CLUTTER ADAPTIVE TRACKING OF MULTIASPECT TARGETS IN IRAR IMAGERY

Marcelo G. S. Bruno

Electrical Engineering Depart., University of São Paulo
P.O. Box 61548, São Paulo SP 05424-970, Brazil
ph:(55-11) 3818-5290; email:bruno@lcs.poli.usp.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 present in this paper a clutter adaptive, multiframe Bayesian algorithm for joint detection and tracking of a multiaspect target in cluttered image sequences. The target template is randomly translated, rotated, scaled and sheared from frame to frame. Tracking performance studies with a sequence generated from real data infrared airbone radar (IRAR) imagery show a reduction in the steady-state position estimation error and in the target acquisition time when the Bayes detector/tracker is compared to the association of a bank of matched filter detectors and a linearized Kalman-Bucy tracker.

1. INTRODUCTION

We introduced in [1] an optimal Bayesian algorithm for integrated, multiframe detection and tracking of moving targets in sequences of two-dimensional (2D) digital images generated by remote imaging sensors. In this paper, we generalize the algorithm in [1] to include multiaspect targets. Random changes over time in the aspect of the target of interest may result from rotational motion and/or from variations in the conditions of observation of the target, e.g. due to changes in the relative target-sensor orientation. Previous work [2, 3] considered the problem of classification and identification of multiaspect targets using hidden Markov models (HMMs) to represent the aspect-dependent electromagnetic [2] or acoustical [3] scattering characteristics of the targets. In our work, we consider instead the problem of detecting and tracking multiaspect targets using a finite resolution imaging sensor, e.g. an infrared airbone radar (IRAR) [4]. Instead of processing the target's scattered waveforms, we have as data a preprocessed sequence of digital images in which we use the HMM formalism to model the target's translation and the rotation and/or deformation (scaling or shearing) of the target template from frame to frame. Rather than classification of stationary objects, our goal is to detect man-made moving targets that are distinct from the background, and to sequentially estimate the spatial location of those targets in the surveillance region. Our algorithm fully incorporates in a optimal way the statistical models for target motion, target aspect, and for the background clutter, and enables multiframe detection and tracking decisions using all the information available from the observed image sequence. The generalized algorithm is also clutter adaptive, allowing for on-line learning of the background clutter model parameters from the available test data.

This paper is divided into 5 sections. Section 1 is this introduction. In section 2, we review briefly the models for target aspect, target motion, and clutter that underly our unified approach to detection and tracking. We derive in section 3 the multiframe Bayes detector/tracker for multiaspect targets and discuss clutter parameter estimation. In section 4, we compare the tracking performance of the proposed algorithm to the performance of the suboptimal association of a bank of image correlation detectors (matched to the various aspects of the target) and a linearized Kalman-Bucy filter used for tracking. Finally, section 5 summarizes the contributions of the paper.

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

2. OBSERVATION MODEL

Due to the finite resolution of the sensor, each frame in the image sequence is represented by an $L \times M$ finite lattice where a pixel corresponds to one sensor resolution cell. At frame n, the observations consist of an $L \times M$ matrix, $\mathbf{Y}_n$, such that

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

where $\mathbf{F}$ is the clutter-free target model, which is a function of two *unknown* hidden variables: (a) the target's centroid position at frame n, z_n, and (b) the target aspect at frame n, s_n. The matrix $\mathbf{V}_n$ collects the background clutter returns due to the presence of spurious scatterers in the region of surveillance. We assume that the sequence $\{\mathbf{V}_n\}$, $n \geq 0$, is independent, identically distributed (i.i.d.).

Target Motion Let $\mathcal{Z}$ be an equivalent 1D representation of all possible pixel locations in the $L \times M$ sensor grid and neighboring border regions, see [1]. To build an integrated framework for detection and tracking, we augment the set $\overline{\mathcal{L}}$ with an additional dummy state that represents the absence of a target. We denote this absent target state by $z = L_0$. We extend the observation model in (1) to include the absent target case by making $\mathbf{F}(z_n, s_n) = \mathbf{0}$ if $z_n = L_0$. Denoting by $\widetilde{\mathcal{L}}$ the union of the set $\overline{\mathcal{L}}$ and the absent target state L_0, the translational motion of the target between two consecutive frames (including the probability of the target appearing in or disappearing from the sensor image) is described on $\widetilde{\mathcal{L}}$ by a first-order discrete Markov chain specified by the matrix of transition probabilities

$$T_1(k,r) = \text{Prob}(z_n = k \mid z_{n-1} = r) \quad (k,r) \in \widetilde{\mathcal{L}} \times \widetilde{\mathcal{L}}. \tag{2}$$

