MULTIFRAME BAYESIAN TRACKING OF CLUTTERED TARGETS WITH RANDOM MOTION

Marcelo G. S. Bruno

Electrical Engineering Depart., University of São Paulo
P.O. Box 61548, São Paulo SP 05424-970, Brazil
ph:(55-11) 818-5290; email:bruno@lcs.poli.usp.br

José M. F. Moura *

Massachusetts Institute of Technology, Room 35-203
77 Massachusetts Ave, Cambridge, MA 02139-4703 USA
ph: (617) 253-7250; email: moura@mit.edu

ABSTRACT

We present in this paper a multiframe Bayesian algorithm for detection and tracking of heavily cluttered rigid bodies with random translational and rotational motion. Monte Carlo simulations with synthetic targets and clutter show that the proposed algorithm achieves substantial performance gains over the common association of a maximum likelihood position estimator and a linearized Kalman-Bucy filter.

1. INTRODUCTION

We introduced in [1] a new Bayesian algorithm for optimal multiframe detection and tracking of rigid objects in a sequence of two-dimensional (2D) images. Previous solutions to this problem [2] are based on the ad-hoc separation of the detection and tracking tasks into two separate stages. Typically, a single frame detector generates preliminary estimates of the position of targets of interest. These estimates are subsequently associated to a linearized tracking filter, generally a Kalman-Bucy filter (KBf). By contrast, the algorithm described in [1] is a nonlinear, integrated multiframe detector/tracker that incorporates the models for target motion, target signature, and clutter into a single framework, using as data the sequence of raw sensor images.

The discussion in [1] was restricted to targets with random translational motion and known templates. In the present paper, we extend the algorithm to targets with random rotations. Stochastic rotational motion makes it more difficult to estimate the target position since the orientation of the spatial distribution of target signatures around the target centroid varies from frame to frame and is no longer known to the tracker. In addition, target images are also heavily obscured by clutter arising from spurious reflectors in the view of the sensor. We model the background clutter as a 2D correlated, noncausal, Gauss-Markov random field (GMrf) [3]. This model is capable of accurately describing a wide variety of backgrounds [3], ranging from smooth patterns to highly structured texture. We evaluate the tracking performance of the proposed algorithm through Monte Carlo simulations. Our results show substantial improvements over traditional schemes such as the usual association of a single frame maximum likelihood (ML) position estimator and a linearized KBf tracker.

(*) On sabbatical leave from the ECE Department, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA.

2. PROBLEM FORMULATION

A sensor device, e.g., a high resolution radar or an infrared (IR) camera, generates a sequence of cluttered images of possible targets of interest. The goal is to determine at each frame whether targets are present or not, and, if a target is declared present, to estimate its location. Due to the sensor's finite resolution, the sensor image is discretized by a 2D finite lattice. For simplicity, we consider the situation when there is at most one target present at each sensor frame. The target template rotates randomly from frame to frame. We assume that there is a finite number of possible template states, with each state representing one possible spatial orientation of the target. In addition, the target centroid is also randomly translated from frame to frame. The translational and rotational motions are described by two hidden Markov models (HMMs) with known transition probabilities.

2.1. Sensor and Target Model

We define a 2D rectangular region of size $(r_i + r_s + 1) \times (l_i + l_s + 1)$ that contains all possible rotated templates for the target of interest. To model situations when targets move in and out of the sensor grid, we define the *centroid lattice* $\widehat{\mathcal{L}} = \{(i,j)\colon\ -r_s + 1 \le i \le L + r_i,\ -l_s + 1 \le j \le M + l_i\ \}$, where L and M are the number of resolution cells in each dimension. The centroid lattice collects all possible values of the target centroid

0-7803-6297-7/00/$10.00 © 2000 IEEE

position for which at least one target pixel may lie inside the sensor's image.

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

$$\widetilde{\mathcal{L}} = \{\, l\colon 1 \le l \le (L+r_i+r_s)(M+l_i+l_s)+1\,\} \ . \quad (1)$$

Target Model Let $z_n \in \widetilde{\mathcal{L}}$ denote the target centroid position during the nth frame. Let m be the number of target template states in the sensor image and denote by $s_n \in \mathcal{I} = \{0, 1, \ldots, m-1\}$ the state of the target template at instant n. We model the clutter-free image of a target that is present as the nonlinear function

$$\mathbf{F}\left[z_n(i_n, j_n),\, s_n\right] = \sum_{k=-r_i}^{r_s} \sum_{l=-l_i}^{l_s} a_{k,l}(s_n)\mathbf{E}_{i_n+k, j_n+l} \ . \quad (2)$$

