we propose to fit a GMrf model as in (14) to the residual clutter data at each frame  $n$  by first assuming that the model holds and, then, by estimating the most likely corresponding clutter model parameters  $\beta_{h,n}^c, \beta_{v,n}^c, \sigma_{c,n}^2$  for each observed frame.

Ideally, the time-varying clutter parameters should be jointly estimated with the hidden state variables  $\mathbf{x}_n$  and  $s_n$  in a nonlinear Bayesian framework. For computational simplicity, though, in this paper, we use 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 [26]. The suboptimal estimates  $\hat{\beta}_{h,n}^c, \hat{\beta}_{v,n}^c$ , and  $\hat{\sigma}_{c,n}^2$  are then plugged into (16), (19), and (20) to evaluate the likelihood function for all states in the  $\tilde{\mathcal{L}} \times \mathcal{I}$  lattice in the case of the grid-based algorithms from Section III or, for all particles in the set  $\{\mathbf{x}_n^{(j)}, s_n^{(j)}\}, 1 \leq j \leq N_p$ , in the case of the Monte Carlo filters in Section IV.

Assuming that the target is small compared with the background, we summarize, in Table IV, the AML parameter estimation algorithm given an  $L \times M$   $n$ th observed sensor frame  $\mathbf{Y}_n$ ; see [26] and [28] for further details. We are reminded that, like elsewhere in this paper, the input frame  $\mathbf{Y}_n$  used in the routine in Table IV corresponds to the raw observed image *after suitable preprocessing* to remove the local means.

## VI. PERFORMANCE ANALYSIS

In this section, we investigate the tracking performance of the algorithms proposed in Sections III and IV. In our performance studies, we use simulated image sequences that are generated from real-world infrared laser radar (IRAR) intensity data. The real IRAR data, which were obtained from the Center for Imaging Science at Johns Hopkins University, were collected by an airborne laser imaging sensor [29] flying over the Portage region. Fig. 1 shows an imaged scene where we see a stretch of