Target Aspect We assume that there is a finite number of possible target aspects defined on the set $\mathcal{I} = \{0, 1, \ldots, M\}$, $M > 1$. We model the changes in target aspect over time by a first-order discrete Markov sequence $\{s_n\}$, $n > 1$, specified by the transition probability matrix

$$T_2(p,l) = \text{Prob}(s_n = p \mid s_{n-1} = l) \quad (p, l) \in \mathcal{I} \times \mathcal{I}. \tag{3}$$

Clutter Model To capture the 2D spatial correlation of the background clutter, we use a noncausal, spatially homogeneous Gauss-Markov random field (GMrf) model [5]. The GMrf model is based on the principle of locality, i.e., the assumption that the clutter intensity in a given pixel location is related to the clutter intensities in the neighboring pixels. For a first-order model, the clutter returns at frame n, $V_n(i, j)$, $1 \leq i \leq L$, $1 \leq j \leq M$, are described by the 2D finite difference equation

$$\begin{aligned} V_n(i, j) &= \beta_v^c [V_n(i-1, j) + V_n(i+1, j)] \\ &+ \beta_h^c [V_n(i, j-1) + V_n(i, j+1)] + U_n(i, j) \end{aligned} \tag{4}$$

where $E[V_n(i, j)\, U_n(p, r)] = \sigma_c^2\, \delta(i-p, j-r)$, with δ denoting the 2D Kronecker delta function. The assumption of zero-mean clutter implies a pre-processing of the data that subtracts the mean of the background. A nonzero clutter mean could be accounted for trivially.

The advantages of the GMrf clutter model are twofold. First, by changing the model parameters, we are able to represent a wide variety of both heavily and weakly correlated 2D textures. Second, the inverse of the covariance matrix of a GMrf is highly sparse, see [5], with a well-defined Toeplitz-block-Toeplitz structure. This structure can be explored to design efficient algorithms that avoid costly operations of matrix inversion and multiplication.

3. DETECTION AND TRACKING ALGORITHMS

Let $\mathbf{y}_n$ be an equivalent long vector representation of the nth sensor frame, $\mathbf{Y}_n$, and let $\mathbf{Y}_0^n = [\mathbf{y}_0^T \ldots \mathbf{y}_n^T]^T$, be the collection of all observed frames from instants 0 to n. We derive next an algorithm for the recursive computation of $P(z_n = l, s_n = k \mid \mathbf{Y}_0^n)$, $l \in \widetilde{\mathcal{L}}$, $k \in \mathcal{I}$. We assume as a first approximation that the random sequences $\{z_n\}$ and $\{s_n\}$, $n \geq 0$, are statistically independent, and that both sequences are also independent of the clutter frame sequence $\{\mathbf{V}_n\}$, $n \geq 0$. The algorithm consists of three steps.

Filtering Step From Bayes' law and using the assumption that the sequence of clutter frames $\{\mathbf{V}_n\}$ is i.i.d, we write

$$P(z_n, s_n \mid \mathbf{Y}_0^n) = C_n p(\mathbf{y}_n \mid z_n, s_n) P(z_n, s_n \mid \mathbf{Y}_0^{n-1}) \tag{5}$$

where C_n is a normalization factor that is independent of z_n and s_n.

Template State Prediction Under the assumption that s_n is independent of $\{z_n\}$, and $\{\mathbf{V}_n\}$, $n \geq 0$, and modeling $\{s_n\}$ as a first-order discrete Markov chain, we conclude that, conditioned on s_{n-1}, s_n is independent of $\mathbf{Y}_0^{n-1}$, and therefore we write

$$\begin{aligned} P(z_n, s_n \mid \mathbf{Y}_0^{n-1}) &= \sum_{s_{n-1}} [P(s_n \mid s_{n-1}) \\ &\times P(z_n, s_{n-1} \mid \mathbf{Y}_0^{n-1})] \,. \end{aligned} \tag{6}$$

Translation Prediction Using a similar reasoning as in the previous step, we get

$$\begin{aligned} P(z_n, s_{n-1} \mid \mathbf{Y}_0^{n-1}) &= \sum_{z_{n-1}} [P(z_n \mid z_{n-1}) \\ &\times P(z_{n-1}, s_{n-1} \mid \mathbf{Y}_0^{n-1})] \,. \end{aligned}$$

