Comparison of fusion methods for multiple target tracking

Mark Morelande
Dept. Electrical & Electronic Engineering
The University of Melbourne
Parkville, VIC, Australia
mrmore@unimelb.edu.au

Neil Gordon
ISR Division
Defence Science & Technology Organisation
Edinburgh, SA, Australia
neil.gordon@dsto.defence.gov.au

Abstract – *Fusion of data from multiple sensors for the purposes of tracking multiple targets is considered. Two methods of fusion are considered. The first uses all measurements to perform tracking while the second adds a clustering step which attempts to remove clutter prior to tracking. Clustering is performed using a sequential Bayesian analysis and tracking is performed using a Monte Carlo approximation to the multiple hypothesis tracker (MHT). In performance analyses involving up to 25 targets, tracking with all measurements performs significantly better than tracking with the output of the clustering algorithm.*

1 Introduction

The goal of multi-target tracking is to recursively estimate the number of targets in a surveillance area and their locations using a sequence of observations. In the most common formulation of the problem the tracking algorithm is provided, at each sampling instant, with a set of locations at which targets are supposed to be. Since the detection procedure is imperfect the locations may correspond to non-target related phenomena, or clutter, rather than targets. Further, there is usually no information provided to distinguish measurements due to different targets. There is therefore a measurement origin uncertainty which can be optimally resolved only by enumerating and evaluating all possible hypotheses regarding the origins of the measurements. This optimal approach, referred to as multiple hypothesis tracking (MHT), is not practical because the number of hypotheses increases exponentially as measurements are acquired. A number of practical, approximate MHT algorithms have been proposed [1, 2].

When measurements from multiple sensors are available, the optimal approach is to pass all measurements from all sensors to the tracking algorithm. However, because of the approximations which are used in all practical tracking algorithms, this approach may not result in the best performance in practice. It has been suggested in [3, 4] that better performance may be achieved by exploiting the different characteristics of target and clutter measurements to remove clutter measurements prior to tracking. The reduction in the number of clutter measurements is achieved under the assumption that measurements from a particular target will be similarly positioned for all sensors while clutter measurements will arise at different locations across the collection of sensors. The use of a pre-processing step to remove clutter in multiple sensor data is referred to as fuse-before-track [3].

The aim of this paper is to analyse the value of the fuse-before-track approach to multiple sensor multiple target tracking. A Monte Carlo approximation to the optimal MHT algorithm, similar to that of [5], is used for tracking. Clutter reduction is approached as a clustering problem which in turn is formulated as a mixture estimation problem. A Bayesian framework is adopted as it allows the incorporation of prior information obtained from previous measurement scans. In this mixture estimation problem the number of mixture components is unknown and there is a potentially large number of outliers, i.e., measurements which do not belong to any cluster. The former difficulty has been widely considered in the literature [6] while the latter difficulty appears to have received little attention. Bayesian mixture estimation is commonly carried out using Monte Carlo approaches which are plagued by the label switching problem wherein different samples from the posterior may correspond to different permutations of the component labels [7]. Here the label switching problem is avoided by processing the measurements sequentially using a model adapted from [8].

The paper is organised as follows. The MHT algorithm is reviewed in Section 2 and a Monte Carlo approximation is developed. The Bayesian clustering algorithm is given in Section 3. Section 4 contains the performance comparison between tracking with all sensor measurements and tracking only with the measurements returned by the Bayesian clustering algorithm.

2 Multiple hypothesis tracking

In this section multiple hypothesis tracking (MHT) is reviewed and its approximation using Monte Carlo methods is described. The development assumes the usual linear/Gaussian stochastic dynamic model for the evolution and observation of the target states. The measurements obtained from a collection of d sensors at K sample times are arranged into Kd scans such that the measurements of the kth scan are collected at the $(\lfloor (k-1)/d \rfloor + 1)$th sample time. Let r denote the unknown number of targets and $\boldsymbol{x}_{i,k} = [x_{i,k}, y_{i,k}, \dot{x}_{i,k}, \dot{y}_{i,k}]'$ denote the state of the ith target at the kth scan, where $(x_{i,k}, y_{i,k})$ is the target position in Cartesian coordinates and the dot notation indicates differentiation with respect to time. The target states evolve independently according to

$$\boldsymbol{x}_{i,k} | \boldsymbol{x}_{i,k-1} \sim N(\boldsymbol{F}_k \boldsymbol{x}_{k-1}, \boldsymbol{Q}_k) \tag{1}$$

where

$$\boldsymbol{F}_k = \begin{cases} \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \otimes \boldsymbol{I}_2, & \text{rem}((k-1)/d) = 0, \\ \boldsymbol{I}_4, & \text{otherwise}, \end{cases} \tag{2}$$

$$\boldsymbol{Q}_k = \begin{cases} q \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix} \otimes \boldsymbol{I}_2, & \text{rem}((k-1)/d) = 0, \\ \boldsymbol{0}, & \text{otherwise}. \end{cases} \tag{3}$$

where T is the sampling period. Note that target states change values only every d scans when the sampling instant is advanced by T.

