BAYESIAN SMOOTHING AND FILTERING FOR MULTIFRAME, MULTIASPECT TARGET DETECTION AND TRACKING

Marcelo G. S. Bruno

Instituto Tecnológico de Aeronáutica
São José dos Campos SP 12228-900, Brazil
ph:(55-12) 3947-5878; email:bruno@ele.ita.br

José M. F. Moura

ECE Department, Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213, USA
ph: (412) 268-6341; email: moura@ece.cmu.edu

ABSTRACT

We introduce in this paper a new Bayesian algorithm for joint multiframe detection and tracking of multiaspect targets that move randomly in cluttered digital image sequences. Two versions of the algorithm are derived: a batch Bayes smoother and an on-line Bayes filter. Performance results with a simulated image sequence generated from real infrared airborne radar (IRAR) data show an improvement over the association of a bank of correlation detectors and a Kalman-Bucy tracker in a scenario with a heavily cluttered multiaspect target.

1. INTRODUCTION

We present in this paper a new Bayesian algorithm for integrated, multiframe detection and tracking of multiaspect moving targets in sequences of two-dimensional (2D) cluttered images that are generated by a remote airborne sensor. Random changes in the aspect of the target of interest may result from rotational motion and/or from changes in the conditions of observation of the target due e.g. to variations in the relative target-sensor orientation. Previous literature on aspect-invariant target detection in sensor images, see e.g. [1], 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, e.g. [2], is in turn based on a suboptimal decoupling of the detection and tracking tasks: a preliminary single frame detection stage (typically a correlation detector) generates initial estimates of the target's true position which are subsequently associated to a multiframe linear tracker, generally a variation of the Kalman-Bucy filter. In our work, we propose a different approach where a Bayesian methodology is used to integrate detection and tracking into a single framework using a recursive spatio-temporal algorithm that fully incorporates the dynamical models for target motion and aspect as well as the statistical model for the spatially correlated clutter background.

Bayesian algorithms based on sequential Monte Carlo methods have been successfully applied to shape and motion tracking in digital images, appearing in the computer vision literature under the generic name of condensation algorithms [3]. In this paper, we use an alternative modeling approach to solve the problem of joint Bayesian detection and tracking in remote sensing images. Instead of defining the unknown target position on a continuous-valued space, we take advantage of the sensor's finite resolution to model the target's centroid motion as a two-dimensional hidden Markov model (HMM) defined on the sensor's finite resolution grid. Detection and tracking are easily integrated then by adding a dummy absent target state to the HMM. Similarly, we assume that the target of interest has a finite number of possible aspect states and use a second HMM defined on the discrete aspect state space to model the dynamical aspect changes from frame to frame. This approach to aspect change modeling is similar to the work for target classification in [4] with the difference that, in [4], the targets of interest are stationary and the data are the electromagnetic scattered waveforms of the targets rather than pre-processed cluttered target images generated by an imaging sensor.

This paper is divided into 6 sections. Section 1 is this introduction. Section 2 briefly reviews the observation and clutter models. In section 3, we present a new Bayes smoother that recursively computes the joint posterior distribution of the target's hidden position and aspect state at each frame conditioned on all past, present, and future frames in a given data volume. In section 4, we introduce an alternative on-line Bayes filter that generates sequential estimates of the target's hidden position based on the past and present frames only. In section 5, we compare the tracking performance of the Bayes smoother and filter to the performance of the suboptimal association of a bank of correlation filters and a Kalman-Bucy filter. The performance studies are carried out using a simulated multiaspect target which is added to a sequence generated from real clutter infrared airbone radar (IRAR) data. The clutter model parameters are adaptively learned from the observed data. Finally, section 6 summarizes the contributions of the paper.

2. OBSERVATION MODEL

The raw sensor measurements at instant n are sampled and processed to form a digital sensor image represented by the

The first author's work was partially supported by FAPESP, Brazil.

$L \times M$ matrix

$$\mathbf{Y}_n = \mathbf{F}(z_n, s_n) + \mathbf{V}_n . \quad (1)$$

In (1), $\mathbf{F}(z_n, s_n)$ is the clutter-free target image which is a function of the target's unknown (hidden) centroid position, z_n, and of the target's unknown (hidden) aspect, s_n. The matrix $\mathbf{V}_n$ represents the background clutter.