In (2), for $1 \le i \le L$, $1 \le j \le M$, $\mathbf{E}_{i,j}$, is an L x M matrix whose entries are all zero, except for the element (i,j) which is one. For any $(i,j) \notin \mathcal{L}_1 \times \mathcal{L}_2$, where $\mathcal{L}_1 = \{l\colon 1 \le l \le L\}$ and $\mathcal{L}_2 = \{l\colon 1 \le l \le M\}$, $\mathbf{E}_{i,j}$ is identically zero. Finally, when no target is present, the target model returns a null image, i.e., if $z_n = (L+r_i+r_s)(M+l_i+l_s)+1$, we make $\mathbf{F}(z_n,\, s_n) = \mathbf{0}_{L\times M}$, $\forall s_n \in \mathcal{I}$.

The coefficients $a_{k,l}$ in (2) are called the *target signature* coefficients and vary according to the template model s_n. For a given s_n, the signature coefficients are the product of a binary parameter $c_{k,l}(s_n) \in \mathcal{B} = 0, 1$, which defines the target shape, and a real coefficient $\phi_{k,l}(s_n) \in \Re$, that specifies the pixel intensities of the target. For simplicity, we assume in this paper that the pixel intensities are constant and time-invariant for all template models. Randomly varying pixel intensities are subject of ongoing research.

2.2. Measurements and Clutter Model

Spurious reflectors and environmental noise in the view of the sensor corrupt the measurements with clutter. For a single target scenario, the model for the nth sensor frame is the L x M matrix

$$\mathbf{Y}_n = \mathbf{F}(z_n, s_n) + \mathbf{V}_n \quad (3)$$

where z_n is the position of the target centroid in the equivalent 1D extended lattice (including the absent state), $\mathbf{F}(.)$ is the 2D extended target model described in the previous subsection and $\mathbf{V}_n$ is the background clutter frame.

In general, each clutter frame $\mathbf{V}_n$ may exhibit a spatial correlation. We describe the clutter's spatial correlation using the spatially homogeneous Gauss-Markov random field (GMrf) model [3]. Since there is no preferred direction in space, we allow the neighborhood region around pixel $(i,\, j)$ in the GMrf model to be *noncausal* with respect to all possible orderings in the 2D plane. For example, a a *first order* noncausal GMrf is described by the finite difference equation model

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

where $U_n(i,\, j)$ is a Gaussian prediction error term that is statistically orthogonal to all $V_n(k,\, l)$, $(k,\, l) \neq (i,\, j)$. A set of boundary conditions is added to specify equation (4) near the boundaries of the lattice.

2.3. Motion and Template State Models

The probability of a displacement of the target centroid between two consecutive sensor frames is described by the matrix $\mathbf{T}_1$ such that

$$T_1(k,r) = \text{Prob}(z_n = k \mid z_{n-1} = r) \qquad (k,r) \in \ \widetilde{\mathcal{L}}\times\widetilde{\mathcal{L}} \, . \quad (5)$$

On the other hand, the changes in the target template from one state to another are specified by the matrix $\mathbf{T}_2$, where

$$T_2(k,r) = \text{Prob}(s_n = k \mid s_{n-1} = r) \qquad (k,\, r) \in \mathcal{I} \times \mathcal{I} \, . \quad (6)$$

3. DETECTION/TRACKING ALGORITHMS

Let $\mathbf{y}_n$ be the 1D row-lexicographed vector representation of the 2D sensor image $\mathbf{Y}_n$ and introduce the vector $\mathbf{Y}_0^n = \left[\mathbf{y}_0^T\ \mathbf{y}_1^T\ \ldots\ \mathbf{y}_n^T\right]^T$. 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 $\{s_n\}$, $n \ge 0$, are statistically independent, and that both sequences are also independent of the clutter frames sequence $\{\mathbf{V}_n\}$, $n \ge 0$. 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 independent, identically distributed (i.i.d), we write

$$P(z_n, s_n \mid \mathbf{Y}_0^n) = C_n p(\mathbf{y}_n \mid z_n,\, s_n)\, P(z_n,\, s_n \mid \mathbf{Y}_0^{n-1}) \quad (7)$$

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

Rotation 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

$$P(z_n, s_n \mid \mathbf{Y}_0^{n-1}) = \sum_{s_{n-1}} [P(s_n \mid s_{n-1}) \times P(z_n, s_{n-1} \mid \mathbf{Y}_0^{n-1})] \ . \tag{8}$$

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

$$P(z_n, s_{n-1} \mid \mathbf{Y}_0^{n-1}) = \sum_{z_{n-1}} [P(z_n \mid z_{n-1}) \times P(z_{n-1}, s_{n-1} \mid \mathbf{Y}_0^{n-1})] \ . \tag{9}$$

The *marginal* posterior probability of the centroid position z_n conditioned on the observations is

$$P(z_n \mid \mathbf{Y}_0^n) = \sum_{s_n \in \mathcal{I}} P(z_n, s_n \mid \mathbf{Y}_0^n) \ . \tag{10}$$

We present in the sequel the multiframe detection and tracking algorithms.