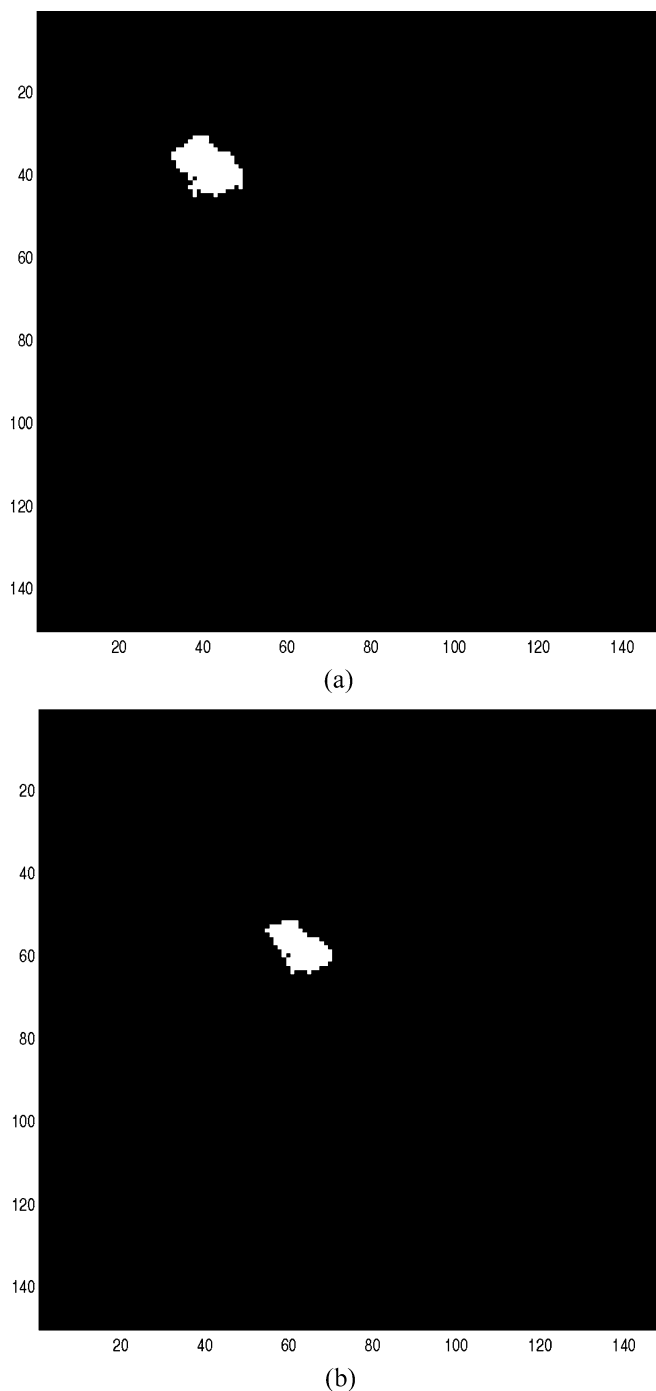


Fig. 3. Clutter-free target templates for Figs. 2(a) and (b) respectively, shown as binary images.

terrain next to a body of water on the bottom left corner of the image. Brighter areas indicate stronger laser returns.

In order to simulate a moving target sequence, we extracted the spatially variant local mean along the image and fitted a first-order GMrf model to the background by estimating the model parameters using the AML estimator. The background clutter movie is then generated by adding to the matrix of previously stored local means a sequence of synthesized random GMrf samples that are generated using the estimated background clutter parameters and the GMrf simulation algorithm in [25].

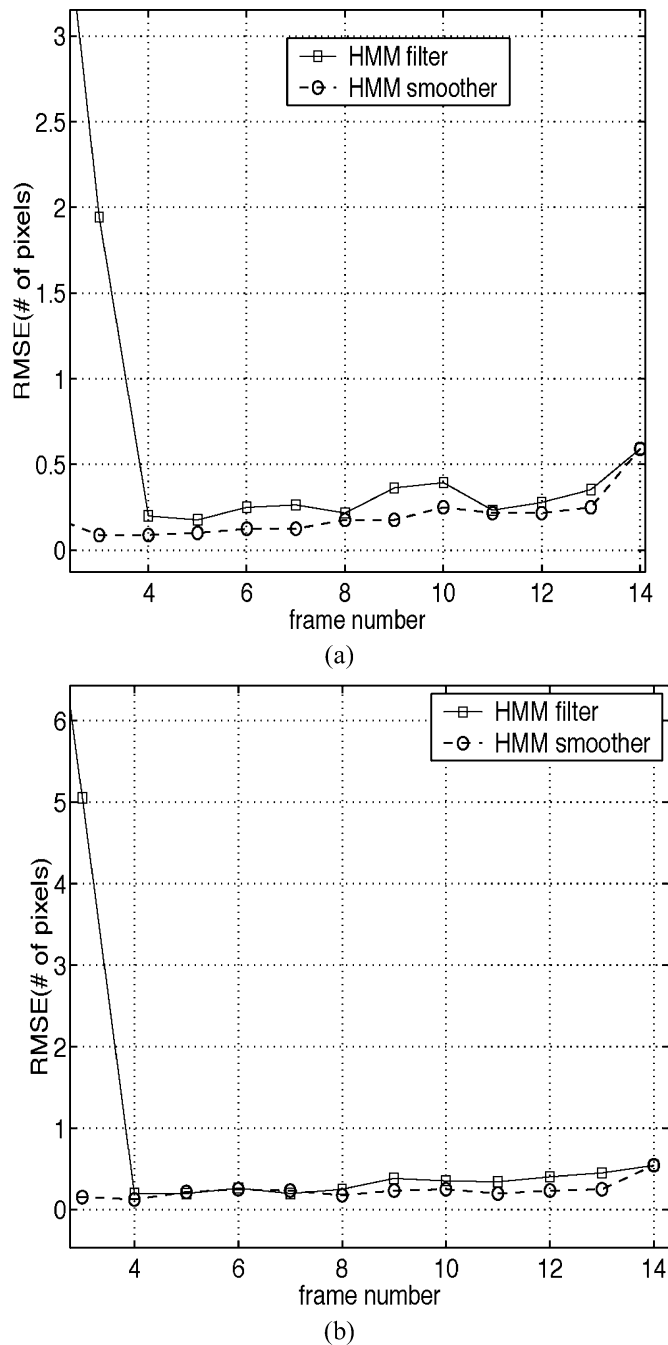


Fig. 4. Position estimate RMSE for the HMM filter (solid) and HMM smoother (dashed)  $\text{PTCR} = -3.6$  dB. (a) Vertical coordinate. (b) Horizontal coordinate.

