Survey Paper

Tracking Methods in a Multitarget Environment

YAAKOV BAR-SHALOM, MEMBER, IEEE

Abstract—**The objective of this paper is to survey and put in perspective the existing methods of tracking in multitarget environment. In such an environment the origin of the measurements can be uncertain: they could have come from the target(s) of interest or clutter or false alarms or be due to the background. This compact and unified presentation of the state-of-art in multitarget tracking was motivated by the recent surge of interest in this problem. It is also hoped to be useful in view of the need to adapt and modify existing techniques before using them for specific problems. Particular attention is paid to the assumptions underlying each algorithm and its applicability to various situations.**

I. Introduction

IT WAS recognized as far back as 1964 [26] that in tracking targets there can be an uncertainty associated with the measurements in addition to their inaccuracy, which is usually modeled by some additive noise. This additional uncertainty is related to the origin of the measurements: a measurement, which is to be used in the tracking algorithm, might not have originated from the target of interest. This situation can occur in a surveillance system when a sensor such as a radar, sonar, or optical one is operating in an environment in which there is clutter, or the false-alarm rate is high. It can also happen when several targets are in the same neighborhood and one cannot associate with certainty the observed detections, (assumed to have been resolved) which yield the measurements, with the various targets. A similar situation can occur in the track formation problem when there are several targets but their number is not known and some of the measurements might be spurious. The application of standard estimation algorithms which would use the measurement nearest in some sense to the predicted measurement ("nearest neighbor filter") can lead to very poor results in an environment where spurious measurements occur frequently. This is because such an algorithm does not account for the fact that the measurement used in the filter might have originated from a source different from the target of interest.

Recently, with the proliferation of surveillance systems and their increased sophistication, there has been a great deal of interest in the multitarget tracking problem. Several algorithms that have been published in the literature point to the current activity in this area. The purpose of this paper is to survey and put in perspective the recent work on multitarget tracking and present in more detail the most important algorithms. The (subjective) criterion of what is important is based upon what is believed to be of general interest and potential usefulness for application to real problems. Particular attention is given to the assumptions underlying the various algorithms and to what problems they can be applied to.

The pioneering work of Sittler [26] was motivated by the need to find a reasonable way of incorporating measurements of uncertain origin into existing tracks. The estimation algorithm he considered was of the type used before the Kalman filter became popular. The method consisted of splitting the track whenever more than one return (detection) was observed in the neighborhood of the predicted measurement. Then the likelihood function of each trajectory was computed and those whose likelihood was below a threshold were dropped. Similar approaches were developed within the framework of Kalman filtering by Fraser and Meier [10] for active sonar tracking and by Smith and Buechler [27] for radar tracking. This track split or branching algorithm, described in Section II, is suitable both for track initiation or track update but its computational and memory requirements can become very large in dense environments.

A method particularly suitable for track formation for several targets in the same neighborhood was developed by Morefield [18]. This method, discussed in Section III, is also based upon likelihood functions and converts the association of measurements to form tracks into an integer programming problem.

Both of the above methods are essentially non-Bayesian: they make decisions to accept or reject trajectories and then estimate the state conditioned upon the correct-

Manuscript received April 25, 1977; revised March 9, 1978. Paper recommended by R. Monopoli, Past Chairman of the Adaptive, Learning Systems, Pattern Recognition Committee. This research was supported in part by NSF Grant ENG 77-08177.

The author is with the University of Connecticut, Storrs, CT 06268.

0018-9286/78/0800-0618$00.75

ness of these decisions. The resulting state estimates and covariances do not account for the possibility that these decisions can be incorrect. Bayesian methods for estimation in dynamic systems where the measurement was either the state plus noise or (zero-mean) noise only were developed for several models by Nahi [19] and Jaffer and Gupta [13], [14]. Ackerson and Fu [1] and Akashi and Kumamoto [4] studied the problem of estimation in linear systems in a switching environment where the noise statistics evolved according to a Markov chain. These methods are related, but only conceptually, to the problem of tracking with measurement origin uncertainty so they will not be discussed in detail here.

Bayesian procedures were first developed for tracking problems with measurement origin uncertainty by Singer, Sea, and Stein [22]–[24]. They proposed using the "nearest neighbor" of the predicted measurement and modified the (Kalman) filter to account for the *a priori* probability that this measurement might be spurious. Subsequently, Jaffer and Bar-Shalom [15] developed the nearest neighbor modified filter with the *a posteriori* probability of the observation being incorrect by using its location relative to the predicted measurement.

The need to incorporate all the observations lying in the neighborhood of the predicted measurement was pointed out by Bar-Shalom and Jaffer in [5] where a suboptimal algorithm using *a posteriori* probabilities was presented. The final version of this algorithm was developed by Bar-Shalom and Tse [6]. An important feature of this algorithm, discussed in Section IV, is that its computational and memory requirements are not increasing with time and only slightly exceed those of the standard Kalman filter. Singer, Sea, and Housewright [25] presented the full-blown "all neighbors" approach and obtained the optimal algorithm which involves splitting the tracks all the way back to the initial time and recombining them. Since this involves a growing memory, suboptimal versions that split the track only for the last N scans and thus prevent the memory requirements from growing were also presented. The case $N=0$, i.e., no splitting, coincides with the algorithm of [6]. This is discussed in Section V. Since the *a priori* and nearest-neighbor algorithms [22]–[24], [15] have been superseded by the all-neighbors *a posteriori* algorithms [6], [25] they are not discussed here. Athans *et al.* [3] developed a tracking algorithm similar to the one of [6] for the specific problem of tracking a ballistic missile in the presence of wake noise, which can cause several returns to be detected by the radar. Their filter had also an adaptive feature consisting of real-time estimation of the measurement noise intensity due to the wake. This was used to modify the model based upon which state estimation was done with an extended Kalman filter. Because of the specialized nature of this algorithm it is not discussed here in more detail.