The centroid position z_n of a target that is present at frame n is defined on a finite *centroid lattice* $\overline{\mathcal{L}}$ whose sites are either one pixel in the sensor's $L \times M$ finite resolution image, or a pixel in the vicinity of the image's borders, see [5] for details. To integrate detection and tracking, we augment the centroid lattice with a dummy absent target state denoted $z_n = L_1$ in this paper. We refer to the union of $\overline{\mathcal{L}}$ and the absent target state as the *extended lattice*, $\widetilde{\mathcal{L}}$. Similarly, we define the unknown (hidden) aspect at frame n, s_n, on a finite discrete aspect set $\mathcal{I} = \{0, 1, \ldots, J-1\}$ where each element in $\mathcal{I}$ is an index linking to one of J possible target templates. For each $z_n \in \overline{\mathcal{L}}$, the clutter-free target image model $\mathbf{F}$ in (1) returns an aspect-dependent spatial distribution of intensities which is centered on the location z_n. For $z_n = L_1$, function $\mathbf{F}$ returns a null image, indicating that the target is absent from the scene.

Clutter Model We capture the 2D spatial correlation of the background clutter using a noncausal, spatially homogeneous Gauss-Markov random field (GMrf) model [6]. The clutter returns at frame n, $V_n(i, j)$, $1 \le i \le L, 1 \le j \le M$, are described by the 2D finite difference equation

$$\begin{aligned} V_n(i, j) &= \beta_v^c\,[V_n(i-1, j) + V_n(i+1, j)] \\ &+ \beta_h^c\,[V_n(i, j-1) + V_n(i, j+1)] + U_n(i, j) \quad (2) \end{aligned}$$

where $E\,[V_n(i, j)\,U_n(p, r)] = \sigma_c^2\,\delta(i-p, j-r)$. The assumption of zero-mean clutter implies a pre-processing of the data that subtracts the mean of the background.

3. BAYES SMOOTHER

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

$$\begin{aligned} \alpha_n(i, j) &= p(z_n = i,\, s_n = j, \mathbf{Y}_0^n) \quad (3) \\ \beta_n(i, j) &= p(\mathbf{Y}_{n+1}^N \mid z_n = i,\, s_n = j) \quad (4) \\ S_n(i, j) &= p(\mathbf{y}_n \mid z_n = i,\, s_n = j) \quad (5) \end{aligned}$$

for $i \in \widetilde{\mathcal{L}}$ and $j \in \mathcal{I}$. We derive in the sequel an algorithm for the recursive computation of $P(z_n = i,\, s_n = j \mid \mathbf{Y}_0^N)$ using the matrices defined in (3)-(5). We make the following assumptions in the derivation:

- The sequence of clutter frames $\{\mathbf{V}_k\}$, $k \ge 0$, is independent indentically distributed (i.i.d.) and also statistically independent of the sequences of target centroid positions, $\{z_k\}$, and target aspect states, $\{s_k\}$, $k \ge 0$.
- The sequences $\{z_k\}$ and $\{s_k\}$ $k \ge 0$, are *mutually independent* first-order discrete Markov processes described respectively by the transition probability matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ such that

$$\begin{aligned} T_1(i,j) &= P(z_n = i \mid z_{n-1} = j) \qquad (i,j) \in \widetilde{\mathcal{L}} \times \widetilde{\mathcal{L}}\,, \\ T_2(l,r) &= P(s_n = l \mid s_{n-1} = r) \qquad (l, r) \in \mathcal{I} \times \mathcal{I}\,. \end{aligned}$$

Forward Recursion From the memoryless observation model in (1) and using the previous assumptions on the sequences $\{z_k\}$, $\{s_k\}$ and $\{\mathbf{V}_k\}$, we use Bayes' law and the Theorem of Total Probability to write

$$\begin{aligned} \alpha_{n+1}(i, j) &= p(\mathbf{y}_{n+1} \mid z_{n+1} = i,\, s_{n+1} = j) \\ &\times \sum_l \Big[P(z_{n+1} = i \mid z_n = l) \sum_r P(s_{n+1} = j \mid s_n = r) \\ &\times p(z_n = l, s_n = r, \mathbf{Y}_0^n) \Big] \\ &= S_{n+1}(i, j) \sum_l T_1(i, l) \left[\sum_r \alpha_n(l, r)\, T_2^T(r, j) \right] \end{aligned}$$

or, in compact matrix notation,

$$\underline{\alpha}_{n+1} = \mathbf{S}_{n+1} \odot \left[\mathbf{T}_1 \underline{\alpha}_n \mathbf{T}_2^T \right] \quad (6)$$

where $\odot$ denotes pointwise multiplication and the superscript T stands for the transpose of a matrix. We initialize the forward recursion (6) with $\alpha_0(i,j) = S_0(i,j)P(z_0 = i)P(s_0 = j)$.

Backward Recursion Using again the previous assumptions on $\{z_k\}$, $\{s_k\}$, and $\{\mathbf{V}_k\}$, it follows from Bayes' law and the Theorem of Total Probability that

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

or, in compact matrix notation,

$$\underline{\beta}_n = \mathbf{T}_1^T \left[\mathbf{S}_{n+1} \odot \underline{\beta}_{n+1} \right] \mathbf{T}_2 \,. \quad (7)$$

The backward recursion in (7) is initialized with $\beta_N(i,j) = 1, \forall i \in \widetilde{\mathcal{L}}, \forall j \in \mathcal{I}$.