The recursion is initialized by making $P(z_0, s_0 \mid \mathbf{Y}_0^{-1}) = P(z_0)\, P(s_0)$. The *marginal* posterior probability of the centroid position z_n conditioned on the observations is given by

$$P(z_n \mid \mathbf{Y}_0^n) = \sum_{s_n \in \mathcal{I}} P(z_n, s_n \mid \mathbf{Y}_0^n) \,. \tag{7}$$

Minimum Probability of Error Detector Let H_0 denote the hypothesis that the target is absent at frame n and H_1 denote the hypothesis that the target is present during the nth sensor scan. The minimum probability of error Bayes detector follows the decision rule

$$P(z_n = L_0 \mid \mathbf{Y}_0^n) \underset{H_1}{\overset{H_0}{\gtrless}} 1 - P(z_n = L_0 \mid \mathbf{Y}_0^n) \tag{8}$$

where L_0 is the dummy absent target state, see section 2.

MAP Tracker If hypothesis H_1 is declared true, we compute the conditional probability vector

$$\begin{aligned} Q_l^f[n] &= P(z_n = l \mid \text{target is present}, \mathbf{Y}_0^n) \quad l \in \overline{\mathcal{L}} \\ &= \frac{P(z_n = l \mid \mathbf{Y}_0^n)}{1 - P(z_n = L_0 \mid \mathbf{Y}_0^n)} \end{aligned} \tag{9}$$

The MAP estimate of target's centroid position is

$$\hat{z}_{n|n} = \arg \max_{l \in \mathcal{L}} Q_l^f[n] \ . \tag{10}$$

Clutter Adaptation As a first approximation, since the target is very small compared to the background, we simply ignore the presence of the target and estimate the GMrf clutter parameters directly from each available sensor frame $\mathbf{Y}_n$ using a variation of *approximate maximum likelihood* (AML) parameter estimation algorithm introduced in [5]. Table 1 summarizes the AML parameter estimation algorithm given the L x M nth sensor frame $\mathbf{Y}_n$, see [5] for further details. Note that the algorithm described in Table 1 is a fast, single frame, on-line estimator that is applied directly to each frame of the test data. No training data or previous learning are required.

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:
- $\widehat{\beta_h^c} = \frac{\epsilon \chi_h}{|\chi_v| \cos(\frac{\pi}{L+1}) + \alpha |\chi_h| \cos(\frac{\pi}{M+1})}$.
- $\widehat{\beta_v^c} = \frac{\epsilon \chi_v}{|\chi_v| \cos(\frac{\pi}{L+1}) + \alpha |\chi_h| \cos(\frac{\pi}{M+1})}$.

e) Clutter power estimate
- $\widehat{\sigma_c^2} = \frac{1}{LM} (S_y - \widehat{\beta_h^c} * X_h - \widehat{\beta_v^c} * X_v)$.

Table 1. Summary of the AML parameter estimation algorithm for an $L \times M$ Gauss-Markov random field.

Remark: In real scenarios, the parameters of the target aspect model are also unknown to the tracker. If the motion model parameters are known and training data is available, the parameters of the HMM that describes the changes in target aspect may be learned from the data using the Baum-Welch reestimation (or Expectation-Maximization) algorithm, see [6, 7]. We omit this discussion here for lack of space.

4. TRACKING PERFORMANCE

We study in the sequel the tracking performance of the proposed clutter adaptive, multiframe detector/tracker with a multiaspect target observed in a simulated image sequence that was generated from real-world infrared airbone radar (IRAR) intensity imagery. The IRAR data is from the MIT Lincoln Laboratory's Portage database and was obtained via the Center for Imaging Science at Johns Hopkins University. Figure 1 shows an imaged scene where we see two different stretches of terrain next to a body of water (the dark region on the top right 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 clutter parameters β_h^c, β_v^c and σ_c^2, using the AML estimator. The background clutter movie is then generated by adding to the matrix of previously stored local means a sequence of synthetized random GMrf samples generated using the estimated background clutter parameters. 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 the target centroid position changing from frame to frame according to a near-constant velocity translation model. The near-constant velocity model consists of 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 randomly initialized with an unknown aspect from the template library and then changed over time according to a first-order Markov chain. At any given frame, the true aspect of the target and the clutter parameters are *unknown* to the detector/tracker. The target pixel intensity is on the other hand time-invariant and known, and was set according to a desired low level of contrast between the template and the background. Figures 2 (a) and (b) show two simulated frames, respectively at instants $n = 0$ and $n = 6$.