The sensors produce noisy measurements of the target positions in Cartesian coordinates as well as measurements due to non-target related phenomena, or clutter. Targets are detected with probability P_D and the clutter measurements are uniformly distributed with the number in a given area Poisson distributed with density λ. The measurements in the kth scan are denoted by $\boldsymbol{y}_{k,1}, \ldots, \boldsymbol{y}_{k,m_k}$ where m_k is the number of measurements received. Let $\nu_{k,j} = i > 0$ denote the event that the jth measurement is due to the ith target and $\nu_{k,j} = 0$ denote the event that this measurement is due to clutter. Then, measurements are generated independently as

$$\boldsymbol{y}_{k,j} | (\boldsymbol{x}_{i,k}, \nu_{k,j} = i) \sim \begin{cases} 1/V, & i = 0, \\ N(\boldsymbol{y}_{k,j}; \boldsymbol{H}\boldsymbol{x}_{i,k}, \boldsymbol{R}), & \\ & i = 1, \ldots, r, \end{cases} \tag{4}$$

where V is the area of the surveillance region, $\boldsymbol{R}$ is a positive definite covariance matrix and $\boldsymbol{H} = [1\ 0] \otimes \boldsymbol{I}_2$. Let $\boldsymbol{y}_k = [\boldsymbol{y}'_{k,1}, \ldots, \boldsymbol{y}'_{k,m_k}]'$ denote the vector of measurements in the kth scan.

2.1 Review of MHT

A measurement origin hypothesis accounts for each of the measurements in a scan. A valid measurement origin hypothesis assigns at most one target to each measurement and at most one measurement to each target. We represent a measurement origin hypothesis by a vector $\boldsymbol{\nu}_k = [\nu_{k,1}, \ldots, \nu_{k,m_k}]$ where $\nu_{k,j} \in \mathbb{Z}$ represents the origin of the jth measurement. A sequence of such hypotheses is written as $\boldsymbol{\nu}_{1:k}$. The MHT involves recursively computing the posterior probabilities of each valid hypothesis sequence and the posterior PDF of the target states presumed to exist under each hypothesis sequence.

Assume that the MHT has been applied to the measurement sequence $\boldsymbol{y}_{1:k-1}$ to give the posterior probabilities $\mathsf{P}(\boldsymbol{\nu}_{1:k-1} | \boldsymbol{y}_{1:k-1})$ and conditional posterior PDFs $\pi_{k-1|k-1,i}(\cdot | \boldsymbol{\nu}_{1:k-1})$ for $i = 1, \ldots, r(\boldsymbol{\nu}_{1:k-1})$, where $r(\boldsymbol{\nu}_{1:k-1})$ is the number of targets presumed to exist under $\boldsymbol{\nu}_{1:k-1}$. Given measurements $\boldsymbol{y}_k$ it is desired to calculate the posterior probabilities and conditional posterior PDFs for each sequence $\boldsymbol{\nu}_{1:k}$. Following [2] it is assumed that new targets appear uniformly in the surveillance region and the number of new targets is Poisson distributed with density β. For the hypothesis $\boldsymbol{\nu}_k$, let c denote the number of clutter measurements, b the number of new targets and e the number of existing targets detected. The set of indices of measurements due to existing targets is denoted as $D = \{j : \nu_{k,j} \in \{1, \ldots, r\}\}$. The dependence of these quantities on the current hypothesis $\boldsymbol{\nu}_k$ has been omitted for brevity. The dependence of the number of existing targets on $\boldsymbol{\nu}_{1:k-1}$ is also omitted in the following expressions. The posterior probability of $\boldsymbol{\nu}_{1:k}$ is

$$\mathsf{P}(\boldsymbol{\nu}_{1:k} | \boldsymbol{y}_{1:k}) \propto \tau_k(\boldsymbol{\nu}_{1:k}) \mathsf{P}(\boldsymbol{\nu}_{1:k-1} | \boldsymbol{y}_{1:k-1}) \tag{5}$$

where the multiplicative update is

$$\tau_k(\boldsymbol{\nu}_{1:k}) = (1 - P_D)^r \left[\frac{P_D}{\lambda(1 - P_D)} \right]^e (\beta/\lambda)^b \times \prod_{j \in D} \psi_{k,\nu_{k,j}}(\boldsymbol{y}_{k,j}) \tag{6}$$

The prior measurement PDF is given by

$$\psi_{k,i}(\boldsymbol{y}) = \int N(\boldsymbol{y}; \boldsymbol{H}\boldsymbol{x}, \boldsymbol{R}) \pi_{k|k-1,i}(\boldsymbol{x} | \boldsymbol{\nu}_{1:k-1}) \, d\boldsymbol{x} \tag{7}$$

where $\pi_{k|k-1,i}(\cdot | \boldsymbol{\nu}_{1:k-1})$ is the prior target state PDF. The prior measurement and target state PDFs can be calculated in closed-form for the linear/Gaussian stochastic dynamic system used here. It remains to calculate the conditional posterior PDFs. If the ith existing target is assigned to the jth measurement then

$$\pi_{k|k,i}(\boldsymbol{x} | \boldsymbol{\nu}_{1:k}) \propto N(\boldsymbol{y}_{k,j}; \boldsymbol{H}\boldsymbol{x}, \boldsymbol{R}) \pi_{k|k-1,i}(\boldsymbol{x} | \boldsymbol{\nu}_{1:k-1}) \tag{8}$$

If the ith existing target is not assigned to a measurement under $\boldsymbol{\nu}_k$ then $\pi_{k|k,i}(\boldsymbol{x} | \boldsymbol{\nu}_{1:k}) = \pi_{k|k-1,i}(\boldsymbol{x} | \boldsymbol{\nu}_{1:k-1})$.

