A Particle Filter Algorithm for Target Tracking in Images

Marcelo G. S. Bruno
Instituto Tecnológico de Aeronáutica, São José dos Campos SP, Brazil

Abstract—We present in this paper a new algorithm for target tracking in cluttered image sequences using the bootstrap particle filter. The proposed algorithm incorporates the models for target signature, target motion and clutter correlation and allows for direct tracking from the image sequence. Monte Carlo simulation results show that the bootstrap tracker outperforms the association of a single frame maximum likelihood position estimator and a Kalman-Bucy filter in a scenario with a heavily cluttered, dim target.

I. INTRODUCTION

We introduced in [1] a Bayesian algorithm for automatic target tracking in digital image sequences. In scenarios of heavily cluttered targets, the Bayes tracker was shown to outperform the association of an image correlator and a Kalman-Bucy filter that was previously proposed in the literature, see [2]. However, we assumed in [1] a discrete-time, discrete-valued target motion model where the target centroid position was allowed to take values only on a finite grid. In many real-world situations, such modeling strategy may prove too restrictive to describe the actual target dynamics. Unlike in [1], we use in this paper an alternative discrete-time, but continuous-valued motion model where the unknown target centroid position and velocity in both dimensions of the plane are allowed to be real numbers. To solve the Bayesian estimation problem for this discrete-time, continuous-valued state space model, we resort then to sequential importance sampling [3].

Sequential importance sampling (SIS) filtering, also known as particle filtering, is a recursive Monte Carlo simulation approach where the desired posterior probability density function (pdf) of the hidden state vector is represented at each instant by a set of samples (or particles) with associated importance weights. The particles are drawn sequentially from a proposed importance function while the importance weights are updated recursively at each time step. A selection (or resampling) step [5] is added to prevent the distribution of importance weights from getting skewed as time increases. In recent years, particle filters have been successfully applied to several practical problems including mobile robot localization [7], computer vision [8], terrain navigation [9], and many others.

II. TARGET MOTION 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

\[
x_i(t) = \left[ x_{i}(t) \quad \frac{dx_{i}(t)}{dt} \right]^T, \quad t \in \mathbb{R}
\]

be a continuous-time state vector that collects the position and velocity of the target centroid at instant \(t\) in dimension \(i\). Let now

\[
x_{n,i} = x_i(n \Delta) \quad n \in \mathbb{Z} \quad i = 1, 2
\]

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

\[
x_n = \left[ x_{n,1}^T \quad x_{n,2}^T \right]^T
\]

where \(x_{n,i}, i = 1, 2\), are defined as in (2). We assume further that the centroid motion in dimension \(i = 1\) is statistically independent of the motion in dimension \(i = 2\) and that
the system input noise is Gaussian, zero mean, and identically distributed in both dimensions. The white noise acceleration motion model is given by [2]

\[
x_{n+1} = \begin{bmatrix} F & 0 \\ 0 & F \end{bmatrix} x_n + \begin{bmatrix} v_{n,1} \\ v_{n,2} \end{bmatrix}
\]

(4)

where

\[
F = \begin{bmatrix} 1 & \Delta \\ 0 & 1 \end{bmatrix}
\]

(5)

and

\[
E \left[ v_{n,i} v_{n,j}^T \right] = q \begin{bmatrix} \Delta^2 & 0 \\ 0 & \Delta^2 \end{bmatrix} \delta_{n-l, i-j}
\]

(6)

In (6), \(E[\cdot]\) denotes expected value or ensemble average, \(q\) is a positive real number and \(\delta_{n,s}\) is the 2D discrete Dirac delta function such that \(\delta_{n,s} = 1 \text{ if } (r,s) = (0,0)\) and zero otherwise.

A. Markovian Transition Kernel

From the assumption of statistical independence of the target motion in each coordinate, it follows that the Markovian transition kernel, \(p(x_{n+1} \mid x_n)\), factors as

\[
p(x_{n+1} \mid x_n) = p(x_{n+1,1} \mid x_{n,1}) p(x_{n+1,2} \mid x_{n,2})
\]

(7)

On the other hand, from (4), we write

\[
p(x_{n+1,1} \mid x_{n,1}) = N(F x_{n,1}, Q)
\]

(8)

where \(N(m, R)\) denotes the multivariate normal (Gaussian) distribution with mean \(m\) and covariance matrix \(R\), and matrix \(Q\) is defined as in (6).