**Target Simulation:** To simulate the target, we took an artificial template representing a military vehicle and generated a library of transformations of that model 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 a linear white noise acceleration model [4]. The white noise acceleration motion model is a particular case of the model in (5) (see Section II-A), where we make

$$\mathbf{x}_{n+1,i} = \mathbf{F}\mathbf{x}_{n,i} + \mathbf{u}_{n,i}, \quad i = 1, 2 \quad (57)$$

with

$$\mathbf{F} = \begin{bmatrix} 1 & \Delta \\ 0 & 1 \end{bmatrix} \quad (58)$$

and

$$E[\mathbf{u}_{n,i} \mathbf{u}_{l,j}^T] = q \underbrace{\begin{bmatrix} \frac{\Delta^3}{3} & \frac{\Delta^2}{2} \\ \frac{\Delta^2}{2} & \Delta \end{bmatrix}}_{\mathbf{Q}} \delta_{n-l,i-j} \quad (59)$$

for  $i, j = \{1, 2\}$ ,  $n \geq 0$ , and  $l \geq 0$  with  $E[\mathbf{u}_{n,1}] = E[\mathbf{u}_{n,2}] = \mathbf{0}$ ,  $\forall n \geq 0$ . For the target simulations in this paper, we set  $\Delta = 0.04$  and  $q = 8$  in (58) and (59). Note that the particular choice of a linear motion model in (57) does not turn the target tracking problem into a linear/Gaussian stochastic filtering problem since the nonlinearity in the observation model (7) will still cause the posterior pdf of the hidden target state to be non-Gaussian. The model choice in (57) does not imply, therefore, any loss of generality.

In order to simulate the target's aspect dynamics, we initialized the target template state  $s_0$  with a random, unknown selection 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 trackers. 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. Figs. 2(a) and (b) show two simulated frames, respectively, at instants  $n = 0$  and  $n = 9$ , with peak target-to-clutter ratio (PTCR) equal to 3.6 dB. The corresponding clutter-free target templates are shown as binary images in Figs. 3(a) and (b), respectively.

**Performance of the Grid-Based Detector/Trackers:** We examine first the performance of the grid-based algorithms in Section III. The simulated target starts from an *unknown* initial position that is assumed to be uniformly distributed over the entire image grid. The target is subsequently detected and tracked over 14 frames using 1) the on-line HMM filter and 2) the batch HMM smoother. A total of five different template models were used for each block of 14 simulated frames. The peak target-to-clutter ratio (PTCR) was lowered to  $-3.6$  dB in the simulated image sequence, and the initial target velocity was set at 10 m/s in each dimension. At each frame, the HMM algorithms first make a hard detection decision on the presence or absence of a target and then, if a target is declared present, estimate the spatial location of the target's centroid. In the implementation of the grid-based trackers, we approximate the true continuous-valued motion model in (57) by a discrete-valued state model consisting of a deterministic drift of 2 pixels/frame in each dimension superimposed to a 2-D random walk model with probability of fluctuation of one pixel in either dimension equal to 15%. The background clutter parameters are adaptively estimated using the AML algorithm.