Both [6] and [25] deal with the case where there is one target of interest and several returns are observed. A generalization of the results of [6] to the case of several targets of interest and an arbitrary number of returns has been presented in [7]. Alspach [2] treated the case where there are m targets of interest and m returns. The recent unpublished work of Reid [21][1] treats the problem of tracking several targets of interest and at the same time track initiation. These methods are briefly discussed in Section VI, where a comparison of the main algorithms is also given.

Some of the mathematical developments are carried out in sufficient detail to emphasize the underlying assumptions and allow their modification for possible future specific applications.

II. The Track Splitting Approach [10], [27]

The state of a target is assumed to evolve in time according to the equation

$$x_{k+1} = F_k x_k + w_k \tag{2.1}$$

and the corresponding measurement is given by

$$z_k = H_k x_k + v_k \tag{2.2}$$

with w_k and v_k being zero-mean mutually independent white Gaussian noises with (known) covariances Q_k and R_k, respectively.

Tracks are usually initialized based upon one or two measurements. Then, for every subsequent observation that falls in a "window" around the location where the new measurement for an existing track is expected, the track is split. A Kalman filter is used to estimate the state of each split track. Since the number of branches can grow exponentially the likelihood function of each split track is computed and the unlikely ones are discarded.

A particular (say, the lth) sequence of measurements up to k is denoted as

$$\theta^{k,l} = \{ z_{1,i_1}, z_{2,i_2}, \cdots, z_{k,i_k} \} \tag{2.3}$$

where $z_{j,i}$ is the ith measurement at time j.

The likelihood function of this sequence of measurements conditioned upon their having originated from the same target can be written in terms of the innovations [16] as follows (the second index is dropped for simplicity of notation)

$$\begin{aligned} \Lambda(\theta^k) &= p(z_1, \cdots, z_k | \theta^k \text{ is a correct track}) \\ &= p(\nu_1, \cdots, \nu_k) = \prod_{i=1}^{k} p(\nu_i) \\ &= c_k \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \nu_i' S_i^{-1} \nu_i \right\} \end{aligned} \tag{2.4}$$

where c_k is a constant, ν_i is the innovation corresponding to z_i when it is assumed to belong to track θ^k and S_i is the innovation covariance matrix obtained from the standard

[1]This was brought to the author's knowledge after this paper was already accepted for publication.

Kalman filter. The probability of detection is assumed to be unity.[2]

The negative log-likelihood function (with the constant omitted) can be easily computed in a recursive form

$$\lambda(\theta^k) = -\log \Lambda(\theta^k) = \lambda(\theta^{k-1}) + \frac{1}{2} \nu_k' S_k^{-1} \nu_k. \quad (2.5)$$

Assuming all the noises and the initial state to be Gaussian, the innovations, which are m-vectors are also Gaussian and white. Then $\lambda(\theta^k)$ will have a chi-square distribution with km degrees of freedom. Tracks for which λ exceeds a certain threshold can be dropped.

This algorithm is of the sequential type and can be used in an environment where the number of targets is arbitrary and unknown. The incorrect measurements, i.e., those that did not originate from any target, can be either fixed or random. The state estimates and covariances are obtained by a standard Kalman filter based on an assumed sequence of measurements; however no probability that a sequence is correct can be obtained. The main problem with this algorithm is that, as mentioned earlier, its computational and memory requirements can grow with time and saturate even large computing systems.

III. The Maximum Likelihood Method [18]

In this approach which yields the most likely set of trajectories it is assumed that there are s targets and each target's state evolves according to the same equation (2.1).[3] The corresponding measurement from a target is given by (2.2); however, when a set of measurements is obtained, one does not know with certainty which measurement originated from which target, if any. The set of measurements obtained at time k is

$$Z_k = \{z_{k,i}\}_{i=1}^{m_k} \quad (3.1)$$

and the set of measurements up to and including time k is denoted as

$$Z^k = \{Z_j\}_{j=1}^{k}. \quad (3.2)$$

Measurements not having originated from a target are assumed to be i.i.d. (independent identically distributed) with uniform density V^{-1} where V is the hypervolume in which the observations are made.

Since this procedure is essentially of the batch processing type (even though it could be modified for sequential processing) the time index k will be omitted for simplicity where it will not create confusion and Z will denote the set of all measurements through the present time. The number of elements in this set is

$$N = \sum_{j=1}^{k} m_j. \quad (3.3)$$

As in (2.3), θ^l will be a sequence of measurements up to the present time, consisting of N_l observations.

The procedure then consists of the following steps.

A. Construction of Feasible ("Not Too Unlikely") Tracks

This is done to reduce the number of tracks to be considered in the search for the most likely set of tracks. These tracks are obtained by associating sequentially returns by a simple window procedure that carries out a coarse test. The result consists of tracks that are "reasonable" to be kept for subsequent consideration. The set of feasible tracks is denoted as

$$\Phi = \{\theta^l\}_{l=1}^{L}. \quad (3.4)$$

B. Feasible Partitions of Z

Define the set of tracks

$$\tau = \{\theta^{l_i}\}_{i=0}^{I} \quad (3.5)$$

where θ^{l_i} is a feasible track for $i > 0$ and θ^{l_0} consists of the measurements not associated in this partition with any of the tracks θ^{l_i}, $i = 1, \cdots, I$. The set τ is a feasible partition of Z if

$$Z = \bigcup_{i=0}^{I} \theta^{l_i} \quad (3.6)$$

and

$$\theta^{l_i} \cap \theta^{l_j} = \varnothing \qquad \forall i \neq j, \quad (3.7)$$

i.e., a measurement cannot belong to more than one track. The set of all feasible partitions of Z is denoted as

$$T = \{\tau^j\}_{j=1}^{M}. \quad (3.8)$$

C. The Optimization Criterion

The posterior probability of one such partition τ being correct is

$$P\{\tau | Z\} = \frac{p(Z|\tau) P\{\tau\}}{p(Z)} \quad (3.9)$$

where $P\{\tau\}$ is the prior probability of the partition τ. In the absence of target signature or other prior information one usually assumes the priors to be equal for all feasible partitions. In this case searching for the most likely trajectory based upon (3.9) coincides with maximizing the likelihood function of the measurements over all the feasible partitions, i.e.,

$$\max_{\tau \in T} p(Z|\tau). \quad (3.10)$$

[2]Assuming it to be less than one entails an additional split track at every sampling time.