Multiframe Detection Let $L_1 = (L+r_i+r_s)\,(M+l_i+l_s)$. Denote by H_0 the hypothesis that the target is absent and, by H_1, the hypothesis that the target is present. Assuming equal cost for misses and false alarms and zero cost for correct decisions, the minimum probability of error Bayes detector is the test

$$\frac{P(H_0 \mid \mathbf{Y}_0^n)}{P(H_1 \mid \mathbf{Y}_0^n)} \underset{H_1}{\overset{H_0}{\gtrless}} 1 \Leftrightarrow \frac{P(z_n = L_1 + 1 \mid \mathbf{Y}_0^n)}{1 - P(z_n = L_1 + 1 \mid \mathbf{Y}_0^n)} \underset{H_1}{\overset{H_0}{\gtrless}} 1 \ . \tag{11}$$

Multiframe Tracking Introduce the conditional probability vector $\mathbf{Q}^f\,[n]$ such that, for all $l \in \overline{\mathcal{L}}$

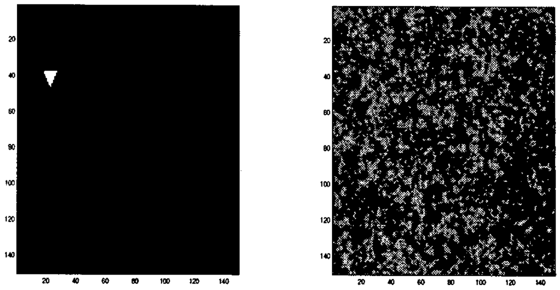
$$\begin{aligned} Q_l^f\,[n] &= P(\mathbf{z}_n = l \mid \text{target is present}, \mathbf{Y}_0^n) \\ &= \frac{P(z_n = l \mid \mathbf{Y}_0^n)}{1 - P(z_n = L_1 + 1 \mid \mathbf{Y}_0^n)} \end{aligned} \tag{12}$$

where $\overline{\mathcal{L}}$ is the 1D equivalent centroid lattice, see section 2. The maximum a posteriori (MAP) estimate of the the target's centroid position assuming that the target is present is

$$\hat{z}_{\text{map}}\,[n] = \arg\max_{l \in \overline{\mathcal{L}}} Q_l^f\,[n] \ . \tag{13}$$

4. TRACKING PERFORMANCE

We examine next the tracking performance of the Bayes algorithm using synthetic data. The illustrative simulated targets are 2D triangular-shaped objects with constant pixel intensity and size 9 x 9. The targets are cluttered by a first order, highly correlated GMrf background with $\beta_h = \beta_v = 0.24$.

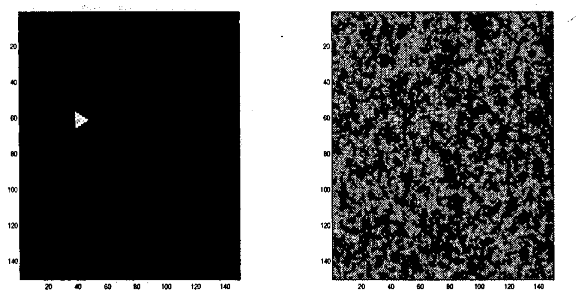


(a) (b)

Figure 1: Initial frame: (a) Clutter-free target image, (b) Simulated sensor image, PSNR = 3 dB.

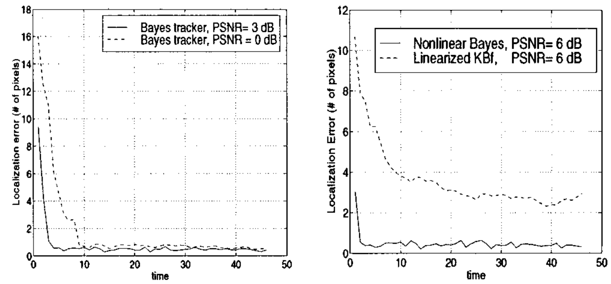
At each sensor scan, there is only a single target present. The target moves in a 150 x 150 discrete grid with constant nominal velocities of 2 resolution cells per frame in both the horizontal and vertical directions. The target centroid position fluctuates around its nominal location according to a first order 2D random walk model. If the nominal centroid position is the pixel (i, j), there is an equal probability $p = 0.20$ that the real centroid position be any of the pixels $(i-1, j)$, $(i+1, j)$, $(i, j+1)$, or $(i, j-1)$. In addition to the translational motion, the triangular template of a target that is present rotates randomly around its centroid.