In an exact MHT implementation the expressions (6) and (8) are evaluated for all valid hypotheses

$\boldsymbol{\nu}_k$ branching from each existing hypothesis $\boldsymbol{\nu}_{1:k-1}$. The process is then repeated for the measurements $\boldsymbol{y}_{k+1}, \boldsymbol{y}_{k+2}, \ldots$. This procedure cannot be performed in practice because the number of valid hypotheses increases exponentially with each new collection of sensor measurements. Many techniques, such as hypothesis pruning, track clustering and hypothesis merging, have been proposed to address this problem [1, 2].

2.2 Monte Carlo MHT

It is proposed to approximate the posterior hypothesis probabilities by a Monte Carlo procedure in which measurement origin hypotheses are drawn from an importance distribution and assigned appropriate weights. This approximation is referred to as Monte Carlo MHT (MC-MHT). Let $\boldsymbol{\nu}^a_{1:k} \sim Q(\cdot|\boldsymbol{y}_{1:k})$, $a = 1, \ldots, n$ denote the samples of k-length association hypotheses drawn from the importance distribution Q. The sample weights are computed as

$$w^a_k = C_k \, \mathsf{P}(\boldsymbol{\nu}^a_{1:k}|\boldsymbol{y}_{1:k}) / Q(\boldsymbol{\nu}^a_{1:k}|\boldsymbol{y}_{1:k}) \tag{9}$$

where C_k is such that the weights sum to one. The posterior probabilities are approximated by

$$\widehat{\mathsf{P}}(\boldsymbol{\nu}_{1:k}|\boldsymbol{y}_{1:k}) = \sum_{\{a:\boldsymbol{\nu}^a_{1:k}=\boldsymbol{\nu}_{1:k}\}} w^a_k. \tag{10}$$

A sequential procedure is obtained by choosing the importance distribution to factorise as [9]

$$Q(\boldsymbol{\nu}_{1:k}|\boldsymbol{y}_{1:k}) = Q(\boldsymbol{\nu}_k|\boldsymbol{\nu}_{1:k-1}, \boldsymbol{y}_{1:k}) Q(\boldsymbol{\nu}_{1:k-1}|\boldsymbol{y}_{1:k-1}) \tag{11}$$

Then, $\boldsymbol{\nu}^a_{1:k} = [\boldsymbol{\nu}^a_{1:k-1}, \boldsymbol{\nu}^a_k]$ where

$$\boldsymbol{\nu}^a_k \sim Q(\cdot|\boldsymbol{\nu}^a_{1:k-1}, \boldsymbol{y}_{1:k}) \tag{12}$$

and the weights are computed sequentially as

$$w^a_k = B_k \, w^a_{k-1} \tau_k(\boldsymbol{\nu}^a_{1:k}) \, / Q(\boldsymbol{\nu}^a_k|\boldsymbol{\nu}^a_{1:k-1}, \boldsymbol{y}_{1:k}) \tag{13}$$

where B_k is a normalising constant. Repeated application of (12) and (13) will eventually lead to sample degeneracy. This can be avoided by performing resampling [10]. Resampling can be formulated as the drawing of an auxiliary variable, the sample index, in addition to the measurement origin hypothesis. Thus, for $a = 1, \ldots, n$, $s^a \in \{1, \ldots, n\}$, we draw

$$(\boldsymbol{\nu}^a_k, s^a) \sim \vartheta^{s^a}_k Q(\boldsymbol{\nu}^a_k|\boldsymbol{\nu}^{s^a}_{1:k-1}, \boldsymbol{y}_{1:k-1}) \tag{14}$$

where $\sum_{s=1}^n \vartheta^s_k = 1$, and calculate the weight

$$w^a_k = B_k \, w^{s^a}_{k-1} \tau_k(\boldsymbol{\nu}^a_{1:k}) \Big/ \vartheta^{s^a}_k Q(\boldsymbol{\nu}^a_k|\boldsymbol{\nu}^{s^a}_{1:k-1}, \boldsymbol{y}_{1:k}) \tag{15}$$

where $\boldsymbol{\nu}^a_{1:k} = [\boldsymbol{\nu}^{s^a}_{1:k-1}, \boldsymbol{\nu}^a_k]$. The sampled auxiliary variables $s^1, \ldots, s^n$ are discarded.

Under some regularity conditions the importance sampling approximations of the posterior hypothesis probabilities become exact as the sample size $n \to \infty$. However, for a finite sample size the importance distribution has a great influence on the accuracy of the approximation. Ideally, measurements would be assigned based on joint association probabilities. However the practical impossibility of calculating these probabilities is the reason sampling is necessary at all. It is proposed to instead associate the individual measurements sequentially. Let r^s denote the number of targets presumed to existed under $\boldsymbol{\nu}^s_{1:k-1}$. Then, sampling conditional on $\boldsymbol{\nu}^s_{1:k-1}$, we have $\nu_{k,j} \in \{0, \ldots, r^s + 1\}$ where $\nu_{k,j} = 0$ assigns the jth measurement to clutter, $\nu_{k,j} = i \in \{1, \ldots, r^s\}$ assigns the jth measurement to the ith existing target and $\nu_{k,j} = r^s + 1$ means that the jth measurement is due to a new target. The proposed importance density is

$$\vartheta^s_k = w^s_{k-1} \tag{16}$$

$$Q(\boldsymbol{\nu}_k|\boldsymbol{\nu}^s_{1:k-1}, \boldsymbol{y}_k) = \prod_{j=1}^{m_k} \rho^s_{\nu_{k,j},j} \tag{17}$$