III. OBSERVATION AND CLUTTER MODEL

A remote sensing device, e.g., an infrared airborne radar, sequentially generates raw sensor measurements of a given surveillance region that contains both targets of interest and undesired spurious reflectors (clutter). For simplicity, we assume in this paper that there is only one single target of interest present at the scene at each sensor scan. The raw sensor measurements at instant \(n\) are sampled and processed to form a 2D digital sensor image, referred to as a frame. Frame \(n\) is modeled by the \(L \times M\) matrix

\[
Y_n = H(x_n^*) + V_n
\]

(9)

where matrix \(V_n\) represents the background clutter, and matrix \(H(x_n^*)\) is the clutter-free target image, which is a function of the 2D pixel location of the target centroid, \(x_n^*\). The two-dimensional hidden vector \(x_n^*\) is defined on the finite sensor grid \(L = \{(i,j) \mid 1 \leq i \leq L, 1 \leq j \leq M\}\) and is obtained from the four-dimensional continuous-valued state vector \(x_n\) in (3) by making

\[
x_n^*(1) = \text{round}\left(\frac{x_n(1)}{\Delta_1}\right)
\]

(10)

\[
x_n^*(2) = \text{round}\left(\frac{x_n(2)}{\Delta_2}\right)
\]

(11)

where \(\Delta_1\) and \(\Delta_2\) are the image resolutions respectively in dimensions \(i = 1\) and \(i = 2\).

Target Model We assume that, any given frame, the clutter-free target image 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 \((i,j) \in L\), the nonlinear function \(H\) in (9) returns a spatial distribution of (real-valued) pixel intensities \(\{a_{k,l}\}, -r_i \leq k \leq r_s, -l_i \leq l \leq l_s\) centered at \((i,j)\). Formally, we write

\[
H(i_n, j_n) = \sum_{k=-r_i}^{r_i} \sum_{l=-l_i}^{l_i} a_{k,l} E_{n+k,j_n+l}
\]

(12)

where \(E_{r,s}\) is an \(L \times M\) matrix whose entries are all equal to zero, except for the element \((r,s)\) which is equal to 1. The coefficients \(\{a_{k,l}\}\) in (12) are referred to as the target signature parameters. As a first approximation, we assume in this paper that the signature parameters are deterministic, known and frame-invariant.

Remark To write (12), 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 defining the extended image grid \(\bar{L} = \{ (i,j) \mid -r_s + 1 \leq i \leq L + r_i, -l_s + 1 \leq j \leq M + l_i \}\) and changing the summation limits accordingly in (12) for centroid locations near the borders, see [1] for details.

Clutter Model We capture the 2D spatial correlation of the background clutter using a noncausal, spatially homogeneous Gauss-Markov random field (GMrf) model [10]. 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

\[
V_n(i, j) = \beta_n^c [V_n(i-1, j) + V_n(i+1, j)] + \beta_n^c [V_n(i, j-1) + V_n(i, j+1)] + U_n(i,j)
\]

(13)

where \(E[V_n(i,j) U_n(p, r)] = \sigma_n^2 \delta_{i-p, j-r}\). The assumption of zero-mean clutter implies a pre-processing of the data that subtracts the mean of the background. We also assume that, after the removal of the background mean, the clutter frames \(\{V_n\}\) are statistically independent.
A. Likelihood Function

Let $y_n$ be a 1D long-vector representation of the frame $Y_n$ obtained by either row or columnwise scanning. Assuming a 2D GMrf background as in (13) and deterministic signature parameters $\{a_{k,l}\}$, we use the results in [1] to write the likelihood function of the observed $n$th frame as

$$p(y_n | x_{n,1}, x_{n,2}) \propto \exp \left( \frac{2\lambda(x_{n,1}, x_{n,2}) - \rho}{2\sigma^2} \right).$$

where $\rho$ is a target energy term that is constant away from the image borders, see [1] for details. The function $\lambda$ in (14) is in turn given by [1]

$$\lambda(x_{n,1}, x_{n,2}) = \sum_{k=-r}^{r} \sum_{l=-l}^{l} a_{k,l} \mu(x_n^*(1) + k, x_n^*(2) + l)$$

where $x_n^*(i), i = 1, 2,$ are obtained respectively from equations (10) and (11), and $\mu(p, r)$ is the output of the differential operator

$$\mu(p, r) = Y_n(p, r) - \beta_n Y_n(p, r - 1) + Y_n(p, r + 1) - \beta_n^2 Y_n(p - 1, r) + Y_n(p + 1, r)$$

with Dirichlet (identically zero) boundary conditions. Equation (15) 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 (15) must be varied accordingly as explained in [1]. Intuitively, we can interpret the function $\lambda$ in (15) as the concatenation of two linear filtering operations: first, we pass the image through a noncausal differential filter to generate a residual error image and, then, we apply to this error image a 2D correlation filter that is matched to the (known) 2D target template.