[3]The method, as presented in [18], can be applied in principle to the case where each target is described by a different difference equation. Since this complicates the notations (and implementation) substantially, only the simpler case is discussed here for clarity.

D. Integer Programming Formulation

The negative log-likelihood of each feasible track is computed according to (2.5). The following modified log-likelihood function is then defined

$$\lambda_l = \lambda(\theta^l) - N_l \log V^{-1} \qquad l = 1, \cdots, L \tag{3.11}$$

where N_l is the number of points in track θ^l and V is the hypervolume in which the observations are made. The last term in (3.11) is introduced in order to account for the returns not associated with any track, as will be seen in the sequel. These returns are assumed i.i.d. with uniform density V^{-1}. The vector $\boldsymbol{\lambda}$ is defined as

$$\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_L]'. \tag{3.12}$$

Let ψ be a binary vector of dimension N, the total number of measurements in the set Z. This vector indicates with units in the corresponding elements which measurements belong to track l. Let $\boldsymbol{\rho}$ be a binary vector of dimension L that indicates which of the feasible tracks from (3.4) belong to a certain feasible partition. Define the following $N \times L$ matrix

$$A \triangleq [\psi^1, \cdots, \psi^L]. \tag{3.13}$$

Then condition (3.7) that a partition be feasible can be expressed as

$$A\boldsymbol{\rho} \leqslant \mathbf{1} \tag{3.14}$$

where $\mathbf{1}$ is an N-vector of ones. Then, since

$$\boldsymbol{\rho}'\boldsymbol{\lambda} = \sum_{l=1}^{L} \rho_l \lambda_l = \sum_{l=1}^{L} \rho_l \lambda(\theta^l) - \sum_{l=1}^{L} \rho_l N_l \log V^{-1} \tag{3.15}$$

and

$$\begin{aligned} -\log p(Z|\tau) &= \sum_{l=1}^{L} \rho_l \lambda(\theta^l) + \lambda(\theta^0) \\ &= \sum_{l=1}^{L} \rho_l \lambda(\theta^l) + \left(N - \sum_{l=1}^{L} \rho_l N_l \right) \log V^{-1} \end{aligned} \tag{3.16}$$

it follows that

$$\boldsymbol{\rho}'\boldsymbol{\lambda} = -\log p(Z|\tau) + N \log V^{-1} \tag{3.17}$$

where the last term is a constant.

Therefore maximization of the likelihood function as in (3.16) is equivalent to

$$\min_{\boldsymbol{\rho}} \boldsymbol{\rho}'\boldsymbol{\lambda} \tag{3.18}$$

subject to the feasibility condition (3.14) and the fact that $\boldsymbol{\rho}$ is a binary vector. Thus the track formation has been converted into an integer programming problem. While this is not a trivial computational matter, the feasibility of this method was illustrated in [18] for 20 targets with 4-point tracks, i.e., total number of measurements $N = 80$. No false measurements were generated in these simulations and the probability of detection was unity. Less than unity probability of detection would require using the "absent return" hypothesis at each time when forming the feasible tracks.

Once the most likely set of trajectories has been chosen from (3.18), the state estimates and covariances are computed in this method from a corresponding set of standard Kalman filters.

IV. A Suboptimal Bayesian Approach: The Probabilistic Data Association Filter [5], [6]

The methods discussed in the previous two sections are based on the likelihood function rather than Bayesian approach. The first one keeps all the "reasonably likely" tracks, the second yields the most likely set of tracks. While they are capable, at least in principle, of handling an arbitrary (and unknown) number of targets, *the estimates and covariances they yield are conditioned on the "accepted" tracks being true*. The algorithms of this and the next section incorporate sequentially into a track clusters of measurements by attaching to each a probability of being correct. Such an approach yields *estimates and covariances that account for the measurement origin uncertainty*.

It is assumed that there is only one target of interest, whose track has been already initiated, which is observed in the presence of additional measurements that can originate from clutter or other targets. The incorrect measurements are assumed to be i.i.d. random variables with uniform spatial distribution. The target of interest and the measurements on it are described by (2.1) and (2.2), respectively.

The notation Z_k, defined in (3.1), will be used here for the set of measurements lying in a neighborhood of the predicted location of the observation from the target. As in (3.2), the accumulated set of these measurements is denoted by Z^k. The best estimate of the target's state is the conditional mean based upon all the observations that with some nonzero probability originated from the target, i.e.,

$$\begin{aligned} \hat{\boldsymbol{x}}_{k|k} = E\{\boldsymbol{x}_k|Z^k\} &= \sum_{i=0}^{m_k} \beta_{k,i} E\{\boldsymbol{x}_k|\mathcal{X}_{k,i}, Z^k\} \\ &\triangleq \sum_{i=0}^{m_k} \beta_{k,i} \hat{\boldsymbol{x}}_{k|k,i} \end{aligned} \tag{4.1}$$

where $\mathcal{X}_{k,i}$ denotes the event that the ith measurement, $z_{k,i}$ is correct ($i = 1, \cdots, m_k$), $\mathcal{X}_{k,0}$ is the event that none of them is correct and

$$\beta_{k,i} \triangleq \boldsymbol{P}\{\mathcal{X}_{k,i}|Z^k\} \qquad i = 0, 1, \cdots, m_k. \tag{4.2}$$[4]

Equation (4.1), which is the essence of this Bayesian, "all-neighbors" approach, follows from the total probability theorem [20]. This approach does not allow track

[4]The number of observed measurements is subsumed in the conditioning.

splitting—at every time it combines all the latest measurements into the estimate.

