Estimating Unknown Clutter Intensity for PHD Filter

In most of the existing probability hypothesis density (PHD) filters, the clutter is modeled as a Poisson random finite set (RFS) with a known intensity. The clutter intensity is characterized as a product of the average number of clutter (false alarm) points per scan and the probability density of clutter spatial distribution. The PHD filter is generalized to the problem of multi-target tracking (MTT) in clutter with an unknown intensity. In the proposed approach, the unknown clutter intensity is first estimated for the PHD filter. Estimation of the clutter intensity involves the estimation of the average clutter number per scan and the estimation of the clutter density. The clutter density is estimated as finite mixture models (FMM) via either expectation maximum (EM) or Markov chain Monte Carlo (MCMC) algorithm. Then, the estimated intensity is used directly in the PHD filter to perform multi-target detecting and tracking. Monte Carlo (MC) simulation results show that the proposed approach outperforms the naive PHD filter of assuming uniform clutter distribution significantly especially when the nominal clutter model is obviously different from the ground truth.

I. INTRODUCTION

The objective of multi-target tracking (MTT) with a varying number of targets is to jointly estimate the number of targets and their states at each scan from the observation sets in the presence of clutter (false alarms), association uncertainty, detection uncertainty and noise. The traditional approaches to the MTT problem are the association-based algorithms [1–3], such as nearest neighbor (NN) [4], joint probabilistic data association (JPDA) [5], and multiple hypothesis tracking (MHT) [6]. In recent years, the random finite set (RFS) based approaches to MTT [7–9] have attracted substantial interest. The probability hypothesis density (PHD) filter, developed by Mahler [10, 11], has been shown to be a computationally tractable alternative to full multi-target Bayes filter in RFS framework. The sequential Monte Carlo (SMC) implementation for the PHD filter was devised by Zajic, et al. [12], Sidenbladh [13], and Vo, et al. [14]. And the relevant multi-target state extraction techniques from the particle PHD were discussed by Clark, et al. [15], and Liu, et al. [16]. Vo and Ma [17] devised the Gaussian mixture (GM) implementation for the PHD filter under the linear, Gaussian assumption on target dynamics and birth process. Proofs of convergence for the PHD filter were established by Johansen, et al. [18], and Clark, et al. [19, 20]. Clark, et al. [21], Panta, et al. [22–24], and Lin, et al. [25] discussed the data association and track continuity problems for the PHD filter. Punithakumar, et al. [26], and Vo, et al. [27] proposed a multiple model implementation of the PHD filter for tracking maneuvering targets. A physical-space interpretation of the PHD filter was formulated by Erdinc, et al. [28]. Nandakumaran, et al. [29, 30] proposed a PHD smoothing algorithm to improve the capability of the PHD filter. The PHD-based tracking approaches have been successfully applied in many practical problems, such as passive radar tracking [31], terrain vehicle tracking [32, 33], sonar image tracking [34], multi-target visual tracking [35, 36], and simultaneous localisation and mapping (SLAM) problem [37].

In the PHD filter, the clutter is modeled as a Poisson RFS and thus it can be completely characterized by the intensity, which is formed as a product of the average number of clutter points per scan and the probability density of the clutter spatial distribution. Most existing MTT approaches rest on the simplifying assumption that the clutter intensity is known a priori. Especially for the clutter density, it is commonly assumed to be uniformly distributed over the observation region. But in some practical applications, such an assumption is not realistic. For example, in the applications of ground and maritime target tracking, cruise missiles tracking over land, battle target tracking, and visual tracking, both the average clutter number per scan and the clutter spatial distribution might become unknown due to the complicated ground and sea cluttered backgrounds, multi-path reverberation, undiscovered jamming, electronic countermeasure, and other disturbances. When the actual and modeled clutter intensities differ significantly, the performance of the naive approaches to MTT is likely to deteriorate badly. Therefore, how to detect and track multiple targets in clutter with an unknown model is still a very challenging topic.

To the best of the authors' knowledge, some approaches which are not based on the PHD filter have been proposed for this problem. In [38, 39], Musicki, et al. suggested an extension of the integrated probabilistic data association (IPDA) algorithm to the problem. In their method, the number of the clutter within the selection gate at each scan is estimated adaptively under the assumption that

Manuscript received July 9, 2008; revised January 11, 2009; released for publication October 1, 2009.

