OPTIMAL MULTIFRAME DETECTION AND TRACKING IN DIGITAL IMAGE SEQUENCES

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 Bayesian algorithm for optimal multiframe detection and tracking of small extended targets in two-dimensional (2D) finite resolution images. The algorithm integrates detection and tracking into a single framework using as data a sequence of cluttered sensor snapshots. Performance studies using Monte Carlo simulations show substantial improvements when the proposed Bayes tracker is compared to the association of a correlation filter and a linearized Kalman-Bucy filter. Likewise, there are significant detection performance gains of up to 6 dB in peak signal-to-noise ratio (PSNR) when the multiframe Bayes detector is compared to a single frame likelihood ratio test (LRT) detector.

1. INTRODUCTION

We introduced in [3] a new Bayesian algorithm for integrated multiframe detection and tracking of extended targets in one-dimensional (1D) finite discrete grids. In this paper, we extend the algorithm to two dimensions (2D) and present comprehensive 2D detection and tracking performance studies based on Monte Carlo simulations.

A sensor device (e.g., an imaging radar or an infrared camera) generates a sequence of finite resolution images of a surveillance region. The goal is to decide whether targets are present or not at any given frame of the sequence, and, if targets are declared present, to track them across the successive frames. Previous solutions to this problem propose a suboptimal separation of the detection and tracking stages [1]. A single frame predetection stage generates a preliminary estimate of the targets' position. These preliminary estimates are then statistically associated to a linear tracking filter, typically a Kalman-Bucy filter (KBf). By contrast, we use nonlinear stochastic filtering to design the optimal multiframe detector/tracker that integrates detection and tracking into a common framework.

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

The optimal detector/tracker benchmarks the performance that can be achieved by any alternative algorithm. We quantify this benchmark using Monte Carlo simulations that generate optimal detection and tracking performance curves in the 2D case. These curves provide a bound that indicates the performance gain to be had over any other suboptimal scheme. We show in the paper that there is a significant margin of improvement in tracking performance in comparison with the common association [1] of a correlation filter and a linearized KBf. We also show that there are detection performance gains of up to 6 dB in PSNR to be had over the usual single frame likelihood ratio test (LRT) detector.

The paper is divided into 6 sections. Section 1 is this introduction. Section 2 reviews briefly the target signature, target motion, and clutter models that underly our proposed algorithm. Section 3 presents the detection and tracking algorithms. Further details are found in [4]. Sections 4 and 5 present the performance results, respectively for tracking and detection. Finally, section 6 summarizes the main contributions of the paper.

2. PROBLEM SETUP

We consider as targets of interest 2D rigid bodies with translational motion. The target templates are bounded by a 2D rectangular region of size $(r_i + r_s + 1) \times (l_i + l_s + 1)$. For simplicity, we assume that there is at most one target present at each sensor frame.

2.1. Sensor and Target Model

Due to sensor's finite resolution, the surveillance space is discretized by a uniform 2D finite lattice. To model situations when targets move in and out of the sensor grid, we define the *centroid lattice* $\widehat{\mathcal{L}} = \{(i,j)\colon -r_s+1 \leq i \leq L+r_i,\ -l_s+1 \leq j \leq M+l_i\ \}$, where L and M are the number of resolution cells in each dimension. The centroid lattice collects all possible values of the target centroid position for which at least one target pixel lies inside the sensor's surveillance space.

Let $\overline{\mathcal{L}}$ be an equivalent 1D representation of the centroid lattice $\widehat{\mathcal{L}}$ obtained by row lexicographic ordering. To build an integrated framework for detection and tracking, we augment $\overline{\mathcal{L}}$ 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

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

Extended Target Model We represent the noise-free image of a target centered at $z_n \in \overline{\mathcal{L}}$ as a spatial distribution of pixel intensities around the centroid position. Formally, for $z_n = (i_n + r_s - 1)(M + l_i + l_s) + (j_n + l_s)$, $(i_n, j_n) \in \widehat{\mathcal{L}}$, the target image is the L x M matrix $\mathbf{F}(z_n)$ such that

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

In (2), for $1 \leq i \leq L$, $1 \leq j \leq M$, $\mathbf{E}_{i,j}$, 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 $\mathcal{L}_1 = \{l : 1 \leq l \leq L\}$ and $\mathcal{L}_2 = \{l : 1 \leq l \leq M\}$, $\mathbf{E}_{i,j}$ is identically zero. The coefficients $a_{k,l}^{n}$ are referred to as the *target signature* coefficients. They specify both the target shape and its pixel intensities. For simplicity, we assume in this paper that the signature coefficients are deterministic and known.