Figures 1 (a) and (b) show examples respectively of the clutter-free target image and the target plus clutter (sensor) image with peak signal-to-noise ratio (PSNR) equal to 3 dB at instant $n = 0$. The target is centered at the coordinates (42, 23). Figures 2 (a) and (b) show respectively the clutter-free and target plus clutter images of the same target at instant $n = 11$ for the same level of PSNR. The target centroid has moved randomly to position (61, 43), while the target template has undergone a 90 degrees random rotation. The simulated target departures at instant zero from an *unknown* random location in the 50 x 50 upper left corner of the image and is subsequently tracked over 45 consecutive sensor frames. For computational simplicity, but without loss of generality, we consider in the simulation only 4 different angular positions for the template, respectively 0, 90, 180 and 270 degrees. The probability of a 90 degrees rotation between two frames is set to 0.8. For the nonlinear Bayes tracker, figure 3(a) shows the evolution over time of the standard deviation of the error in the centroid's vertical position estimate. The standard deviation is expressed in number of pixels and evaluated by repeating the experiment 130 times with



(a) (b)

Figure 2: Eleventh frame: (a) Clutter-free target image, (b) Simulated sensor image, PSNR = 3 dB.



(a) (b)

Figure 3: (a) Performance of the nonlinear Bayes tracker in correlated GMrf clutter; (b) Performance of the nonlinear Bayes tracker vs the linearized KBf.

two values of PSNR, respectively 3 and 0 dB. The corresponding curves for the horizontal position estimate are qualitatively similar and are omitted for lack of space. We see from figure 3(a) that the initial localization error declines with time as new measurements become available to the tracker. The initial and steady-state errors, as well as the target acquisition time, increase as PSNR decreases.

Next, we compare the nonlinear Bayes tracker with the alternative suboptimal association of a single frame maximum likelihood (ML) tracker and a linearized Kalman Bucy filter. The single frame ML estimate of the centroid position, assuming the target is present and using the hypothesis of independence between s_n and z_n, is given by

$$\hat{z}_{ML} = \arg\max_{z_n} \sum_{s_n} p(\mathbf{y}_n \mid s_n, z_n)\, P(s_n) \ . \qquad (14)$$

The "a priori" probabilities of the template state s_n are computed from the Markov chain that describes the sequence $\{s_n\}$. On the other hand, for a given value of s_n and assuming GMrf clutter, the kernel $p(\mathbf{y}_n \mid z_n, s_n)$, for all $z_n \in \overline{\mathcal{L}}$, is basically computed [1, 4] using a noncausal differential operator (whose weights depend on the GMrf parameters β_h and β_v), followed by a 2D correlation filter that is matched to the template s_n. The output of the single frame ML tracker is treated as a preliminary position estimate which is subsequently incorporated as a noisy measurement of the true target position into a multiframe, linearized Kalman filter (KBf). The KBf is used to compensate the large errors in the ML tracker's position estimates.

Figure 3(b) plots the standard deviation over time of the error in the vertical position estimate for the nonlinear Bayes tracker from section 3 and the suboptimal linearized KBf tracker described in the previous paragraphs, in a scenario of PSNR equal to 6 dB. We see from the plot that the KBf tracker has much higher initial and steady state position estimate errors and a longer target acquisition time when compared to the Bayes tracker.

5. SUMMARY

We present in this paper a Bayesian algorithm for multiframe detection and tracking of extended targets in a sequence of 2D cluttered images. The targets move randomly in a finite 2D grid and also have randomly-rotating templates. Performance studies using Monte Carlo simulations show that there is a significant improvement over traditional trackers such as the usual association of a maximum likelihood position estimator and a linearized Kalman-Bucy filter.

ACKNOWLEDGEMENT

This work was funded by ONR grant N00014-97-1-0800. The work of the first author was also partially funded by FAPESP, São Paulo, Brazil.

6. REFERENCES

[1] M. G. S. Bruno and J. M. F. Moura, "The optimal 2D multiframe detector/tracker", *AEÜ Int. J. Electron. Commun.*, vol. 56, no. 2, pp. 346–355, December 1999.

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

[3] J. M. F. Moura and N. Balram, "Recursive structure of noncausal Gauss-Markov random fields", *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 334–354, March 1992.

[4] M. G. S. Bruno, *Joint Detection and Tracking of Moving Targets in Clutter*, Ph.D. thesis, Carnegie Mellon University, 1998.