Optimal Bayes Smoother From the modeling assumptions,

$$\begin{aligned} \gamma_n(i, j) &= P(z_n = i,\, s_n = j \mid \mathbf{Y}_0^N) \\ &= C_N\, [p(\mathbf{Y}_{n+1}^N \mid z_n = i, s_n = j) \\ &\times p(z_n = i,\, s_n = j,\, \mathbf{Y}_0^n)] \\ &= C_N \alpha_n(i, j)\, \beta_n(i, j) \quad (8) \end{aligned}$$

where $C_N = 1/p(\mathbf{Y}_0^N)$ is a normalization constant. From γ_n, we obtain the marginal posterior distribution for the hidden centroid position z_n,

$$\widetilde{\gamma}_n(i) = P(z_n = i \mid \mathbf{Y}_0^N) = \sum_j \gamma_n(i, j) . \qquad (9)$$

Multiframe Target Detection The minimum probability of error multiframe detector at frame n is the test

$$\widetilde{\gamma}_n(L_1) \underset{H_1}{\overset{H_0}{\gtrless}} 1 - \widetilde{\gamma}_n(L_1), \qquad (10)$$

where $z_n = L_1$ is the absent target state, and H_0 and H_1 denote the hypotheses that the target is respectively absent from and present at frame n.

Multiframe Target Tracking The multiframe maximum a posteriori (MAP) estimate of the target's centroid position z_n at frame n is

$$\widehat{z}_{n|N+1} = \arg\max_{l \in \overline{\mathcal{L}}} \widetilde{\gamma}_n(l) \qquad (11)$$

where $\overline{\mathcal{L}}$ is the centroid lattice, see section 2.

Remark The forward and backward recursions in (6) and (7) may be interpreted as a generalization of the BCJR algorithm [7] from digital coding theory. The main novelty in our proposed method is that we do *simultaneous* smoothing of two hidden Markov sequences and process the observed 2D sensor images directly using a likelihood function $p(\mathbf{y}_n \mid z_n, s_n)$ that incorporates the 2D models for target signature and clutter. Details on the computation of the likelihood function with a 2D GMrf clutter model are found in [5].

4. BAYES FILTER

The multiframe detector/tracker of section 3 is a batch algorithm that uses forward and backward recursions to compute the joint posterior distribution 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 *on-line* algorithm that is based on the recursive computation of the joint posterior *filtering* distribution

$$f_{n|n}(i,j) = P(z_n = i,\, s_n = j \mid \mathbf{Y}_0^n) \qquad (12)$$

conditioned only on present and past data. Note however that

$$f_{n|n}(i,j) = \frac{p(z_n = i, s_n = j, \mathbf{Y}_0^n)}{p(\mathbf{Y}_0^n)} = C_n \alpha_n(i,j) . \quad (13)$$

From (13), we conclude that, barring a normalization constant, the recursion for the computation of $\mathbf{f}_{n|n}$ coincides with the forward recursion of the Bayes smoother, i.e.,

$$\mathbf{f}_{n+1|n+1} = K_{n+1} \mathbf{S}_{n+1} \odot \left[\mathbf{T}_1 \mathbf{f}_{n|n} \mathbf{T}_2^T \right] \qquad (14)$$

where $K_{n+1} = 1/p(\mathbf{y}_{n+1} \mid \mathbf{Y}_0^n)$ 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

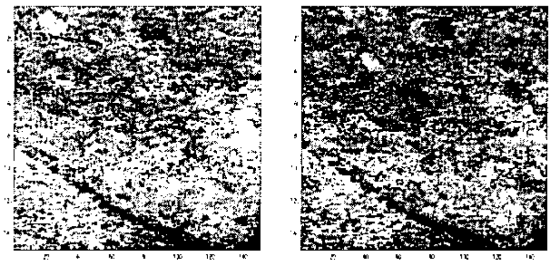
$$\widetilde{f}_{n|n}(i) = \sum_j f_{n|n}(i, j) . \qquad (15)$$

The multiframe detection and tracking steps are now identical to the detector and estimator in (10) and (11), but replacing $\widetilde{\gamma}_n$ with $\widetilde{f}_{n|n}$.

Remark: Parameter Estimation In practical situations, the clutter and target model parameters must be learned from the data. The GMrf model for real clutter can be easily estimated from test data using a single-frame, on-line version of the approximate maximum likelihood (AML) estimator introduced in [6]. No training data is required. The HMM parameters for motion and aspect can be in turn estimated from a large collection of training data using a variation of the expectation-maximization (EM) algorithm [8].