The zoomed-in plots in Figs. 4(a) and (b) show the root mean-square error (RMSE) in number of pixels of the estimates of the centroid position, respectively, in the vertical and horizontal

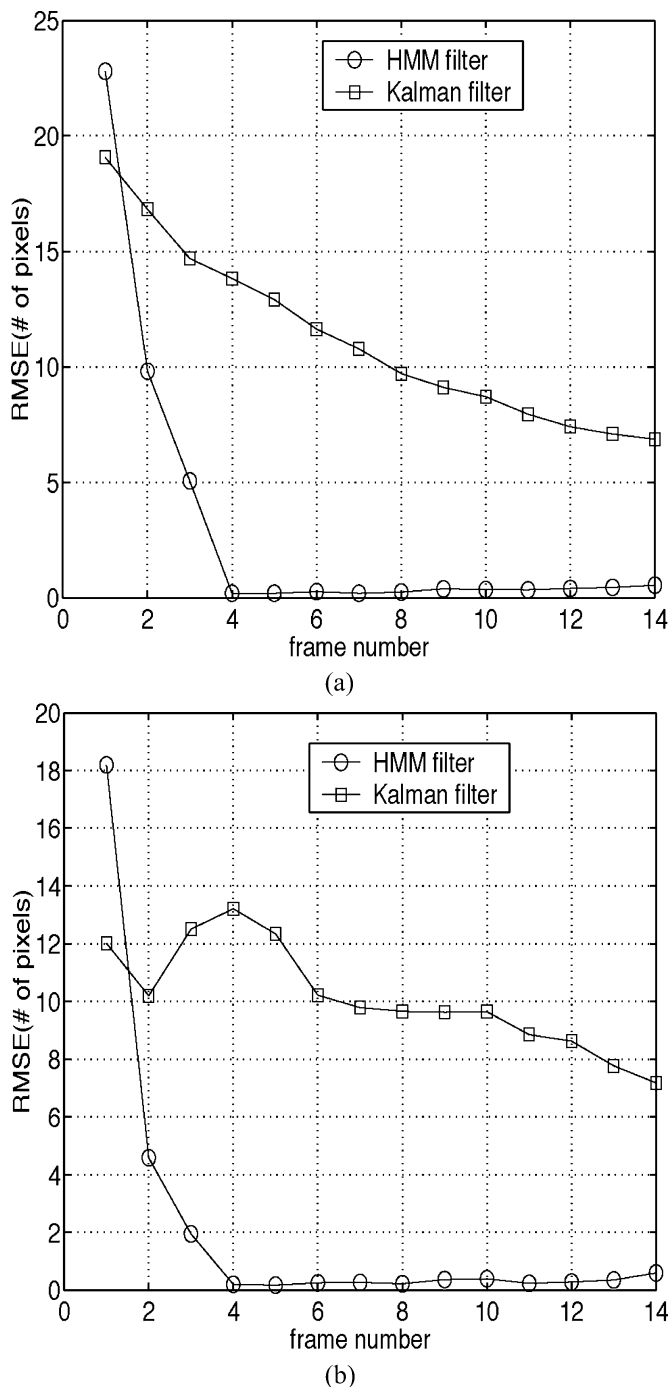


Fig. 5. RMSE for the HMM filter and Kalman filter.  $\text{PTCR} = -3.6$  dB. (a) Horizontal coordinate. (b) Vertical coordinate.

dimensions in Fig. 2. The curves in Figs. 4(a) and (b) were obtained from 130 Monte Carlo runs (i.e., 1820 simulated image frames). We note from the plots in Fig. 4 that the online HMM filter has an initial error, probably due to its poor initialization with a uniform (noninformative) prior. However, as additional data frames are processed, the algorithm quickly acquires the target and eventually reaches a low steady-state localization error. In turn, the batch HMM smoother takes advantage of both past and future information in the data volume to attenuate the large initial errors of the HMM filter, thus achieving near-perfect tracking for this particular level of peak target-to-clutter ratio.

Note also that both the HMM filter and the HMM smoother had no misses (i.e., present targets that are incorrectly declared absent) over all 1820 simulated frames.

For comparison purposes, using the same level of PTCR ( $-3.6$  dB), we plot in Figs. 5(a) and (b) the RMSE of the horizontal and vertical coordinate estimates obtained by the online HMM filter and by the suboptimal association of a bank of correlation filters (each one matched to one possible target template) and a linear Kalman–Bucy filter (KBf) used for sequential tracking. The plots in Fig. 5 show that the correlator/KBf association performs poorly in low levels of PTCR, exhibiting a very long target acquisition time in contrast with the quick acquisition and low steady-state localization for the nonlinear HMM filter.