where $\rho^s_{i,j} = \alpha^s_{i,j} / u^s_j$ with

$$\alpha^s_{i,j} = \begin{cases} 1, & i = 0 \\ P_D \psi^s_{k,i}(\boldsymbol{y}_{k,j}) / [\lambda(1 - P_D)], & i = 1, \ldots, r^s, \\ \beta/\lambda, & i = r^s + 1, \end{cases} \tag{18}$$

$$u^s_j = \sum_{i \in A^s_j} \alpha^s_{i,j}. \tag{19}$$

In (18), $\psi^s_{k,i}(\cdot)$ is the prior measurement PDF of (7) conditional on $\boldsymbol{\nu}^s_{1:k-1}$. The set $A^s_j \subseteq \{0, \ldots, r^s + 1\}$ contains the possible values of the jth association hypothesis. It is given by $A^s_1 = \{0, \ldots, r^s + 1\}$ and $A^s_j = A^s_{j-1} \setminus (\{1, \ldots, r^s\} \cap \{\nu_{k,j-1}\})$ for $j = 2, \ldots, m_k$. The weights for the importance density of (16)-(17) are

$$w^a_k = B_k \, (1 - P_D)^{r^{s^a}} \prod_{j=1}^{m_k} u^{s^a}_j \tag{20}$$

A draw from (17) is made by sampling an association hypothesis for each measurement in turn. As seen in (18), the sampling probabilities for an association hypothesis are determined by strengths of association between the measurement and each possible origin hypothesis, without taking into account neighbouring measurements. The bias which results from sampling association hypotheses singly rather than jointly is accounted for in the weight calculation of (20).

An important characteristic of the sampling procedure of (17) is demonstrated in an example with two existing targets and two measurements. The association strengths, defined in (18), are listed in Table 1.

It is desirable for the importance sampling probabilities to match the posterior hypothesis probabilities as closely as possible since this reduces the variance

Table 1: Association strengths for the importance sampling example.

	Measurement 1	Measurement 2
Clutter	1	1
Target 1	100	100
Target 2	5	0.1
New target	1	1

of the importance weights and therefore the variance of the importance sampling approximation [11]. The posterior probabilities and importance sampling probabilities for the 14 valid joint hypotheses arising from the scenario of Table 1 are plotted in Figure 1. Two importance distributions are considered. The first importance distribution Q_1 samples an association for measurement 1 and then for measurement 2. The order in which associations are sampled is reversed for the second importance distribution Q_2. It can be seen that Q_2 is much closer to the posterior probability distribution than Q_1. Consequently, the importance sampling approximation to the posterior probabilities obtained using Q_2 is much more accurate than that obtained using Q_1: the approximation variance using Q_2 is about $1/10$ of that obtained using Q_1.

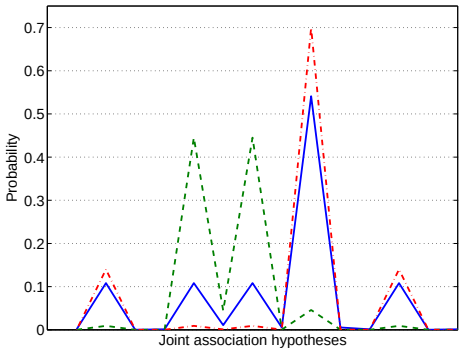

Figure 1: Posterior (——) and importance sampling probabilities (– – – for Q_1 and –·– for Q_2) for the joint association hypotheses of the example of Table 1.

The example of Table 1 demonstrates the dependence of performance on the order in which associations are sampled. In this simple example it would be possible to determine which of the two orderings is preferred. However, to do so would require a computational expense similar to that of the association problem we are trying to solve. Two methods have been explored for addressing the order-dependent performance of the sampling scheme. First, a random ordering of measurements, independently generated for each sample, can be used. In the example of Table 1 this results in an approximation variance which is about $1/11$ of that obtained using Q_1. Interestingly, for this example randomly selected orders do better than either of the fixed orders. In the second method associations for clusters of neighbouring measurements are sampled jointly. The approximation variance for joint association sampling is about $1/20$ of that for Q_1. The first method, random measurement permutations, is often effective and can be implemented with little additional computational cost. The second method is also effective but entails a significant computational expense and can really only be applied to small clusters of measurements. Although these methods can be used together, i.e., joint sampling with random ordering, in this paper single associations are sampled in a random order. A detailed analysis of the performance improvements offered by joint sampling and random ordering is given in [12].

2.3 Discussion

The collection of hypotheses evaluated by the MC-MHT must be summarised to provide a useful output. Here, the output of the MC-MHT are the tracks in the most probable hypothesis which have more than a certain number of detections in their association history. The threshold on the number detections required for output of a track can be used to control the number of false tracks. Despite its simplicity this method seems to perform better than more sophisticated alternatives.

In approximate MHT implementations it is not uncommon for an erroneous track to become attached to an otherwise good hypothesis. A means of identifying such tracks, so that they may be removed, is required. This can be done checking that a track's recent detection history is compatible with that expected of a target track. In particular, let d denote the number of detections received by a particular track over a trajectory of length h. Then, the track is declared false, and removed from all hypotheses in which it appears, if

$$B(d; h, P_D) \leq \Lambda \tag{21}$$

where $B(\cdot; n, p)$ is the cumulative distribution function of a binomial random variable with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$ and Λ is a pre-defined threshold. Usually, the threshold Λ is quite small, say 10^{-3}.