IEEE Log No. T-AES/46/4/938818.

Refereeing of this contribution was handled by V-N. Vo.

This work was supported by the State Key Program for Basic Research of China (973) (2007CB311006) and the National Natural Science Foundation of China (60602026).

clutter is uniformly distributed over the selection gate. Some approaches for target background estimation in nonlinear filtering are also promising for solving the problem. For example, Rozovsky and his group, who are interested in nonlinear filtering for spatial-temporal processes by application of the stochastic partial differential equations (SPDEs) [40], proposed a method for tracking and detecting targets against complicated sea cluttered backgrounds [41].

As any tracking filter based on models, the PHD filter will also not perform well when its nominal clutter model is significantly different from the actual clutter process. Therefore, this paper aims to solve the problem of multi-targets detecting and tracking in clutter with unknown intensity using a novel extension of the PHD filter. In the proposed method, the clutter intensity is first estimated and then the estimated intensity is used directly in the PHD filter to perform multi-target detecting and tracking. The estimation of the clutter intensity involves the estimation of the average clutter number per scan and the estimation of the clutter density. In general, the latter is much more difficult than the former. The proposed approach estimates the clutter density by application of finite mixture models (FMM) [42]. The parameters of the mixture are estimated via either expectation maximum (EM) [43] or Markov chain Monte Carlo (MCMC) [44] algorithms. The strategy of pruning and merging is used to manage the component number of the mixture.

For the purpose of illustration, we present an example of MTT in clutter with unknown intensity. In this example, the actual clutter density is nonuniform and presents complex multimodality over the observation region. The proposed approach is compared against the naive PHD filter, where the clutter is nominally assumed to be uniformly distributed over the region. Monte Carlo (MC) simulation results show the proposed approach performs MTT correctly although neither the average clutter number per scan nor the clutter density is known. Its performance is significantly better than that of the naive PHD filter which presumes uniform clutter distribution. The latter deteriorates badly especially in estimation of the target number.

The proposed approach can be extended directly to the cardinalized PHD filter (CPHD) [45, 46] and the traditional association-based methods.

The rest of the paper is organized as follows. Section II summarizes the PHD filter and describes the clutter with unknown intensity. In Section III, the clutter intensity is estimated for the PHD filter. Here estimation of the clutter intensity involves the estimation of the average clutter number per scan and the estimation of the clutter density. In Section IV, an example is presented with simulation results to compare the performance of the proposed method with that of the naive PHD filter. Finally, the conclusions and the future work are given in Section V.

II. BACKGROUND AND PROBLEM FORMULATION

A. The PHD Filter

At time k, let $\mathbf{x}_k$ be the single-target state vector and $\mathbf{z}_k$ be the measurement vector. Multi-target states and sensor measurements can be represented naturally as finite sets [8–11]

$$X_k = \{\mathbf{x}_k^1, \ldots, \mathbf{x}_k^{T_k}\}, \qquad Z_k = \{\mathbf{z}_k^1, \ldots, \mathbf{z}_k^{M_k}\} \tag{1}$$

where at time k, T_k is the number of targets and M_k is the number of sensor measurements; the multi-target state set X_k is the union of the surviving target states, the spawned target states, and the spontaneous birth target states; the sensor measurement set Z_k is the union of clutter and target-generated measurements.

Let $f_{k|k}(X_k \mid Z_{1:k})$ denote multi-target posterior density in RFS framework, where $Z_{1:k} := Z_1, \ldots, Z_k$ are the cumulative measurements up to time k. The PHD is the first-order moment of the full multi-target posterior [10].

$$D_{k|k}(\mathbf{x}_k \mid Z_{1:k}) = \int_{\mathbf{x}_k \in X_k} f_{k|k}(X_k \mid Z_{1:k}) \delta X_k \tag{2}$$

where $\int \cdot \delta X_k$ denotes an ordinary set integral.