*Particle Filter Tracking Performance—Multiaspect Target:* We examine next the performance of the *multiaspect*, *mixed-state* SIR, and APF trackers introduced in Section IV. The target simulation parameters are identical to those used for the grid-based trackers, except that instead of assuming that initial target position is uniformly distributed over the entire image grid, we assume that the vertical initial target centroid position is uniformly distributed between pixels 20 and 60 within the image resolution. Similarly, we assume that the horizontal initial target centroid position is uniformly distributed between pixels 20 and 40. We use 5000 particles and track the object of interest over 13 consecutive frames. For a  $150 \times 150$  image and assuming five target template states and a maximum template size equal to  $20 \times 20$ , the particle filter trackers with 5000 particles have a computational cost  $O(2015000)$  in number of floating-point multiplications per frame compared with an approximate cost of order  $O(45337500)$  for the raw implementation of the HMM trackers.

Figs. 6(a) and (b) show the position estimate RMSE measured in number of pixels for the SIR tracker (solid line) and for the APF tracker (dashed line), respectively, in the vertical and horizontal dimension. Setting  $q = 8$  (very noisy trajectory) and  $\text{PTCR} = -3.6$  dB (invisible target), the SIR tracker diverged (i.e., failed to acquire the correct track within 13 frames) in eight out of 135 Monte Carlo runs. The APF tracker, on the other hand, diverged in seven out of 144 Monte Carlo runs. The error curves in Figs. 6(a) and (b) were computed excluding the diverging tracks from the average.

We see from the plots in Fig. 6 that the particle filters also have a relatively high initial position estimation error, but roughly 95% of the time, they quickly acquire the target after a few frames and converge to a low RMS steady-state error. When the target is acquired, the APF tracker has a slightly better performance than the SIR algorithm, as reflected in a smoother error curve in Figs. 6(a) and (b). However, it was impossible in this simulation to detect any statistically meaningful improvement in terms of number of lost tracks when we compare the APF algorithm to the SIR tracker. We are currently investigating possible enhancements to the particle filter trackers to reduce the percentage of lost tracks without increasing the number of particles. A possible alternative is to introduce an additional Monte Carlo Markov chain (MCMC) move step to steer the particles back to the “correct” region of