3 Clutter reduction using measurement clustering

The availability of measurements from multiple sensors presents the opportunity of reducing clutter by adding a clustering step prior to tracking. The basic idea is that target-originated measurements form reasonably tight clusters across the sensor collection while clutter measurements are distributed uniformly. The clustering problem is formulated as a mixture estimation problem and a Bayesian solution is developed.

Assume that m measurements $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$ are obtained from a collection of sensors at a given sampling

instant. Each measurement can be assigned to a cluster or deemed to be an outlier. Assume that there are at most $\bar{q}$ clusters and let $\zeta_t \in \{0, \ldots, \bar{q}\}$ denote the label assigned to the tth measurement. The mixture is characterised by the weights $\boldsymbol{w} = [w_0, \ldots, w_{\bar{q}}]'$, $\sum_{i=0}^{\bar{q}} w_i = 1$, and locations $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{\bar{q}}$. The mixture parameters are collected into the vector $\boldsymbol{\theta}$ with prior PDF π_0. The label assignments and measurements satisfy

$$\mathsf{P}(\zeta_t = i) = w_i, \tag{22}$$

$$\ell(\boldsymbol{y}_t|\boldsymbol{\theta}, \zeta_t = i) = \begin{cases} 1/V, & i = 0, \\ N(\boldsymbol{y}_t; \boldsymbol{\mu}_i, \boldsymbol{R}), & i = 1, \ldots, \bar{q}, \end{cases} \tag{23}$$

where V is the area of the region of interest and $\boldsymbol{R}$ is a known covariance matrix. Note that w_0 is the weight assigned to the outliers. The clustering problem amounts to estimating the labels $\zeta_1, \ldots, \zeta_m$.

3.1 Sequential Bayesian clustering

The mixture estimation problem is complicated by identifiability issues. In particular, the value of the likelihood is unchanged when the labels assigned to each component are permuted. This lack of identifiability causes difficulties when performing inference. An example of this is the occurrence of label switching in sampled-based Bayesian approaches [7, 13]. A sequential formulation of the mixture estimation problem can enforce identifiability by constraining the mixture components to appear gradually in the measurement record. Let q_{t-1} denote the number of unique mixture components assigned for a sequence $\zeta_{1:t-1}$ of label assignments. For $i \geq 0$, $q_{t-1} < \bar{q}$, the label probability (22) is replaced by [8]

$$\mathsf{P}(\zeta_t = i|\zeta_{1:t-1}) = \begin{cases} w_i, & i = 0, \ldots, q_{t-1}, \\ \displaystyle\sum_{j=q_{t-1}+1}^{\bar{q}} w_j, & i = q_{t-1}+1, \\ 0, & i = q_{t-1}+2, \ldots, \bar{q}. \end{cases} \tag{24}$$

If $q_{t-1} = \bar{q}$ then $\mathsf{P}(\zeta_t = i|\zeta_{1:t-1}) = w_i$ for $i = 1, \ldots, \bar{q}$. If the prior PDF for the mixture parameters is permutation invariant (24) produces a sequential model which is equivalent to the batch model (22), in a sense described in [14]. More generally, if the prior is not permutation invariant, q_{t-1} in (24) should be the sum of the number of permutation variant components in the prior and the number of permutation invariant components which have been assigned under $\zeta_{1:t-1}$.

It is assumed that the weights are known and only the component means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_q$ need to be estimated. In the context of multiple sensor tracking this amounts to assuming that the clutter density and detection probability are known. The extension to the case of unknown weights can be handled by placing a Dirichlet prior on the weights. The prior PDF is, for $q_0 \leq \bar{q}$,

$$\pi_0(\boldsymbol{\theta}) = \prod_{i=1}^{q_0} N(\boldsymbol{\mu}_i; \hat{\boldsymbol{\mu}}_{i,0}, \boldsymbol{\Sigma}_{i,0}) \prod_{i=q_0+1}^{\bar{q}} N(\boldsymbol{\mu}_i; \hat{\boldsymbol{\mu}}_0, \boldsymbol{\Sigma}_0)$$

Our aim is to compute the posterior PDF of the parameter $\boldsymbol{\theta}$ and label sequence $\zeta_{1:t}$ sequentially for $t = 1, 2, \ldots$. Assume after processing the first $t-1$ measurements that, for each valid $\zeta_{1:t-1}$, we have

$$\pi_{t-1}(\boldsymbol{\theta}, \zeta_{1:t-1}) = \gamma(\zeta_{1:t-1}) \times \prod_{i=1}^{\bar{q}} N(\boldsymbol{\mu}_i; \hat{\boldsymbol{\mu}}_i(\zeta_{1:t-1}), \boldsymbol{\Sigma}_i(\zeta_{1:t-1})) \tag{25}$$