In order to make the filter suitable for real-time implementation (fixed-memory and modest computational requirements) an approximation is made. The probability density of the state conditioned upon past observations is assumed normal with mean $\hat{x}_{k|k-1}$ and covariance $P_{k|k-1}$, i.e.,

$$p(x_k|Z^{k-1}) = \mathcal{N}(x_k; \hat{x}_{k|k-1}, P_{k|k-1}). \tag{4.3}$$

As it will be seen in the next section where the optimum Bayesian procedure is discussed, the above density is actually a mixture of normal densities. In view of (4.3), the set Z_k should be made up of the measurements lying in the "validation" region (window or confidence ellipsoid [9])

$$\{z : (z-\hat{z}_{k|k-1})' S_k^{-1} (z-\hat{z}_{k|k-1}) \leqslant d\} \tag{4.4}$$

where

$$S_k = H_k P_{k|k-1} H_k' + R_k \tag{4.5}$$

is the innovation covariance and d is a threshold from the chi-square tables with m degrees of freedom corresponding to a certain probability α_1, of rejecting the correct return.

Using Bayes' rule, the probabilities (4.2) can be written as [20]

$$\begin{aligned} P\{\mathcal{X}_{k,i}|Z^k, m_k\} &= P\{\mathcal{X}_{k,i}|Z_k, Z^{k-1}, m_k\} \\ &= c_k^{-1} p(Z_k|\mathcal{X}_{k,i}, Z^{k-1}, m_k) \\ &\quad \cdot P\{m_k|\mathcal{X}_{k,i}, Z^{k-1}\} P\{\mathcal{X}_{k,i}|Z^{k-1}\} \end{aligned} \tag{4.6}$$

where the conditioning on the total number of observed returns at time k being equal to m_k has now been written out explicitly, and c_k is a normalization constant.

Note that the probability of observing a *total* number of returns equal to m_k conditioned upon the event that none of the returns is correct is

$$P\{m_k|\mathcal{X}_{k,0}, Z^{k-1}\} = P\{I_k = m_k|Z^{k-1}\} \tag{4.7}$$

where I_k is the number of *incorrect* returns that are observed at time k. The probability of observing a total number of returns equal to m_k when one of them is correct is

$$P\{m_k|\mathcal{X}_{k,i}, Z^{k-1}\} = P\{I_k = m_k - 1|Z^{k-1}\} \quad i = 1, \cdots, m_k. \tag{4.8}$$

If no inference can be made on the number of incorrect returns from past data then the events $I_k = m_k$ and $I_k = m_k - 1$ are treated as having the same probability Π_0. The value of Π_0 is irrelevant because it cancels out in (4.6). Such a "diffuse" distribution, which is improper because its integral is infinite, is often used to describe vague prior knowledge; however, due to the above-mentioned cancellation the resulting posterior *is proper*. A detailed discussion of improper prior distributions can be found in De Groot [8]. Alternatively, a Poisson distribution can be used for I_k, as in [23]–[25]. In an environment where the density of false alarms is poorly known or may change unpredictably, the diffuse prior is an appropriate model. If the false alarm rate is known, the Poisson model is adequate.

For $i = 0$, i.e., when all the validated returns are incorrect ones, their joint density is

$$p(Z_k|\mathcal{X}_{k,0}, Z^{k-1}, m_k) = \prod_{i=1}^{m_k} p(z_{k,i}|\mathcal{X}_{k,0}, Z^{k-1}, m_k) = V_k^{-m_k} \tag{4.9}$$

where V_k is the volume of the validation region, because of our assumption on the incorrect measurements being uniformly distributed.

The probability of $\mathcal{X}_{k,0}$ based on past data is

$$P\{\mathcal{X}_{k,0}|Z^{k-1}\} = \alpha_1 + (1-\alpha_1)\alpha_2 \tag{4.10}$$

where α_1 is the probability that the correct return will not lie in the validation region and α_2 is the probability that the correct return will not be detected.

For $i = 1, \cdots, m_k$ the density on the r.h.s. of (4.6) is

$$p\{Z_k|\mathcal{X}_{k,i}, Z^{k-1}, m_k\} = f_k(z_{k,i}) V_k^{-m_k+1} \tag{4.11}$$

where

$$f_k(z_{k,i}) \triangleq (1-\alpha_1)^{-1} \mathcal{N}(z_{k,i}; \hat{z}_{k|k-1}, S_k) \tag{4.12}$$

is a truncated normal density which is zero outside the validation region.

The probability of $\mathcal{X}_{k,i}$ conditioned upon past data is assumed the same for all i, unless target signature information can be used. If no such information is available, then

$$\begin{aligned} P\{\mathcal{X}_{k,i}|Z^{k-1}\} &= \frac{1 - P\{\mathcal{X}_{k,0}|Z^{k-1}\}}{m_k} \\ &= \frac{(1-\alpha_1)(1-\alpha_2)}{m_k} \qquad i = 1, \cdots, m_k. \end{aligned} \tag{4.13}$$

Inserting (4.9–4.12) into (4.6) yields the equations of the PDA (probabilistic data association) method,

$$\beta_{k,i} = f_k(z_{k,i}) \left[b_k + \sum_{i=1}^{m_k} f_k(z_{k,i}) \right]^{-1} \qquad i = 1, \cdots, m_k \tag{4.14}$$

and

$$\beta_{k,0} = b_k \left[b_k + \sum_{i=1}^{m_k} f_k(z_{k,i}) \right]^{-1} \tag{4.15}$$

where f_k is defined in (4.12) and

$$b_k \triangleq m_k (\alpha_1 + \alpha_2 - \alpha_1\alpha_2) [(1-\alpha_1)(1-\alpha_2) V_k]^{-1}. \tag{4.16}$$[5]

[5]Using a Poisson model for the false observations leads to the same expression except that the (assumed) density of clutter per unit volume replaces $m_k V_k^{-1}$, which is the sample density at time k.