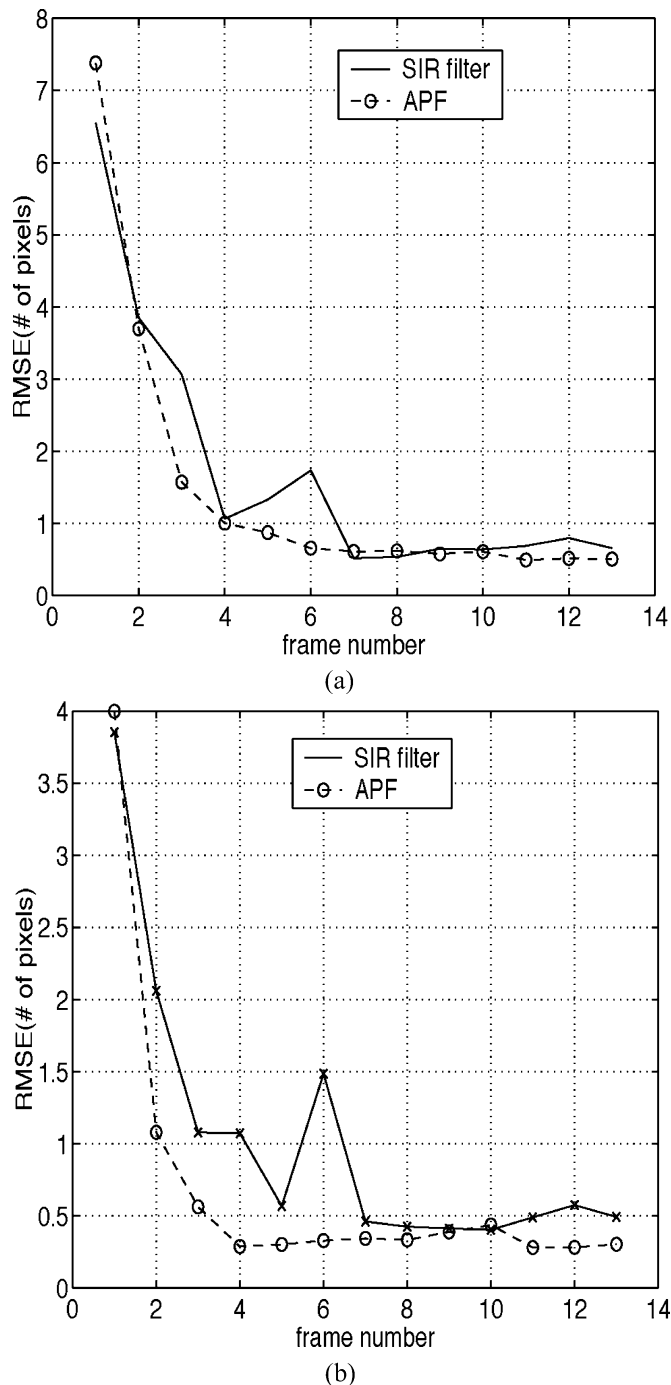


Fig. 6. Position estimate RMSE for the SIR filter (solid) and the auxiliary particle filter (dashed)  $\text{PTCR} = -3.6$  dB. (a) Vertical coordinate. (b) Horizontal coordinate.

the state space once a track has been lost; see, e.g., [38] or, for an unrelated application, [39].

## VII. CONCLUSIONS

In this paper, we discussed different Bayesian algorithms for direct target tracking in image sequences in a situation of random target aspect and unknown clutter parameters. We modeled the changes in target aspect from frame to frame using a discrete-valued hidden Markov chain defined on the space

of possible target templates. The target's translational motion was modeled either as a separate 2-D discrete-valued Markov chain defined on the image grid or as a continuous-valued (possibly nonlinear) first-order Markov dynamic model. The spatial correlation of the clutter background at each frame was represented by a 2-D noncausal GMrf model.

For the discrete-valued target motion model, we integrated detection and tracking by augmenting the 2-D motion HMM with an additional dummy state that represents the absence of a target at the current frame. We introduced a multiframe Bayesian smoother that, given a sequence of observed (cluttered) frames  $\{\mathbf{Y}_0, \mathbf{Y}_2, \dots, \mathbf{Y}_N\}$ , decides at each frame  $n$ ,  $1 \leq n \leq N$  whether a target is present or absent and, if the target is declared present, jointly estimates its position and aspect. We also introduced an alternative online multiframe HMM filter that detects the presence of a target at frame  $n$  and estimates its position and aspect based only on the current and past observed frames.

The integrated detection/tracking framework for grid-based algorithms allows us to propagate the multiframe posterior probability of absence or presence of target together with the multiframe posteriors for the different target position and aspect states. That methodology has obvious advantages in heavily cluttered image sequences where a target of interest shows poor visibility in any given individual frame but may become more visible when seen moving across several different frames. In fact, performance studies with simulated image sequences generated from real IRAR data show that in a scenario of low target-to-clutter ratio and unknown, time-variant target aspect, the HMM detector/tracker had no misses (i.e. no present targets incorrectly declared absent) over 1820 simulated image frames. The HMM smoother and filter also greatly outperformed the association of a bank of multispect image correlators and a Kalman filter tracker in terms of target acquisition time.

Despite their good performance, the use of the proposed HMM trackers is limited, however, by their computational complexity, particularly if the size of the image and the number of possible target models increase. To circumvent this problem, we proposed alternative mixed-state particle filter trackers that assume a continuous-valued motion state and a discrete-valued aspect state. Two versions of the particle filter tracker were introduced: a sampling/importance resampling (SIR) filter and an alternative auxiliary particle filter algorithm. In both algorithms, we track the target directly from the image sequence incorporating the target signature and clutter spatial correlation models into the likelihood function.