In (25), $\gamma(\zeta_{1:t-1})$ is the posterior probability of $\zeta_{1:t-1}$ and $\hat{\boldsymbol{\mu}}_k(\zeta_{1:t-1})$ and $\boldsymbol{\Sigma}_k(\zeta_{1:t-1})$ are the posterior mean and covariance matrix for the location of the kth mixture component conditional on $\zeta_{1:t-1}$. We consider a particular label sequence $\zeta_{1:t-1}$ and compute the posterior PDF of $\boldsymbol{\theta}$ and $\zeta_{1:t} = [\zeta_{1:t-1}, i]$ for $i = 0, \ldots, q_{t-1}+1$. Since the history $\zeta_{1:t-1}$ is held constant it is omitted from the notation. Using Bayes' rule and (23), (24) and (25) the posterior PDF can be found as

$$\pi_t(\boldsymbol{\theta}, \zeta_t = i) \propto \tilde{\gamma}(i) \prod_{k=1}^{\bar{q}} N(\boldsymbol{\mu}_k; \hat{\boldsymbol{\mu}}_k(i), \boldsymbol{\Sigma}_k(i)) \tag{26}$$

where, for $i = 0, \ldots, q_{t-1}+1$, $k = 1, \ldots, \bar{q}$,

$$\tilde{\gamma}(i) = \begin{cases} \gamma w_0/V, & i = 0, \\ \gamma w_i N(\boldsymbol{y}_t; \hat{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_i + \boldsymbol{R}), & i = 1, \ldots, q_{t-1}, \\ \gamma \displaystyle\sum_{k=i}^{\bar{q}} w_k \, N(\boldsymbol{y}_t; \hat{\boldsymbol{\mu}}_i, \boldsymbol{\Sigma}_i + \boldsymbol{R}), & i = q_{t-1}+1, \end{cases} \tag{27}$$

$$\hat{\boldsymbol{\mu}}_k(i) = \hat{\boldsymbol{\mu}}_k + \delta_{k-i} \boldsymbol{G}_k (\boldsymbol{y}_t - \hat{\boldsymbol{\mu}}_k), \tag{28}$$

$$\boldsymbol{\Sigma}_k(i) = \boldsymbol{\Sigma}_k - \delta_{k-i} \boldsymbol{G}_k \boldsymbol{\Sigma}_k, \tag{29}$$

where δ_k is Kronecker's delta and $\boldsymbol{G}_k = \boldsymbol{\Sigma}_k(\boldsymbol{\Sigma}_k + \boldsymbol{R})^{-1}$. The quantity $\tilde{\gamma}(i)$ is the un-normalised posterior probability of the label sequence $[\zeta_{1:t-1}, i]$ and $\hat{\boldsymbol{\mu}}_k(i)$ and $\boldsymbol{\Sigma}_k(i)$ are the posterior mean and covariance matrix for the location of the kth mixture component conditional on $[\zeta_{1:t-1}, i]$.

Since the number of valid label sequences $\zeta_{1:t}$ increases exponentially with t the number of label sequences must be managed by, for example, pruning of unlikely label sequences and/or merging of similar label sequences. The former scheme is adopted here.

Equivalent measurements are obtained by using the clusters identified by the most probable label sequence. Let $\zeta^*_{1:m}$ denote the most probable label sequence. Then, the candidates for equivalent measurements are the estimated component means $\hat{\boldsymbol{\mu}}_1(\zeta^*_{1:m}), \ldots, \hat{\boldsymbol{\mu}}_{q(\zeta^*_{1:m})}(\zeta^*_{1:m})$. Before being accepted as equivalent measurements steps are performed to recognise closely spaced targets merged into a single component and remove components which are not due to

targets. Merging of closely spaced targets is identified by checking the number of measurements assigned to each component. If this number is larger than the number that can be expected to be assigned to a component then the component is subjected to a splitting step. This involves applying the mixture estimation algorithm to the measurements assigned to the component. A check of the number of measurements assigned is also used to remove false components, i.e., mixture components which have been assigned outliers. In particular only mixture components which have been assigned at least a certain number of measurements are output as equivalent measurements. The required number of assigned measurements can be determined from the number of sensors and the detection probability.

3.2 Application to target tracking

The application of the clustering algorithm to multiple sensor multiple target tracking is demonstrated in an example involving 10 targets moving in a 150×150 region. There are $d = 6$ sensors which observe the targets at intervals of 0.5s for 50s. The detection probability is $P_D = 0.9$ and the clutter density is $\lambda = 3 \times 10^{-4}$. The result of applying the MC-MHT to the output of the clustering algorithm is shown in Figure 2. In this plot the black dots are the sensor measurements and the red dots are the equivalent measurements provided by the clustering algorithm. The vast majority of the clutter measurements have been removed while many measurements close to the blue target trajectories remain. As a result the tracking results, shown by the green dashed lines, are good. A more thorough performance analysis is presented in the following section.

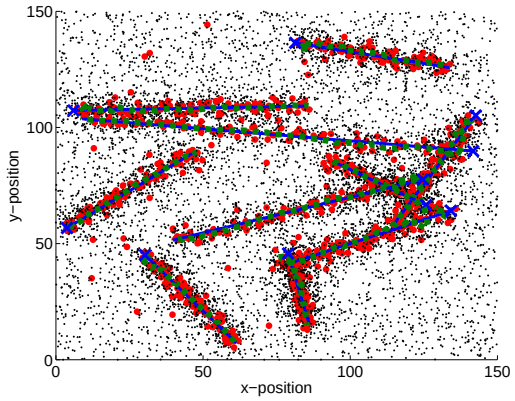

Figure 2: Tracking of 10 targets in clutter with density 3×10^{-4}. The black dots are sensor detections and the red dots are the equivalent sensor detections produced by the clustering algorithm. The blue solid lines are true target trajectories and the green dashed lines are the estimated target trajectories.