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 $\mathbf{F}(z_n) = \mathbf{0}_{L \times M}$.

2.2. Measurements and Clutter Model

The sensor images are corrupted by random clutter originating from spurious reflectors. In a single target scenario, we model the nth sensor frame as the L x M matrix

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

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 matrix, also referred to as the background clutter frame. We assume that the clutter frames $\mathbf{V}_n$, $n = 0, 1, \ldots$, are independent, identically distributed (i.i.d.).

In general, each clutter frame $\mathbf{V}_n$ may exhibit a spatial (or intraframe) correlation. We capture the clutter's spatial correlation using the spatially homogeneous Gauss-Markov random field (GMrf) model [6]. For each pixel (i, j), $1 \leq i \leq L$, $1 \leq j \leq M$, we define its neighborhood system η_{ij}. Under the assumption of spatial homogeneity, $\eta_{ij} = \eta$, $\forall (i, j)$. Since there is no preferred direction in space, in contrast to previous work [2], we allow the neighborhood region η to be *noncausal* with respect to all possible recursive orderings in the 2D plane.

The clutter field, $\mathbf{V}_n$, is a finite order, noncausal, spatially homogeneous GMrf if it is the output, for $1 \leq i \leq L$, $1 \leq j \leq M$, of the two-dimensional finite difference equation [6]

$$V_n(i,j) = \sum_{(i-k,\, j-l)\, \in \eta} -\alpha_k^l V_n(i-k, j-l) + U_n(i,j) \quad (4)$$

where $U_n(i,\, j)$ is a Gaussian input that is statistically orthogonal to $V_n(r,\, l)$, $\forall (r,\, l) \neq (i,\, j)$. For example, a *first order* noncausal GMrf is described by the model

$$\begin{aligned} V_n(i,\, j) &= \beta_v \left[V_n(i-1,\, j) + V_n(i+1,\, j)\right] \\ &+ \ \beta_h \left[V_n(i,\, j-1) + V_n(i,\, j+1)\right] + U_n(i,\, j) \quad (5) \end{aligned}$$

A set of boundary conditions is added to specify equations (4) and (5) near the boundaries of the lattice.

The 2D GMrf model is of particular interest due to the highly structured parametrization of the field inverse covariance matrix [5, 6]. This structure is explored to design fast algorithms, see [4].

2.3. Motion Model

With translational motion, the dynamics of the target centroid is sufficient to describe the target motion. The likelihood of a change between two consecutive frames in the position of the centroid in the extended lattice $\widetilde{\mathcal{L}}$ is specified by a *transition probability matrix*, $\mathbf{T}$, such that, for $(k, r) \in \widetilde{\mathcal{L}} \times \widetilde{\mathcal{L}}$,

$$T(k, r) = \text{Prob}(z_n = k \mid z_{n-1} = r) \ . \quad (6)$$

3. MULTIFRAME DETECTOR/TRACKER

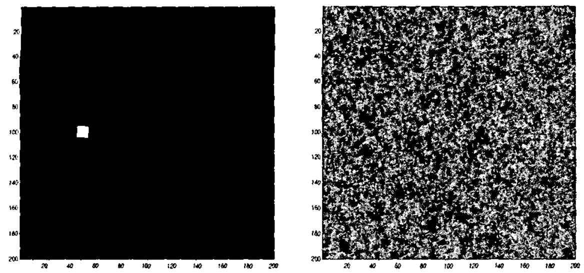
Denote by $\mathbf{y}_n$ the 1D long vector representation of the 2D sensor image $\mathbf{Y}_n$. Introduce also the vector $\mathbf{Y}_0^n = \left[\mathbf{y}_0^T \ \mathbf{y}_1^T \ \cdots \ \mathbf{y}_n^T\right]^T$. We develop next an algorithm for the recursive computation of the posterior probabilities $P(z_n = l \mid \mathbf{Y}_0^n)$, $l \in \widetilde{\mathcal{L}}$. The algorithm is divided into two steps.