The particle filter algorithms are flexible enough to account for nonlinear target motion dynamics and statistical dependence between target motion and target aspect. Moreover, using a limited number of particles, the particle filter trackers have a reduction of over one order of magnitude in computational cost when compared with the grid-based HMM filter. The drawback associated with a particle filter tracker is, however that, for a small number of particles, the algorithm may fail to converge to the true state track. Nevertheless, we observed in our simulations that despite heavily cluttered observations, both the SIR and the APF trackers quickly converged to the true track roughly 95%

of the time. The APF tracker seems to have slightly superior performance to the SIR solution, as reflected in an improved RMSE curve. Excluding those rare occasions when the true track was lost, the steady-state RMS localization error for both the SIR and APF trackers was roughly comparable with that of the HMM filter after the same number of image frames were processed.

#### ACKNOWLEDGMENT

The author would like to thank Prof. José M. F. Moura for useful discussions on HMM-based target tracking.

#### REFERENCES

- [1] D. Casasent and S. Ashizawa, "Synthetic aperture radar detection, recognition, and clutter rejection with new minimum noise and correlation energy filters," *Opt. Eng.*, vol. 36, pp. 2729–2736, Oct. 1997.
- [2] D. Casasent and R. Shenoy, "Synthetic aperture radar detection and clutter rejection MINACE filters," *Patt. Recogn.*, vol. 30, no. 1, pp. 151–161, 1997.
- [3] G. Ravichandran and D. Casasent, "Minimum noise and correlation energy filter," *Applied Opt.*, vol. 31, pp. 1823–1833, Apr. 1992.
- [4] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs, CT: YBS, 1995.
- [5] E. Oron, A. K. Kumar, and Y. Bar-Shalom, "Precision tracking with segmentation for imaging sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 977–987, July 1993.
- [6] M. G. S. Bruno and J. M. F. Moura, "Multiframe detection/tracking in clutter: optimal performance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, pp. 925–946, July 2001.
- [7] P. K. Bhaharadwaj, P. Runkle, L. Carin, J. Berrie, and J. Hughes, "Multiaspect classification of airborne targets via physics-based HMM's and matching pursuits," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, pp. 595–606, Apr. 2001.
- [8] P. K. Bharadwaj, P. Runkle, and L. Carin, "Target identification with wave-based matching pursuits and hidden Markov models," *IEEE Trans. Antennas Propagat.*, vol. 47, pp. 1543–1554, Oct. 1999.
- [9] P. Runkle, L. Carin, L. Couchman, T. J. Yoder, and J. A. Bucaro, "Multi-aspect identification of submerged elastic targets via wave-based matching pursuits and hidden Markov models," *J. Acoust. Soc. Amer.*, no. 104, pp. 937–946, Aug. 1999.
- [10] P. Runkle, P. K. Bharadwaj, L. Couchman, and L. Carin, "Hidden Markov models for multi-aspect target classification," *IEEE Trans. Signal Processing*, vol. 47, pp. 2035–2040, July 1999.
- [11] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–286, Feb. 1989.
- [12] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Stat.*, vol. 37, pp. 1554–1563, 1966.
- [13] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, no. 1, pp. 164–171, 1970.
- [14] Y. Barniv, "Dynamic programming solution for detecting dim moving targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 21, pp. 144–156, Jan. 1985.
- [15] 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, Eds. New York: Springer-Verlag, 2001.
- [16] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, pp. 174–188, Feb. 2002.
- [17] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian, nonlinear state space models," *J. Comput. Graphical Statist.*, vol. 5, no. 1, pp. 1–25, 1996.
- [18] M. Isard and A. Blake, "Condensation—conditional density propagation for visual tracking," *Int. J. Comput. Vision*, vol. 28, no. 1, pp. 5–28, 1998.
- [19] —, "A mixed-state condensation tracker with automatic model switching," in *Proc. Sixth Int. Conf. Comput. Vision*, 1998, pp. 107–112.
- [20] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Proc. Inst. Elect. F*, vol. 140, no. 2, pp. 107–113, 1993.