TABLE I
COMPUTATIONAL REQUIREMENTS OF THE PDAF VERSUS THE STANDARD FILTER

Number of Validated Returns	1	2	3	4
Factor of Increase	1	1.4	1.6	1.8

In view of assumption (4.3), $\hat{x}_{k|k,i}$ from (4.1) is obtained by a Kalman filter utilizing the measurement $z_{k,i}$. Incorporating this into (4.1) yields the probabilistic data association filter (PDAF)

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + W_k \nu_k \tag{4.17}$$

where

$$\nu_k \triangleq \sum_{i=1}^{m_k} \beta_{k,i} \nu_{k,i} \tag{4.18}$$

is the weighted innovation which uses all the validated measurements, and W_k is the weighting matrix given by

$$W_k = P_{k|k-1} H_k' S_k^{-1}. \tag{4.19}$$

The covariance associated with the estimate (4.17) is

$$P_{k|k} = \beta_{k,0} P_{k|k-1} + (1-\beta_{k,0}) P_{k|k}^* + W_k \left[\sum_{i=1}^{m_k} \beta_{k,i} \nu_{k,i} \nu_{k,i}' - \nu_k \nu_k' \right] W_k' \tag{4.20}$$

where $P_{k|k}^*$ is the covariance of the update if we have only one return. The last term above is a positive semidefinite matrix which shows the effect of the incorrect measurements by increasing the covariance of the update. The algebraic details related to (4.20) can be found in [5]. As can be seen from (4.20) the confidence on the estimate is a function of the actual number of validated returns and their location, i.e., is data dependent. Also note that the estimate $\hat{x}_{k|k}$ is a nonlinear function of the observations via the coefficients $\beta_{k,i}$.

An important feature of the PDAF is that its computational requirements are the same as the standard filter's when only one return falls in the window and increase only slightly when the need of processing multiple returns arises. Based upon a computer instruction count, Table I shows the approximate increase of the computational requirements of the PDAF compared to the standard filter[6] as a function of the number of returns in the validation region.

Simulation results were presented in [6] for a target with a seven-dimensional state similar to the one of [17] observed by two angle-only sensors (a two-dimensional measurement each) in a three-dimensional space. The PDAF gave substantially better probability of not losing the target, based upon 50 Monte Carlo runs over 100 sampling periods when compared to the nearest-neighbor standard Kalman filter.

V. OPTIMAL BAYESIAN APPROACH [25]

As in the previous section it is assumed that there is one target and an arbitrary number of observations with uncertain origin. The target and measurement models are the same. The main difference between this method and the one presented in the previous section is the following: the decomposition of the state estimate (according to the total probability theorem) is done in terms of all combinations of measurements from initial to present time[7] rather than only in terms of the latest measurements as in (4.1). A combination of measurements ("track history") at time k will be denoted as $\theta^{k,l}$ according to the definition (2.3). Consequently one can write

$$\theta^{k,l} = \left\{ \theta^{k-1,s}, z_{k,i_l} \right\}. \tag{5.1}$$

Let the event that such a history is true be $\mathcal{X}^{k,l}$. Its a posteriori probability is denoted as

$$\beta^{k,l} = P\left\{ \mathcal{X}^{k,l} | Z^k \right\} \tag{5.2}$$

Then the conditional mean of the state at time k will be expressed as

$$\hat{x}_{k|k} = E\left\{ x_k | Z^k \right\} = \sum_{l=1}^{L_k} \beta^{k,l} E\left\{ x_k | \mathcal{X}^{k,l}, Z^k \right\} = \sum_{l=1}^{L_k} \beta^{k,l} \hat{x}_{k|k}^l \tag{5.3}$$

which is the generalized version of (4.1). In the above equation L_k is the total number of track histories

$$L_k = \prod_{j=1}^{k} (1+m_j) \tag{5.4}$$

where m_j is the number of measurements observed at time j. The possibility that none of the returns is correct is also accounted for.

The vector whose elements are the number of measurements at each time through k is denoted as

$$m^k = [m_1, \cdots, m_k]'. \tag{5.5}$$

[6]The standard filter utilizes only one return (the "nearest neighbor") as if it were the correct one, i.e., the covariance and gain are not modified.

[7]This is equivalent to splitting the track from the initial time to the current and recombining this into a single estimate. A procedure that goes back N scans, i.e., does splitting in a moving window was also presented; this is discussed at the end of the section.