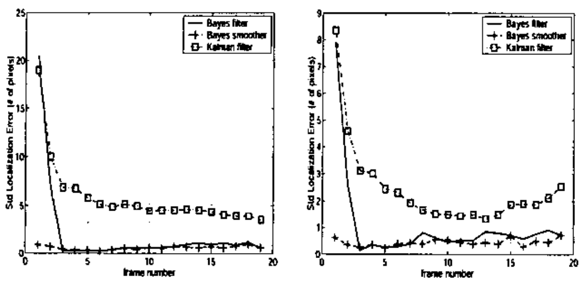
5. PERFORMANCE RESULTS

We study next the tracking performance of the proposed clutter adaptive Bayes smoother and filter with a multiaspect target that is observed in a simulated image sequence generated from real clutter infrared airborne radar (IRAR) data. The IRAR intensity imagery is from the MIT Lincoln Laboratory's Portage database and was obtained from the Center for Imaging Sciences at Johns Hopkins University. To simulate the target, we took an artificial template representing a military vehicle and generated a library of linear transformations of that template using composite operations of rotation, scaling and shearing. We then added the artificial target to the background sequence with a simulated centroid translation model that consists of two time-invariant horizontal and vertical drifts equal to 2 pixels/frame perturbed by a 2D first-order random walk where the probability of fluctuation of one pixel in both dimensions was set at 20 %. The template state was initialized with a random *unknown* aspect state in the template library and then randomly changed over time according to a first-order Markov chain. The target pixel intensity was set according to a desired low level of contrast between the template and the background. Figures 1 (a) and (b) show two simulated frames, respectively at instants $n = 0$ and $n = 6$ with peak target-to-clutter ratio (PTCR) equal to 6.3 dB. The simulated target starts from an *unknown* initial position in the sensor image and is subsequently tracked over 19 frames using (1) the on-line Bayes filter; (2) the association of a bank of correlation filters (each matched to one of the target aspect views) and a linear Kalman-Bucy filter (KBf); and (3) the batch Bayes smoother. The standard deviations of the position estimation errors in the vertical and horizontal



(a) (b)

Fig. 1. Cluttered target sequence, PTCR=6.3 dB: (a) first frame, (b) seventh frame with random target translation, rotation, scaling, and shearing.



(a) (b)

Fig. 2. Standard deviation of the localization error in number of pixels, PTCR = 3 dB: (a) vertical coordinate, (b) horizontal coordinate; Bayes filter (solid), Bayes smoother(+), Kalman Filter(□).

coordinates as a function of the number of frames are plotted in figures 2(a) and (b) respectively, with PTCR lowered to 3 dB. The standard deviations are measured in number of pixels and were obtained from 43 Monte Carlo runs. The clutter parameters at each frame were adaptively estimated using the AML algorithm [6].

We note from the plots in figure 2 that the correlation filter/KBf association performs poorly compared to the Bayes trackers, exhibiting a longer target acquisition time in both coordinates and a relatively high final position estimation error in the vertical dimension. The on-line Bayes filter has a high initial error, but, as more frames are processed, the algorithm quickly acquires the target and reaches a low steady-state localization error. Finally, the batch Bayes smoother takes advantage of both past and future information in the data volume to attenuate the large initial errors of the Bayes filter, thus achieving near-perfect tracking for this particular level of peak target-to-clutter ratio.

6. CONCLUSIONS

We introduced in this paper a new HMM-based Bayesian methodology for clutter adaptive, joint multiframe detection and tracking of randomly moving multiaspect targets in sequences of 2D digital images. The performance of the Bayes algorithm was investigated with simulated image sequences generated from real clutter IRAR data. The on-line detector/tracker based on Bayesian filtering exhibits a large initial localization error, but quickly acquires the target as more frames are processed, converging to a low steady-state estimation error. The batch version of the detector/tracker based on Bayesian smoothing improves the tracking performance by using future data information to attenuate the filter's high initial errors. Both the Bayes filter and the Bayes smoother outperform the association of a bank of correlation detectors with a linear Kalman-Bucy filter.

7. REFERENCES

[1] D. Casasent and R. Shenoy,"Synthetic aperture radar detection and clutter rejection MINACE filters," *Pattern Recognition*, 30(1), pp.151-161, 1997.

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

[3] M. Isard and A. Blake,"Condensation: conditional density propagation for visual tracking," *International Journal of Computer Vision*, 29(1), pp.5-28, 1998.

[4] P. K. Bhaharadwaj, P. Runkle, L. Carin, J. Berrie, and J. Hughes, "Multiaspect classification of airborne targets via physics-based HMMs and matching pursuits," *IEEE Trans. on Aerospace and Electronic Systems*, vol.37, n.2, pp.595-606, April 2001.

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

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

[7] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Information Theory*, pp.284-287, March 1974.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statistical Society*, Ser. B, vol. 39, n.1, pp 1-38, 1977.