- [21] N. J. Gordon, D. J. Salmond, and C. Ewing, "Bayesian state estimation for tracking and guidance using the bootstrap filter," *J. Guidance, Contr., Dyn.*, vol. 18, no. 6, pp. 1434–1443, 1995.
- [22] M. K. Pitt and N. Shephard, "Filtering via simulation: auxiliary particle filters," *J. Amer. Statist. Assoc.*, vol. 94, no. 446, pp. 590–599, 1999.
- [23] M. G. S. Bruno, "Sequential importance sampling filtering for target tracking in image sequences," *IEEE Signal Processing Lett.*, vol. 10, pp. 246–249, Aug. 2003.
- [24] J. Y. Chen and I. S. Reed, "A detection algorithm for optical targets in clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 23, pp. 46–59, Jan. 1987.
- [25] J. M. F. Moura and N. Balram, "Recursive structure of noncausal Gauss Markov random fields," *IEEE Trans. Inform. Theory*, vol. 38, pp. 334–354, Mar. 1992.
- [26] —, "Noncausal Gauss-Markov random fields: parameter structure and estimation," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1333–1355, July 1993.
- [27] J. M. F. Moura and M. G. S. Bruno, "DCT/DST and Gauss-Markov fields: conditions for equivalence," *IEEE Trans. Signal Processing*, vol. 46, pp. 2571–2574, Sept. 1998.
- [28] S. M. Schweitzer and J. M. F. Moura, "Hyperspectral imagery: clutter adaptation in anomaly detection," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1855–1871, Aug. 2000.
- [29] J. K. Bounds, "The Infrared Airborne Radar Sensor Suite," Cambridge, MA, RLE Tech. Rep. no. 610, Dec. 1996.
- [30] M. G. S. Bruno, "Joint detection and tracking of moving targets in clutter," Ph.D. dissertation, Elect. Comput. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA, 1998.
- [31] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc., ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
- [32] S. Thurn, W. Bugard, and D. Fox, "A probabilistic approach to concurrent mapping and localization for mobile robots," *Machine Learning*, vol. 31, pp. 29–53, 1998.
- [33] A. Blake, B. Basclé, M. Isard, and J. MacCormick, "Statistical models of visual shape and motion," *Philos. Trans. R. Soc. A*, vol. 356, pp. 1283–1302, 1998.
- [34] B. Li and R. Chellapa, "Simultaneous tracking and verification via sequential Monte Carlo methods," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2000.
- [35] N. Bergman, L. Ljung, and F. Gustafsson, "Terrain navigation using Bayesian statistics," *IEEE Contr. Syst. Mag.*, vol. 19, pp. 33–40, Mar. 1999.
- [36] T. Semerdjiev, V. Jilkov, and D. Angelove, "Target tracking using Monte Carlo simulation," *Math. Comput. Simulation*, vol. 47, pp. 441–447, 1998.
- [37] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Trans. Signal Processing*, vol. 50, pp. 736–746, Mar. 2002.
- [38] 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.
- [39] C. Andrieu, N. de Freitas, and A. Doucet, "Sequential MCMC for Bayesian model selection," in *Proc. IEEE Higher Order Statistics Workshop*, Caesarea, Israel, 1999, pp. 130–134.

**Marcelo G. S. Bruno** (M'00) received the bachelor's and master's degrees in electrical engineering from the University of São Paulo, São Paulo, Brazil, respectively, in 1992 and 1993, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 1998.

From January 1999 until September 2001, he was a postdoctoral fellow at the Laboratory of Communications and Signal Processing, Polytechnic School, University of São Paulo. Since October 2001, he has been with the Electrical Engineering Division, Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, Brazil, where he is an Assistant Professor. He was also a visiting research scholar with the Electrical and Computer Engineering Department, Carnegie Mellon University, in the summers of 1999 and 2001 and in the winter of 2002. His research interests are in statistical signal/image processing, particularly Markov random fields (Mrfs), hidden Markov models (HMMs), particle filters/sequential Monte Carlo methods, Markov Chain Monte Carlo (MCMC), and their applications in target tracking, image processing, computer vision, mobile robotics, and telecommunications. He is also interested in distributed detection and optimal sensor fusion/management in sensor networks.

Dr. Bruno currently serves as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.