4 Performance analysis

We consider a scenario in which the initial target positions are uniformly distributed over a 150m square region. Each target moves with a constant velocity which is drawn from a uniform distribution such that the target remains in the surveillance region throughout the observation interval. The targets are observed for 50s with measurements generated every 0.5s so that 100 measurement scans are generated by each sensor. The target detection probability is $P_D = 0.9$ and the measurement noise has covariance matrix $\boldsymbol{R} = 16\boldsymbol{I}_2$. The number of targets, number of sensors and clutter density are varied between experiments.

Two algorithms are considered in the performance analysis. The first is the MC-MHT with the full collection of measurements. The second algorithm is the MC-MHT with the equivalent measurements produced by the Bayesian clustering algorithm of Section 3. We refer to the former as the full MC-MHT (FMC-MHT) and the latter as the clustered MC-MHT (CMC-MHT). Use of the CMC-MHT requires the selection of suitable values for the effective detection probability, clutter density and measurement noise covariance matrix. The effective detection probability and clutter density are taken from a performance analysis, not show here, of the clustering algorithm. The effective measurement noise covariance matrix is taken as $3\boldsymbol{R}/d$ where d is the number of sensors and $\boldsymbol{R}$ is the single sensor measurement noise covariance matrix.

Three measures of algorithm performance are used: the mean number of targets in track over the first 30 sampling instants, the mean number of targets in track over the remainder of the surveillance period and the RMS position error of targets in track. The performance statistics are collected under a constraint on the mean number of false tracks, controlled by varying the number of detections required for a track to be confirmed. The mean number of false tracks for each scan is constrained to be at most 1/50.

In the first set of experiments there are $r = 10$ targets and analyses are performed for clutter densities of $\lambda = 10^{-4}$ and 3×10^{-4} with $d = 4, 6$ and 8 sensors. Both algorithms are implemented with a sample size of 200. The CMC-MHT uses a maximum of 100 hypothesised label sequences for clustering. The results, obtained by averaging over 100 realisations, are shown in Tables 2 and 3 for clutter densities of 10^{-4} and 3×10^{-4}, respectively. Each Monte Carlo realisation involves randomly generating both the tracking scenario, i.e., initial target positions and target velocities, and the sensor measurements. It is clear that clustering to produce equivalent measurements prior to tracking, i.e., fuse-before-track, provides no performance benefits in the scenarios considered here. As expected, the performances of both algorithms are worse for the higher clutter density. The drop in performance is more severe for the CMC-MHT

than for the FMC-MHT suggesting that tracking with clustered measurements is less robust to increases in the amount of clutter than tracking with all sensor measurements.

The outcomes obtained here are inconsistent with the comments in [3, 4] regarding the superiority of fuse-before-track. It should be noted that the FMC-MHT performs very well here, even with a moderate sample size, so there is little room for improvement anyway. Even so, the relatively poor performance of the CMC-MHT is worth discussing. The CMC-MHT uses equivalent measurements generated from a clustering algorithm. Typically, clustering algorithms have difficulty distinguishing closely spaced clusters. In the multi-sensor tracking context, this means that the assumed detection probability and clutter density are often incorrect in areas surrounding dense target configurations. The latter anomaly is particularly detrimental when the assumed clutter density is very low, as it is here. Then, two consecutive over-estimates of the number of clusters in a group of closely spaced clusters are sufficient for a false track to become strongly established.

Table 2: Tracking results for $r = 10$ targets, $P_D = 0.9$ and $\lambda = 10^{-4}$.

	Number of sensors	FMC-MHT	CMC-MHT
Start true tracks (%)	4	95.7	94.1
	6	97.2	95.5
	8	98.4	95.8
Mean true tracks (%)	4	99.9	98.7
	6	99.9	99.7
	8	99.9	99.9
RMS position error	4	1.26	1.48
	6	1.07	1.26
	8	0.97	1.16

Table 3: Tracking results for $r = 10$ targets, $P_D = 0.9$ and $\lambda = 3 \times 10^{-4}$.

	Number of sensors	FMC-MHT	CMC-MHT
Start true tracks (%)	4	91.3	84.8
	6	93.9	88.0
	8	95.3	89.2
Mean true tracks (%)	4	99.9	98.8
	6	99.9	99.7
	8	99.9	99.7
RMS position error	4	1.38	1.67
	6	1.16	1.38
	8	1.03	1.26

In the second set of experiments the number of targets is increased to $r = 25$. The results of the simulation analysis are shown in Tables 4 and 5 for clutter densities of 10^{-4} and 3×10^{-4}, respectively. It is of interest to compare these results with those shown in Tables 2 and 3 for $r = 10$ targets. Both algorithms perform worse for $r = 25$ targets than for $r = 10$ targets, particularly in the initiation of new tracks. This is because the increased incidence of false tracks for $r = 25$ requires an increase in the number of detections required for confirmation of a track as a target. For $\lambda = 10^{-4}$, the performance deterioration from $r = 10$ to $r = 25$ is significantly greater for the CMC-MHT than for the FMC-MHT. For $\lambda = 3 \times 10^{-4}$, both algorithms degrade similarly in their ability to quickly initiate target tracks as r increases from 10 to 25. The FMC-MHT is more robust than the CMC-MHT in the other aspects of performance. Thus, in the examples considered here, an increase in the severity of the scenario, due to either an increase in the number of targets or in the clutter density, has a greater adverse affect on the CMC-MHT than the FMC-MHT.