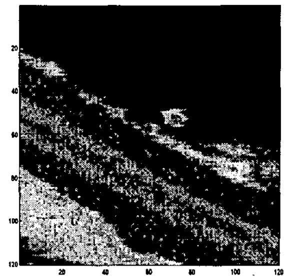
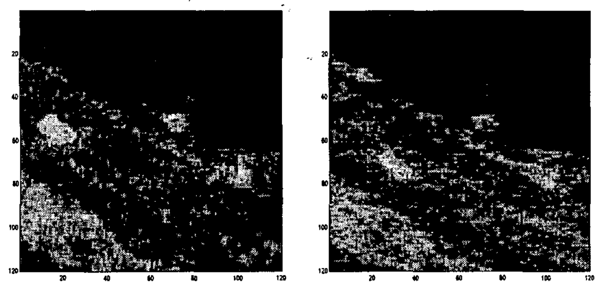


Fig. 1. IRAR intensity data (Portage database).

To have a quantitative assessment of the performance of the Bayes tracker, we ran a Monte Carlo simulation with a total of 60 runs. At each Monte Carlo run, the simulated target departs from a *unknown* location in the land portion of the background and is subsequently tracked over 18 consecutive frames using the clutter adaptive, multiframe Bayes detector/tracker. A library of seven different target aspects was used for the 18-frame simulations. For performance comparison purposes, we also tracked the target using the suboptimal association of a bank of 2D matched filters (each



(a) (b)

Fig. 2. Cluttered target sequence: (a) first frame, (b) seventh frame with random target translation, rotation, scaling, and shearing.

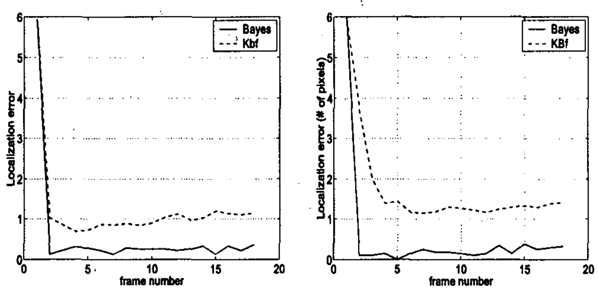


Fig. 3. Standard deviation in number of pixels of the position estimate error: (a) vertical coordinate, (b) horizontal coordinate.

one matched to one of the possible target aspects) and a linearized Kalman-Bucy filter (KBf). The preliminary position estimate generated by the matched filter bank is treated by the KBf as a noisy observation of the true target position.

The Bayes tracker and the matched filter/KBf association were initialized with the same initial position estimation error so that we could compare both the steady-state error and the target acquisition time (i.e., the number of frames that are necessary for the tracker to reach a steady-state) for the two algorithms. The standard deviation (in number of pixels) of the localization error in the vertical coordinate is shown in Figure 3(a). The corresponding curves for the horizontal coordinate are shown in Figure 3(b). The performance curves for the Bayes tracker are shown in solid line, while the curves for the KBf are shown in dashed line. The plots in Figure 3 show that the Bayes tracker has a lower steady-state position estimation error than the matched filter/KBf association. The curves for the horizontal dimension in Figure 3(b) also highlight that the KBf has a longer target acquisition time than the proposed Bayes tracker.

5. SUMMARY

We introduced in this paper a clutter adaptive, multiframe Bayesian algorithm for joint detection and tracking of moving targets in 2D cluttered image sequences. The algorithm is designed for tracking multiaspect targets whose templates are randomly translated, rotated, scaled and/or sheared from frame to frame. The clutter adaptive multiframe detector/tracker was tested on a simulated image sequence generated from real data infrared airborne radar (IRAR) imagery. The corresponding tracking performance with a multiaspect target was compared to the performance of the suboptimal association of a bank of matched filters and a linearized Kalman-Bucy filter. The performance curves show an improvement in the steady-state accuracy of the 2D target position estimate and in the target acquisition time when the Bayes detector/tracker is used.

6. REFERENCES

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

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

[3] P. Runkle, P. K. Bharadwaj, L. Couchman, and L. Carin, "Hidden Markov models for multi-aspect target classification," *IEEE Transactions on Signal Processing*, vol.47, pp 2035-2040, July 1999.

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

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

[6] L. R. Rabiner,"A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol.77, n.2, pp 257-286, February 1989.

[7] 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, n.1, pp 164-171, 1970.