Prediction Step From the Theorem of Total Probability, we write

$$P(z_n \mid \mathbf{Y}_0^{n-1}) = \sum_{z_{n-1}} P(z_n \mid z_{n-1}) P(z_{n-1} \mid \mathbf{Y}_0^{n-1}) \quad (7)$$

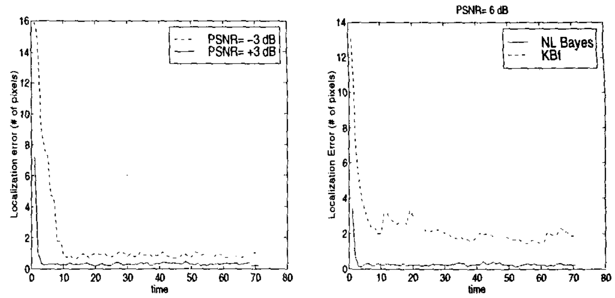
Filtering Step Using Bayes' Law, we get

$$P(z_n \mid \mathbf{Y}_0^n) = C_n p(\mathbf{y}_n \mid z_n) P(z_n \mid \mathbf{Y}_0^{n-1}) \quad (8)$$



(a) (b)

Figure 1: (a) Clutter-free target image, (b) Simulated sensor image, PSNR = 0 dB.



(a) (b)

Figure 2: (a) Performance of the nonlinear Bayes tracker in correlated GMrf clutter; (b) Performance of the nonlinear Bayes tracker vs the linearized KBf.

where C_n is a normalization constant. We now consider detection and tracking.

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 detector is the test

$$\frac{P(H_0 \mid \mathbf{Y}_0^n)}{P(H_1 \mid \mathbf{Y}_0^n)} \underset{H_1}{\overset{H_0}{\gtrless}} 1 \Leftrightarrow \frac{P(z_n = L_1 + 1 \mid \mathbf{Y}_0^n)}{1 - P(z_n = L_1 + 1 \mid \mathbf{Y}_0^n)} \underset{H_1}{\overset{H_0}{\gtrless}} 1 \ . \tag{9}$$

Tracking Introduce the conditional probability

$$\begin{aligned} Q_l^f[n] &= P(\mathbf{z}_n = l \mid \text{target is present}, \mathbf{Y}_0^n) \qquad l \in \overline{\mathcal{L}} \\ &= \frac{P(z_n = l \mid \mathbf{Y}_0^n)}{1 - P(z_n = L_1 + 1 \mid \mathbf{Y}_0^n)} \end{aligned} \tag{10}$$

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

$$\hat{z}_{\text{map}}[n] = \arg\max_{l \in \overline{\mathcal{L}}} Q_l^f[n] \ . \tag{11}$$

4. TRACKING PERFORMANCE

We examine first the tracking-only performance of the Bayes algorithm using synthetic data. The simulated targets are 2D rectangular objects with constant pixel intensity and size 9 x 9. The targets are cluttered by a first order, heavily correlated GMrf background with $\beta_h = \beta_v = 0.24$. Figures 1 (a) and (b) show examples respectively of the clutter-free target image and the target plus clutter (sensor) image when the target is centered at pixel (100, 50) with PSNR= 0 dB. At each sensor scan, there is only a single target present. The target moves in a 200 x 200 discrete grid with constant nominal velocities of 2 resolution cells/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 $(i, j-1)$.

At instant zero, the simulated target departures from an *unknown* random location in the 50 x 50 upper left corner of the image. The target is subsequently tracked over 70 consecutive sensor frames. For the nonlinear Bayes tracker, figure 2(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 150 times with two values of PSNR, respectively +3 and -3 dB. The corresponding curves for the horizontal position estimate are qualitatively similar and are omitted for lack of space. Figure 2(a) shows that there is an initial localization error which declines over time as new measurements become available. The target acquisition time (i.e., the number of sensor scans for the error to reach its steady state) increases as the PSNR decreases. Likewise, the initial and steady state localization errors also increase with decreasing PSNR.

Next, we compare the nonlinear Bayes tracker with the alternative suboptimal association, see [3, 4], of a modified correlation filter (matched to the target template) and a linearized KBf. Figure 2(b) plots the standard deviation over time of the error in the vertical position estimate for both trackers, with PSNR equal to 6 dB. We see from the plot that the KBf tracker has higher initial and steady state position estimate errors and a longer target acquisition time.