While (5.4) indicates that the memory and computational requirements are increasing with time, this approach is of interest because it requires no approximations. Furthermore, suboptimal versions that are intermediate between this and the one from the previous section can be obtained, as discussed later. Prior to obtaining the set of measurements Z_k (3.1) one has the state prediction $\hat{x}_{k|k-1}$ and associated covariance $P_{k|k-1}$. The predicted states $\hat{x}^s_{k|k-1}$, covariances $P^s_{k|k-1}$, predicted measurements $\hat{z}^s_{k|k-1}$ and corresponding innovation covariances $S^s_{k|k-1}$, conditioned upon track history $\theta^{k-1,s}$ being correct are also assumed available together with the probabilities $\beta^{k-1,s}$, $s=1,\cdots,L_{k-1}$. The set Z_k is made up of the measurements falling in a window (validation region) around the predicted measurement

$$\hat{z}_{k|k-1}=H_k\hat{x}_{k|k-1}. \tag{5.6}$$

The computation of the probabilities $\beta^{k,l}$ is done via Bayes' rule as follows, with the conditioning on $\boldsymbol{m}^k$ defined in (5.5) written out explicitly

$$\begin{aligned}\beta^{k,l}&=P\left\{\mathcal{X}^{k,l}|Z^k,\boldsymbol{m}^k\right\}\\&=P\left\{\mathcal{X}^{k-1,s},\mathcal{X}_{k,i_l}|Z_k,m_k,Z^{k-1},\boldsymbol{m}^{k-1}\right\}\\&=c_k^{-1}p\left(Z_k|\mathcal{X}_{k,i_l},m_k,\mathcal{X}^{k-1,s},Z^{k-1},\boldsymbol{m}^{k-1}\right)\\&\quad\cdot P\left\{\mathcal{X}_{k,i_l},m_k|\mathcal{X}^{k-1,s},Z^{k-1},\boldsymbol{m}^{k-1}\right\}\\&\quad\cdot P\left\{\mathcal{X}^{k-1,s}|Z^{k-1},\boldsymbol{m}^{k-1}\right\}\end{aligned} \tag{5.7}$$

where

$$c_k=P\left\{Z_k,m_k|Z^{k-1},\boldsymbol{m}^{k-1}\right\} \tag{5.8}$$

is a normalization constant.

The density on the r.h.s. of (5.7), denoted as $f_k(i_l)$ is for $i_l=0$ (none of the returns at k correct)

$$f_k(0)=V_k^{-m_k} \tag{5.9}$$

where V_k is the volume of the validation region at time k. Otherwise, for $i_l=1,\cdots,m_k$

$$f_k(i_l)=(1-\alpha_1)^{-1}\mathcal{N}\left(z_{k,i_l};\hat{z}^s_{k|k-1},S^s_{k|k-1}\right)V_k^{-m_k+1} \tag{5.10}$$

where α_1 is the probability that the correct return will not lie in the validation region.

The second term on the r.h.s. of (5.7) is

$$P\left\{\mathcal{X}_{k,i},m_k|\mathcal{X}^{k-1,s},Z^{k-1},\boldsymbol{m}^{k-1}\right\}$$

$$=\begin{cases}\dfrac{1}{m_k}(1-\alpha_1)(1-\alpha_2)P\left\{I_k=m_k-1\right\} & i_l=1,\cdots,m_k\\(\alpha_1+\alpha_2-\alpha_1\alpha_2)P\left\{I_k=m_k\right\} & i_l=0\end{cases} \tag{5.11}$$

where α_2 is the probability that the correct return will not be detected. The probability distribution of the number of incorrect returns, I_k, can be taken as Poisson or diffuse as pointed out in the previous section. The third term on the r.h.s. of (5.7) is $\beta^{k-1,s}$, which is available. This completes the calculation of the *a posteriori* probability that a track history is true.

The updated state estimate conditioned upon track $\theta^{k,l}=\{\theta^{k-1,s},z_{k,i_l}\}$ being correct is

$$\hat{x}^l_{k|k}=\hat{x}^s_{k|k-1}+W^l_k\left(z_{k,i_l}-\hat{z}^s_{k|k-1}\right) \tag{5.12}$$

where the filter gain is

$$W^l_k=P^s_{k|k-1}H_k\left(S^s_{k|k-1}\right)^{-1} \qquad \text{if } i_l\geqslant 1 \tag{5.13}$$

and zero if $i_l=0$ (none of the measurements at time k assumed correct). The covariance of the above updated state is

$$P^l_{k|k}=\left(I-W^l_kH_k\right)P^s_{k|k-1}. \tag{5.14}$$

The removal of the conditioning on a certain track history being correct is done by inserting (5.12) into (5.3) which then yields the estimate of the target's state conditioned upon all the observations.

The covariance of the estimate (5.3) is

$$P_{k|k}=\sum_{l=1}^{L_k}\beta^{k,l}P^l_{k|k}+\sum_{l=1}^{L_k}\beta^{k,l}\hat{x}_{k|k}\left(\hat{x}^l_{k|k}\right)'-\hat{x}_{k|k}\hat{x}'_{k|k}. \tag{5.15}$$

The derivation of the above is similar to the one of (4.20). An alternative expression for (5.15) can be obtained in terms of

$$\mu^l_k=\hat{x}^l_{k|k}-\hat{x}_{k|k-1}. \tag{5.16}$$

The resulting expression, which resembles more (4.20), is

$$\begin{aligned}P_{k|k}=&\sum_{l=1}^{L_k}\beta^{k,l}P^l_{k|k}+\sum_{l=1}^{L_k}\beta^{k,l}\mu^l_k\left(\mu^l_k\right)'\\&-\left(\sum_{l=1}^{L_k}\beta^{k,l}\mu^l_k\right)\left(\sum_{l=1}^{L_k}\beta^{k,l}\mu^l_k\right)'.\end{aligned} \tag{5.17}$$

The predicted state and covariance are obtained via the standard Kalman filter equations. This completes the description of the optimal Bayesian filter.

The problem of (exponentially) increasing memory and computational requirements due to (5.4) can be alleviated by the following procedure: all tracks that have identical histories for the past N scans before the current one are to be combined. This procedure will prevent the memory and computational requirement from increasing. The expected number of tracks to be stored and updated in this "N-scan-back" filter is

$$I=\prod_{i=1}^{N}\left(1+E\left\{m_i\right\}\right) \tag{5.18}$$

which makes the algorithm implementable for small N. The algorithm of Section IV corresponds to $N=0$: no previous scans are reexamined and all the current measurements are combined immediately into the estimate.