The integral of the PHD on any region S of the state space gives the expected number of targets contained in S.

$$\hat{T}_{k|k} = \mathrm{E}[T_k] = \int_S D_{k|k}(\mathbf{x}_k \mid Z_{1:k}) d\mathbf{x}_k. \tag{3}$$

At time k, the estimates of the multi-target states $\hat{X}_k = \{\hat{\mathbf{x}}_k^1, \ldots, \hat{\mathbf{x}}_k^{\bar{T}_k}\}$ are derived by rounding the expected number of targets $\hat{T}_{k|k}$ off to the nearest integer $\bar{T}_{k|k}$ and then looking for the $\bar{T}_{k|k}$ largest local maxima of the PHD.

At time k, by modeling the clutter, the birth targets, and the spawned targets as Poisson RFSs with the intensities $\kappa_k(\mathbf{z}_k)$, $\gamma_k(\mathbf{x}_k)$, and $\beta_{k|k-1}(\mathbf{x}_k \mid \mathbf{x}_{k-1})$, respectively, the recursion of PHD filter is derived as [10]

$$\begin{aligned} &D_{k|k-1}(\mathbf{x}_k \mid Z_{1:k-1}) \\ &\quad = \gamma_k(\mathbf{x}_k) + \int [p_{S,k|k-1}(\mathbf{x}_{k-1}) f_{k|k-1}(\mathbf{x}_k \mid \mathbf{x}_{k-1}) \\ &\qquad + \beta_{k|k-1}(\mathbf{x}_k \mid \mathbf{x}_{k-1})] D_{k-1|k-1}(\mathbf{x}_{k-1} \mid Z_{1:k-1}) d\mathbf{x}_{k-1} \end{aligned} \tag{4}$$

$$\begin{aligned} &D_{k|k}(\mathbf{x}_k \mid Z_{1:k}) \\ &\quad = (1 - p_{D,k}(\mathbf{x}_k)) D_{k|k-1}(\mathbf{x}_k \mid Z_{1:k-1}) \\ &\qquad + \sum_{\mathbf{z}_k \in Z_k} \frac{p_{D,k}(\mathbf{x}_k) f_{k|k}(\mathbf{z}_k \mid \mathbf{x}_k) D_{k|k-1}(\mathbf{x}_k \mid Z_{1:k-1})}{\kappa_k(\mathbf{z}_k) + \int p_{D,k}(\mathbf{x}_k) f_{k|k}(\mathbf{z}_k \mid \mathbf{x}_k) D_{k|k-1}(\mathbf{x}_k \mid Z_{1:k-1}) d\mathbf{x}_k} \end{aligned} \tag{5}$$

where at time k, $f_{k|k-1}(\mathbf{x}_k \mid \mathbf{x}_{k-1})$ is the single-target transition density, $f_{k|k}(\mathbf{z}_k \mid \mathbf{x}_k)$ is the single-target likelihood, $p_{S,k|k-1}(\mathbf{x}_{k-1})$ is the probability of target survival, and $p_{D,k}(\mathbf{x}_k)$ is the probability of target detection.

B. The Clutter with Unknown Intensity

Since the clutter is modeled as a Poisson RFS in the PHD filter, it can be completely characterized by the intensity $\kappa_k(\mathbf{z}_k)$, which is formed as

$$\kappa_k(\mathbf{z}_k) = \lambda c(\mathbf{z}_k) \tag{6}$$

where λ is the average clutter number per scan, and $c(\mathbf{z}_k)$ is the probability density of clutter spatial distribution. In this paper, we assume the following.

1) Neither λ nor $c(\mathbf{z}_k)$ is known a priori.
2) λ and $c(\mathbf{z}_k)$ do not vary with time in the period for multi-targets detecting and tracking.

In such a case, we have to first estimate the clutter intensity before performing multi-target detecting and tracking using the PHD filter.

III. CLUTTER INTENSITY ESTIMATION

Since the unknown clutter intensity does not vary with time, it can be pre-estimated for the PHD filter according to the cumulative sensor measurements collected within a window $[1,L]$. The length of the window L should be large enough for ensuring that the estimate of the clutter intensity approximates the ground truth sufficiently well. For the sake of clarity, we first formulate the proposed method for estimating the clutter intensity under the assumption that no targets exist over the observation region within the window $[1,L]$. Actually, as demonstrated in the simulation example, if only the clutter is much more than the target generated measurements, the clutter intensity could still be estimated satisfactorily using the proposed method.