IV. PARTICLE FILTER TRACKER

The on-line Bayesian estimation problem can be summarized as follows: given a sequence of observed frames $Y_1^n = \{y_1, y_2, \ldots, y_n\}$ and the probability density function (pdf) $p(x_0)$ of the initial (continuous-valued) target state, $x_0$, compute recursively the posterior pdf $p(x_n | Y_1^n)$, for $n \geq 1$, using the Markovian transition kernel, $p(x_{n+1} | x_n)$, and the likelihood function, $p(y_n | x_n)$. From the posterior pdf, we can then infer the value of the hidden (unobserved) state at instant $n$, $x_n$, using some optimality criteria, e.g. minimum mean-square error (MMSE) or maximum a posteriori (MAP).

Particle filters [3] are a simulation approach to Bayesian estimation in which the posterior pdf is represented at each instant $n$ by a set of particles $\{x_n^{(j)}\}, 1 \leq j \leq N_p$, with associated weights $w_n^{(j)}$. The MMSE estimate of the hidden state $x_n$ is then obtained simply as [3] a weighted average of the particles. Alternatively, the MAP estimate can be obtained [4] from the histogram of the particles.

Ideally, we would like the particles to be samples from the true posterior in which case all weights $w_n^{(j)}$ would be identical and equal to $1/N_p$. In practice, however, either the posterior pdf is unavailable or difficult to sample from. An alternative approach known as sequential importance sampling (SIS) [4] is to sample $x_n^{(j)}$ sequentially from an importance function $\pi(x_n | X_0^{n-1}, Y_1^n)$ and update the weights using the recursion

$$w_n^{(j)} \propto w_{n-1}^{(j)} p(y_n | x_n^{(j)}) p(x_n^{(j)} | X_0^{n-1}, Y_1^n), \quad \sum_{j=1}^{N_p} w_n^{(j)} = 1.$$  

In principle, the importance function $\pi$ can be arbitrary provided that it satisfies two restrictions [4]

1. $\pi(x_n | X_0^{n-1}, Y_1^n)$ must have the same support as $p(X_0^n | Y_1^n)$ and be strictly positive and integrable to 1 in that support.
2. $\pi(x_n | X_0^{n-1}, Y_1^n) = \pi(x_n | X_0^{n-1}, Y_1^n)\pi(x_n | X_0^{n-1} | Y_1^n).$

A. Bootstrap Tracker

The first practical SIS filter was the bootstrap filter [5], where the Markovian transition kernel $p(x_{n+1} | x_n)$ is used as the importance function $\pi(x_n | X_0^{n-1}, Y_1^n).$ The weight update equation reduces then to

$$w_n^{(j)} \propto w_{n-1}^{(j)} p(y_n | x_n^{(j)}) \sum_{j=1}^{N_p} w_n^{(j)} = 1.$$  

The key innovation of the bootstrap filter [5] was to introduce an additional particle selection step to avoid the so-called degeneracy phenomenon [3], i.e., the tendency of the distribution of the particle weights to become skewed as the number of SIS iterations increases. The selection step proposed in [5] consisted simply of resampling $N_p$ times from the set $\{x_n^{(j)}\}$ with replacement according to the particle weights so that low-weight particles are discarded whereas high-weight particles are multiplied. The new weights after the resampling step are all reset then to $1/N_p$.

Using the motion, target and clutter models from sections II and III and recalling the expressions for the Markovian transition kernel in subsection II-A and for the likelihood function in subsection III-A, we present in Table I a bootstrap filter algorithm for target tracking in image sequences.

Clutter Adaptation When the clutter model parameters $\beta_n$, $\beta_n^2$ and $\sigma^2$ are unknown, they must be estimated from the observed data. For computational simplicity, we use in this paper a suboptimal approach to clutter adaptation where we estimate the GMRF clutter parameters directly from each available sensor frame $Y_n$ using a variation of approximate maximum likelihood (AML) parameter estimation algorithm introduced in [10]. Table II summarizes the AML parameter estimation algorithm given the $L \times M$ $n$th sensor frame $Y_n$, see [10] for further details.
Algorithm I: Bootstrap filter for target tracking in 2D cluttered image sequences.