Table 4: Tracking results for $r = 25$ targets, $P_D = 0.9$ and $\lambda = 10^{-4}$.

	Number of sensors	FMC-MHT	CMC-MHT
Start true tracks (%)	4	91.2	84.7
	6	94.1	85.0
	8	95.7	85.3
Mean true tracks (%)	4	99.9	94.7
	6	99.9	98.3
	8	99.9	99.0
RMS position error	4	1.57	2.20
	6	1.37	1.82
	8	1.23	1.69

Table 5: Tracking results for $r = 25$ targets, $P_D = 0.9$ and $\lambda = 3 \times 10^{-4}$.

	Number of sensors	FMC-MHT	CMC-MHT
Start true tracks (%)	4	74.2	69.9
	6	83.9	76.9
	8	86.9	78.5
Mean true tracks (%)	4	99.4	94.0
	6	99.7	97.6
	8	99.7	98.2
RMS position error	4	1.94	2.48
	6	1.68	2.10
	8	1.46	1.90

5 Conclusions

Multiple target tracking by fusing data from multiple sensors was considered. Two approaches to multiple sensor fusion were examined. In the first approach all measurements from all sensors are provided to the tracking algorithm. In the second approach, tracking is performed with the output of a clustering algorithm applied to the sensor measurements. This is known as fuse-before-track (FBT). The aim of FBT is to remove clutter measurements prior to tracking. A novel clustering algorithm was developed using a sequential Bayesian formulation. Tracking was performed using a Monte Carlo approximation to the multiple hypothesis tracker (MHT).

It has been suggested in the literature that FBT can outperform tracking with all measurements, particularly if a large number of sensors is available. This was not observed in the performance analysis conducted in this report in examples with up to eight sensors. Tracking with all measurements was significantly better than FBT in all aspects of performance in several scenarios with different target and clutter densities. The relatively poor performance of FBT can be traced to an increase in the number of both missed and false detections when targets are in close proximity.

At least two objections could be raised to the conclusions made here. First, it may be argued that FBT will be most beneficial in scenarios which are more demanding than those considered here i.e., higher clutter and/or target density. In such situations, implementation of the MHT necessarily becomes more approximate. However, these situations are notoriously demanding for clustering algorithms also. The simulation results presented here suggest that the performance of a clustering algorithm will deteriorate more sharply than that of a tracking algorithm as clutter or target density increases. A second objection is that the performance of FBT could be improved by the use of a better clustering algorithm than that proposed here. A survey of the literature, however, suggests that no existing clustering algorithm is able to reliably estimate closely spaced clusters, each containing only a few points, in the presence of a large number of outliers.

There are a number of possible extensions of this work. It was assumed here that all sensors produce measurements of target positions in Cartesian coordinates. It would be interesting to extend the FBT idea to cases where sensors produce different types of measurements, or even just position measurements in different coordinate systems. FBT, as it is currently used, requires a collection of equivalent measurements from the clustering algorithm. For a Bayesian clustering algorithm, this means that the collection of weighted hypotheses which constitute the posterior density must be summarised. At present this is done by simply selecting the hypothesis with the highest weight. This is almost certainly not the most accurate summary of the Bayesian mixture analysis. An approach to FBT which uses the entire posterior density, something akin to a probabilistic sensor collection, would be desirable. Finally, here it has been assumed that the sensors are registered. In practice, perfect registration cannot be assumed and this should be accounted for when performing fusion of measurements from different sensors.

References

[1] S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.

[2] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.

[3] S. Coraluppi and C. Carthel, "Multi-stage data fusion in military surveillance systems," in *Proceedings of the NATO RTO-SET 133*, 2009.

[4] M. Guerriero, S. Coraluppi, and P. Willett, "Analysis of scan and batch processing approaches to static fusion in sensor networks," in *Proceedings of the SPIE*, vol. 6969, 2008.

[5] J. Vermaak, S. Godsill, and P. Perez, "Monte Carlo filtering for multi-target tracking and data association," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 1, pp. 309–332, 2005.

[6] S. Richardson and P. Green, "On Bayesian analysis of mixtures with an unknown number of components," *Journal of the Royal Statistical Society Series B*, vol. 59, no. 4, pp. 731–792, 1997.

[7] M. Stephens, "Dealing with label switching in mixture models," *Journal of the Royal Statistical Society, Series B*, vol. 62, no. 4, pp. 795–809, 2000.

[8] N. Chopin, "Inference and model choice for sequentially ordered hidden Markov models," *Journal of the Royal Statistical Society, Series B*, vol. 69, no. 2, pp. 269–284, 2007.

[9] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[10] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings Part F*, vol. 140, no. 2, pp. 107–113, 1993.

[11] M.-S. Oh, "Monte Carlo integration via importance sampling: dimensionality effect and an adaptive algorithm," *Contemporary Mathematics*, vol. 115, pp. 165–187, 1991.

[12] M. Morelande, "Joint data association using importance sampling," in *Proceedings of the International Conference on Information Fusion*, Seattle, USA, 2009.

[13] G. Celeux, M. Hurn, and C. Robert, "Computational and inferential difficulties with mixture posterior distributions," *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 957–970, 2003.

[14] N. Chopin, "On the equivalence between standard and sequentially ordered hidden Markov models," *Statistics & Probability Letters*, vol. 78, pp. 2171–2174, 2008.