A. Estimating the Average Clutter Number per Scan

Let the set $\mathcal{M} = \{M_1,\ldots,M_L\}$ denote the sequence of the number of measurements at each scan within the window $[1,L]$. Each element of the set $\mathcal{M}$ is assumed to satisfy the Poisson distribution as

$$f(M_k \mid \lambda) = \frac{e^{-\lambda}\lambda^{M_k}}{M_k!}, \qquad k = 1,\ldots,L \tag{7}$$

where λ is the average clutter number per scan.

Therefore, given the set $\mathcal{M}$ with the assumption that the elements of the set are independent, the likelihood for λ is formed as

$$f(\mathcal{M} \mid \lambda) = \prod_{k=1}^{L} f(M_k \mid \lambda) = \frac{e^{-\lambda L}\lambda^{\sum_{k=1}^{L} M_k}}{\prod_{k=1}^{L} M_k!}. \tag{8}$$

The maximum likelihood (ML) estimate of λ is

$$\hat{\lambda}^{\mathrm{ML}} = \arg\max_{\lambda}\{\log f(\mathcal{M} \mid \lambda)\}. \tag{9}$$

It is obtained as an appropriate root of $\partial \log f(\mathcal{M} \mid \lambda)/\partial\lambda = 0$. Finally, the ML estimate of λ is derived by

$$\hat{\lambda}^{\mathrm{ML}} = \frac{1}{L}\sum_{k=1}^{L} M_k = \frac{M}{L} \tag{10}$$

where M denotes the number of the cumulative measurements. $M = \sum_{k=1}^{L} M_k$. Furthermore, it has been proved that it is also the unbiased minimum variance estimation of λ [47].

B. Modeling the Clutter Density as FMM

For the sake of convenience, we omit the time subscript "k" of the measurement in Sections IIIB, C, D, and E. As a consequence, the cumulative measurements within the window $[1,L]$ can be denoted by the set $\mathcal{Z} = \{\mathbf{z}^1,\ldots,\mathbf{z}^M\}$.

We estimate the clutter density $c(\mathbf{z})$ by application of a N-component FMM as

$$c(\mathbf{z}) = f(\mathbf{z} \mid \theta) = \sum_{j=1}^{N} \pi^j f^j(\mathbf{z} \mid \theta^j) \qquad \text{with} \quad \sum_{j=1}^{N} \pi^j = 1 \tag{11}$$

where $\theta = \{\pi^j, \theta^j\}_{j=1}^{N}$ is the parameter set of the mixture, and π^j and θ^j are, respectively, the mixing proportion and the parameter of the jth component. Let the first component be the uniform density over the observation region for fitting the sparse clutter usually generated by the background noise. Let the other components be the normal density for fitting the dense clutter usually generated by some disturbances.

$$\begin{aligned} &f^1(\mathbf{z} \mid \theta^1) = \mathcal{U}(\mathbf{z} \mid \theta^1) \quad \text{with} \quad \theta^1 = \text{the extension of the observation region;} \\ &f^j(\mathbf{z} \mid \theta^j) = \mathcal{N}(\mathbf{z} \mid \theta^j) \quad \text{with} \quad \theta^j = \{\boldsymbol{\mu}^j, \boldsymbol{\Sigma}^j\}, \; j = 2,\ldots,N. \end{aligned} \tag{12}$$

Furthermore, [48] has justified that the mixture comprised of multiple normal densities and one uniform density is robust to a small quantity of the outliers. Here the target-generated measurements can be supposed to be the outliers in the period for estimating the clutter intensity.