V. LINEARIZED KALMAN-BUCY TRACKER

We compare the proposed nonlinear bootstrap tracker to the association of a single frame maximum likelihood (ML) position estimator and a linear Kalman-Bucy filter. The preliminary ML estimates of the 2D pixel location of the target centroid are given by

\[ \tilde{x}_n = \arg \max_{x_n} p(y_n | x_n), \]  

(19)

The pixel estimates of the centroid location are converted to continuous space and treated as decoupled noisy measurements of the true centroid positions in each dimension. A conventional Kalman-Bucy filter with knowledge of the state dynamics is then used to refine the preliminary ML estimates. Table III summarizes the algorithm. In Table III, \( \sigma^2 \) denotes the variance of the ML centroid position estimation error in the dimension \( i \) and \( C \) is the row vector \( C = [1, 0] \). We denote the mean and covariance matrix of \( x_{0,i} \), \( i = 1, 2 \), respectively by \( \bar{x}_{0,i} \) and \( \Sigma_{0,i} \). The symbols \( \bar{x}_{n,i} \) and \( \Sigma_{n,i} \) denote respectively the Kalman filter estimate of the target state \( x_{n,i} \) at instant \( n \) in dimension \( i \) and its associated error covariance matrix.

VI. PERFORMANCE RESULTS

We compare next the tracking performances of the bootstrap tracker and the KBF tracker using a simulated image sequence that was generated from real infrared airborne radar (IRAR) intensity imagery. The base image, shown in Figure 1, is from the Lincoln Laboratory’s IRAR database and was obtained from the Johns Hopkins University’s Center for Imaging Sciences. In order to simulate the clutter sequence, we extracted the spatially-variant local mean along the base image and fitted a first-order GMrf model to the background by estimating the clutter parameters \( \beta^b_n, \beta^c \) and \( \sigma^c \), 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 generated using the estimated background clutter parameters. Finally, we add to the background sequence a simulated target template that moves according to

<table>
<thead>
<tr>
<th>TABLE I</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALGORITHM I: Bootstrap filter for target tracking in 2D cluttered image sequences.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II</th>
</tr>
</thead>
<tbody>
<tr>
<td>AML parameter estimation algorithm for an ( L \times M ) GMrf.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm II: Association ML estimator + Linear Kalman-Bucy Filter (for comparison purposes only).</td>
</tr>
</tbody>
</table>
a white noise acceleration model, see section II, with parameters $q = 10$ and $\Delta = 4\text{ms}$. The spatial resolution (pixel size) is $\Delta_1 = \Delta_2 = 20\text{cm}$. The image frame extends from 0 to 30 meters (150 pixels) in both the horizontal and vertical dimensions. The target’s initial vertical and horizontal positions are uniformly distributed respectively between 4 and 12 meters, and between 4 and 8 meters. The initial target velocity is $10\text{ m/s}$ in both dimensions. Figures 2(a) and (b) show respectively the simulated cluttered and clutter-free image of a target centered at pixel location $(40; 40)$. The peak target-to-clutter ratio (PTCR) in Figure 2(a) is $7.3\text{ dB}$. Figures 3 and 4 show the root mean-square error (RMSE) in meters of the MAP target centroid position estimates for a 3400-particle bootstrap tracker, respectively in the vertical and horizontal directions. The error curves were obtained from 45 Monte Carlo runs with PTCR = $-5.7\text{ dB}$. We see from the plots in Figures 5 and 6 that the association ML/KBf has a very poor tracking performance, basically failing to acquire and track the target. Conversely, despite the low PTCR and poor visibility of the target, the proposed nonlinear bootstrap filter tracks the simulated vehicle with a final tracking error after 14 frames of less than 1 pixel within the resolution of the image.

VII. CONCLUSIONS

We introduced in this paper a new bootstrap particle filter algorithm for direct target tracking from image sequences. The performance of the algorithm was evaluated using Monte Carlo simulations. The simulation results show good tracking performance for the proposed tracker using 3400 particles and 14 image frames in a scenario of low target-to-clutter ratio. In contrast to the nonlinear particle filter, a linearized Kalman-Bucy tracker fails to acquire and track the target under the same simulation conditions.
ACKNOWLEDGEMENT

This work was partially supported by FAPESP, São Paulo, Brazil.

REFERENCES