5. DETECTION PERFORMANCE

We study in the sequel the detection performance of the Bayes algorithm. As before, we simulate 9 x 9 constant signature targets buried in first order GMrf clutter with $\beta_h = \beta_v = 0.24$. Simulated targets move in and out of a 100 x 100 grid so that, at any given sensor frame, the target may be either present or absent. The motion dynamics of a target that is actually present is described by a first order noncausal 2D random walk with constant mean velocity, like the model we used in section 4. Once a target moves out of the sensor grid, a new target can appear randomly with equal probability in any position of the 100 x 100 grid. In addition to tracking, the Bayes algorithm now also makes detection decisions at each frame using the binary test (9).

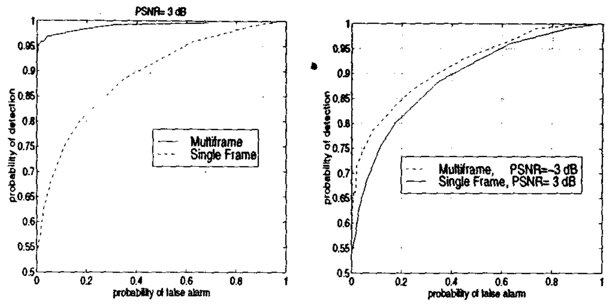
We compare the optimal multiframe detector with a single frame LRT detector that ignores the motion model. The single frame LRT algorithm reduces to the test

$$p(\mathbf{y}_n \mid H_0) \underset{H_1}{\overset{H_0}{\gtrless}} \lambda p(\mathbf{y}_n \mid H_1) \qquad (12)$$

where λ is a threshold that varies according to the desired probability of false alarm. Using a fixed value of PSNR, we vary the thresholds in tests (9) and (12) to plot the receiver operating characteristic curves (ROCs) for both detectors. The ROC curves, estimated from 3,000 Monte Carlo runs, are shown in figure 3(a) for PSNR = 3 dB. The curves in figure 3(a) show that, while the performance of the single frame detector deteriorates significantly in scenarios of low PSNR, the multiframe detector still achieves probabilities of detection around 95 % for false alarm rates of 10^{-2}. Figure 3(b) repeats the ROC curves for both algorithms, except that the PSNR for the multiframe algorithm is lowered to -3 dB. We note from the plot that the -3 dB multiframe ROC curve is closer to the 3 dB single frame ROC, but still lies slightly above the latter. That indicates a substantial gain in PSNR of over 6 dB when we use multiple frames in the detection process.

6. SUMMARY

We presented in this paper an optimal Bayesian algorithm for multiframe detection and tracking of extended targets in a sequence of 2D digital images. Performance studies using Monte Carlo simulations show that there is a significant improvement over existing trackers such as the usual association of a correlation filter and a linearized Kalman-Bucy filter. We also show that there is a substantial detection performance improvement over the single frame LRT detector.



(a) (b)

Figure 3: Single frame vs multiframe ROCs in correlated clutter: (a) PSNR= 3 dB (both detectors); (b) PSNR= 3 dB (single frame), PSNR= - 3 dB (multiframe)

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.

7. REFERENCES

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

[2] A. P. Tzannes. Detection of point targets in image sequences by hypothesis testing: a temporal test first approach. *Proceedings IEEE ICASSP' 99*, IMDSP7.9, vol.6, pp. 3377-3380.

[3] M. G. S. Bruno and J. M. F. Moura. Multiframe detection and tracking in clutter: optimal performance. Submitted to *IEEE Trans. on Aerosp. and Electr. Systems*, under review.

[4] M. G. S. Bruno and J. M. F. Moura. The optimal 2D multiframe detector/tracker. *AEÜ Int. J. Electron. Commun.*, December 1999.

[5] J. M. F. Moura and M. G. S. Bruno. DCT/DST and Gauss-Markov Fields: Conditions for Equivalence. *IEEE Trans. on Signal Processing*, vol. 46, n.9, pp. 2571-2574, September 1998.

[6] J. M. F. Moura and N. Balram. Recursive Structure of Noncausal Gauss Markov Random Fields. *IEEE Trans. on Inf. Theory* IT-38(2), pp. 334-354, March 1992.