Simulations reported in [25] on a two-dimensional system for six time periods based upon 1000 Monte Carlo runs indicated that the optimal filter will give a substantial performance improvement over the nearest-neighbor

TABLE II
COMPARISON OF THE MAIN ALGORITHMS

Algorithm		Suitable for Track Formation	Number of Targets of Interest	Assumed Probability of Detection	State Estimates	Covariance of State	Memory and Computational Requirements
Non-Bayesian (based on likelihood function)	Track Split [10], [27]	yes	arbitrary	unity	are conditioned upon accepted trajectories being correct	yielded by standard Kalman filter	increase in time in a dense environment
	Maximum Likelihood [18]				are conditioned upon selected most likely trajectories being correct		
Bayesian (using *a posteriori* probabilities)	PDAF [6] (Suboptimal, zero-scan-back)	track assumed initiated	one (has been extended to arbitrary [7])	arbitrary	account for all neighboring measurements	account for uncertainty in the origin of the measurements used in the estimates	slightly higher than standard Kalman filter
	N-scan-back (Suboptimal [25])		one				higher than standard Kalman filter
	Optimal (full scan-back) [25]						increase in time

standard Kalman filter. The one and two-scan filters were very close, indicating that they were nearly optimal. The "zero-scan-back" version, which is the PDAF from the previous section, was reported as having a lower performance.

VI. Summary

All the four algorithms discussed above can handle, at least in principle, an arbitrary number of false measurements. While the computational and memory requirements of the track split and maximum likelihood methods can easily become excessive in a dense environment, the PDAF will have the least computational requirements; the *N*-scan-back non-Bayesian is likely to yield better performance at a higher cost than the PDAF. Direct comparison between the Bayesian and non-Bayesian algorithms cannot be made because of the underlying philosophical differences in the formulation.

The two Bayesian algorithms described in Sections IV and V assume only a single target of interest. A more general version of the PDAF has been developed in [7] for the case where there are several targets of interest in the same neighborhood. This method consists of computing the probability that each latest measurement belongs to each feasible target taking into account all its possible origins (other targets or clutter). Using this information all the measurements are combined with the proper weighting into each target's estimate. Due to its complexity this method is not discussed here. The algorithm of Alspach [2], which is based on the Gaussian sum approach, can handle m targets but requires the number of measurements to be the same, i.e., unity probability of detection and no false alarms. Since this is a quite specialized case it is not presented here.

Reid [21] recently developed a Bayesian approach that can handle several targets with false alarms as well as new targets (track initiation). A density of true targets as well as one of false alarms (in terms of reports per unit area) are assumed to set up a prior probability distribution for the number of false alarms and new targets. This then allows obtaining posterior probabilities for each measurement association (track). This is a Bayesian version of the maximum likelihood procedure of [18]. The implementation of this method is sequential with provisions of keeping the number of tracks from diverging.

Another recent work is the one of Hurd [12] which deals with a Bayesian updating procedure for a set of targets with discrete-valued states. The measurements in this problem could have originated from more than one of the targets in the set under consideration.

An application work was reported by Fortmann and Baron [11] where a somewhat modified version of the PDAF was implemented and run on real data. Table II presents a comparison of the algorithms discussed earlier, highlighting their main properties.

VII. Conclusion and Topics for Future Research

While no tracking algorithm can work in an arbitrarily dense environment, the recently developed techniques described in this paper extend the range, in terms of target density, in which one can achieve reliable tracking performance. The viability of these algorithms will have to be tested in real applications. Compromises between computational requirements and performance will be necessary in many cases, especially where the data rate is high and the available computing power is limited, as in airborne systems. The assumptions made in the various algorithms discussed here will probably be modified according to the needs of specific applications. New algorithms that are more versatile will also be developed. One such algorithm seems to be the method presented in [21].

A number of problems that are yet open can be seen at this time. One such problem is the treatment of maneuver-

ing targets in a multitarget environment [11]. At this point it is not known whether adaptive filtering techniques can be combined with any of the methods cited above. The utilization of signal processing information in the estimation procedure could reduce the effect of measurement origin uncertainty. The problems of unresolved detections and crossing targets should also be considered. The suitability of modeling the number of false measurements by a diffuse or Poisson distribution can probably be answered only in the context of a real application. A comparison of the performance vs. cost for the several versions of the Bayesian procedures ($N=0,1,\cdots$ back-scans) for some real problems would also be of interest. The problem of computational requirements, which might be severe in some cases, could be alleviated by the use of fast algorithms that have been found useful in Kalman filters.

The recent activity in the area of multitarget tracking indicates that this emerging body of knowledge experiences a rapid transition to becoming relevant to engineering practice by solving real problems.

Acknowledgment

A series of discussions with Dr. Thomas E. Fortmann from BBN that elucidated several obscure points and also helped relate the methods discussed in this paper is gratefully acknowledged.

References

[1] G. A. Ackerson, and K. S. Fu, "On state estimation in switching environments," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 10–17, Feb. 1970.

[2] D. L. Alspach, "A Gaussian sum approach to the multitarget identification-tracking problem," *Automatica*, vol. 11, pp. 285–296, May 1975 (earlier version in *Proc. 4th Symp. Nonlinear Estimation*, 1973).

[3] M. Athans, R. H. Whiting, and M. Gruber, "A suboptimal estimation algorithm with probabilistic editing for false measurements with applications to target tracking with wake phenomena," in *Proc. 6th Symp. Nonlinear Estimation*, Univ. Calif., San Diego, Sept. 1975 and *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 372–384, June 1977.

[4] H. Akashi and H. Kumamoto, "Random sampling approach to state estimation in switching environments," *Automatica*, vol. 13, pp. 429–434, July 1977.

[5] Y. Bar-Shalom and A. Jaffer, "Adaptive nonlinear filtering for tracking with measurements of uncertain origin," in *Proc. 1972 IEEE Conf. Decision and Control*, pp. 243–247, New Orleans, LA, Dec. 1972.

[6] Y. Bar-Shalom, and E. Tse, "Tracking in a cluttered environment with probabilistic data association," in *Proc. 4th Symp. Nonlinear Estimation*, Univ. Calif., San Diego, Sept. 1973, and *Automatica*, vol. 11, pp. 451–460, Sept. 1975.

[7] Y. Bar-Shalom, "Extension of the probabilistic data association filter to multitarget environment," in *Proc. 5th Symp. Nonlinear Estimation*, Univ. Calif., San Diego, Sept. 1974.

[8] M. DeGroot, *Optimal Statistical Decisions*. New York: McGraw-Hill, 1970.

[9] T. S. Ferguson, *Mathematical Statistics*. New York: Academic, 1967.

[10] E. C. Fraser and L. Meier, "Mathematical models and optimum computation for computer-aided active sonar systems," U.S. Navy Electronic Lab., SRI Final Rep. (First Year), San Diego, CA, Contract N123-(953)54486A, Mar. 1967.

[11] T. E. Fortmann and S. Baron, "Octopus, an experimental passive sonar tracking system" (SECRET), in *Proc. ONR Passive Tracking Conf.*, Monterey, CA, May 1977.

[12] H. L. Hurd, "A Bayesian approach to simultaneous tracking of multiple targets," in *Proc. ONR Passive Tracking Conf.*, Monterey, CA, May 1977.

[13] A. G. Jaffer and S. C. Gupta, "Recursive Bayesian estimation with uncertain observation," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 614–616, Sept. 1971.

[14] ——, "Optimal sequential estimation of discrete processes with Markov interrupted observations," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 471–475, Oct. 1971.

[15] A. G. Jaffer and Y. Bar-Shalom, "On optimal tracking in multiple-target environments," in *Proc. 3rd Symp. Nonlinear Estimation Theory and Its Applications*, pp. 112–117, Univ. Calif., San Diego, Sept. 1972.

[16] T. Kailath, "An innovations approach to least-squares estimation, part I: Linear filtering in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 646–654, Dec. 1968.

[17] R. K. Mehra, "A comparison of several nonlinear filters for reentry vehicle tracking," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 307–319, 1972.

[18] C. L. Morefield, "Application of 0-1 integer programming to multitarget tracking problems," in *Proc. IEEE Conf. Decision and Control*, Dec. 1975 and *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 302–312, June 1977.

[19] N. E. Nahi, "Optimal recursive estimation with uncertain observations," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 456–462, July 1969.

[20] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1965.

[21] D. B. Reid, "A multiple hypothesis filter for tracking multiple targets in a cluttered environment," Lockheed Palo Alto Res. Lab. Tech. Rep., LMSC-D560254, Sept. 1977.

[22] R. G. Sea, "An efficient suboptimal decision procedure for associating sensor data with stored tracks in real-time surveillance systems," in *Proc. 1971 IEEE Conf. on Decision and Control*, pp. 33–37, Miami Beach, FL, Dec. 1971.

[23] R. A. Singer and J. J. Stein, "An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance systems," in *Proc. 1971 IEEE Conf. Decision and Control*, pp. 171–175, Miami Beach, FL, Dec. 1971.

[24] R. A. Singer and R. G. Sea, "New results in optimizing surveillance system tracking and data correlation performance in dense multitarget environments," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 571–581, Dec. 1973.

[25] R. A. Singer, R. G. Sea, and K. Housewright, "Derivation and evaluation of improved tracking filters for use in dense multitarget environments," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 423–432, July 1974.

[26] R. W. Sittler, "An optimal data association problem in surveillance theory," *IEEE Trans. Mil. Electron.*, vol. MIL-8, pp. 125–139, Apr. 1964.

[27] P. Smith and G. Buechler, "A branching algorithm for discriminating and tracking multiple objects," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 101–104, Feb. 1975.

Yaakov Bar-Shalom (S'63–M'66) was born on May 11, 1941. He received the B.S. (cum laude) and M.S. degrees from the Technion, Israel Institute of Technology, Haifa, in 1963 and 1967, respectively, and the Ph.D. degree from Princeton University, Princeton, NJ, in 1970, all in electrical engineering.

During 1963–1966 he served in the Israel Defense Army as an Electronics Officer. From 1966 to 1967 he was a Teaching Assistant in the Department of Electrical Engineering at the Technion. Between 1967 and 1970 he was at Princeton University, first as a Research Assistant, then as a Research Associate. During the academic year 1968–1969 he was appointed Charles Grosvenor Osgood Fellow in recognition for outstanding performance. From 1970 to 1976 he was a Senior Research Engineer with Systems Control, Inc., Palo Alto, CA. At SCI he has developed and applied advanced stochastic control and estimation procedures to problems in the areas of defense, environmental and economic systems. Between 1973 and 1976 he taught graduate courses in system theory at the University of Santa Clara, Santa Clara, CA. Since 1976 he has been an Associate Professor of Electrical Engineering and Computer Science at the University of Connecticut, Storrs. He has been consulting to Systems Control, Inc.; ORI, Inc.; System Development Corporation; C.S. Draper Laboratory; and BBN. His research interests are in stochastic adaptive control, resource allocation under uncertainty, estimation theory, decision theory, system identification, and modeling, analysis, and control of environmental and economic systems.

Dr. Bar-Shalom is a member of the CESR (Conference on the Computer in Economic and Social Research), Eta Kappa Nu, and Sigma Xi. During 1976 and 1977 he served as Chairman of the Technical Committee on Stochastic Control and Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and currently is Associate Editor of *Automatica*.