Given the set $\mathcal{Z} = \{\mathbf{z}^1,\ldots,\mathbf{z}^M\}$ with the assumption that the elements of the set are independent, the likelihood for θ is formed as

$$f(\mathcal{Z} \mid \theta) = \prod_{i=1}^{M} f(\mathbf{z}^i \mid \theta) = \prod_{i=1}^{M}\sum_{j=1}^{N} \pi^j f^j(\mathbf{z}^i \mid \theta^j). \tag{13}$$

Then, the ML estimate of θ can be derived by

$$\hat{\theta}^{\mathrm{ML}} = \arg\max_{\theta}\{\log f(\mathcal{Z} \mid \theta)\}. \tag{14}$$

C. Estimating the Parameter Set of the FMM via EM

In this subsection, the EM algorithm is used to derive the ML estimate of the parameter θ given the set $\mathcal{Z}$. In the context of FMM, the missing data of the EM algorithm is the set of component labels $\mathcal{E} = \{\mathbf{e}^1, \ldots, \mathbf{e}^M\}$ [42]. Here $\mathbf{e}^i$ is the N-dimensional vector with $e^{ij} = 1$ $(i = 1, \ldots, M,\ j = 1, \ldots, N)$ or 0, according to whether the measurement $\mathbf{z}^i$ does or does not arise from the jth component of the mixture. The complete data set is therefore declared to be $\mathcal{Y} = \{\mathcal{Z}, \mathcal{E}\}$. Then, the complete data log likelihood can be derived as

$$\log f(\mathcal{Y} \mid \theta) = \sum_{i=1}^{M} \sum_{j=1}^{N} e^{ij} \log \pi^j f^j(\mathbf{z}^i \mid \theta^j). \tag{15}$$

For the mixture with known component number, the EM algorithm proceeds iteratively by expectation-step (E-step) and maximization-step (M-step) until convergence. However, in this paper the component number of the mixture is unknown. Several usual model selection methods such as minimum message length (MML) criterion [49], Bayesian information criterion (BIC) [50], Akaike's information criterion (AIC) [51], and so on, could be used to estimate the component number of the mixture. Despite this, the paper proposes to estimate the component number using the strategy of pruning and merging the normal components of the mixture. It is more intuitively natural than the aforementioned criteria-based methods in the application of the clutter density estimation. The strategy only requires that the initial component number of the mixture is bigger than the ground truth. The pruning and merging thresholds η_T, η_M need to be preset in the strategy. The details of the strategy are not presented due to the space limitation. It can be found in [17]. Finally, see Table I for a complete description of the EM algorithm.

Note: It is well known that the EM algorithm in essence pertains to the deterministic methods. Basically, all deterministic algorithms for fitting mixtures with unknown component number have two drawbacks: one is the initialization issue and the other is the boundary of the parameter space. Reference [52] has demonstrated how different initialization strategies and stopping rules can lead to quite different estimates in the context of fitting mixtures of exponential components via the EM algorithm. Indeed, in some cases where the likelihood is unbounded on the edge of the parameter space, the sequence of estimates $\{\theta(t)\}$ generated by the EM algorithm may diverge if $\theta(0)$ is chosen too close to the boundary.

Therefore, in order to enable the EM algorithm to converge to the global maximum (minimum) points as well as possible, the initial component number of the mixture $N(0)$ should be chosen to be large enough so that the initial points could cover the data as sufficiently as possible. Empirically, it is chosen to satisfy $N(0) > \log\varepsilon / \log(1 - \pi^{\min})$ [53], where $\pi^{\min}$ is the given minimum mixing proportion and ε is the convergence threshold for the EM algorithm. Commonly, $N(0)$ is much bigger than the ground truth. Given $N(0)$, the other parameters of the mixture are initialized as illustrated in Table I. The proposed initialization strategy could partly compensate the EM algorithm for its inherent drawbacks. Several other methods such as specification of an initial parameter value, k-means, random starting values, and so on, could also be used to initialize the parameter set of the mixture [42].

D. Estimating the Parameter Set of the FMM via MCMC

Given the prior density $f(\theta)$, we can write the posterior density of the parameter θ in a Bayesian framework as

$$f(\theta \mid \mathcal{Z}) = \frac{\sum_{\mathcal{E}} f(\mathcal{Y} \mid \theta) f(\mathcal{E} \mid \theta) f(\theta)}{\int \sum_{\mathcal{E}} f(\mathcal{Y} \mid \theta) f(\mathcal{E} \mid \theta) f(\theta) d\theta} \tag{16}$$

where $f(\mathcal{E} \mid \theta)$ is the conditional density of the missing data $\mathcal{E}$ given the parameter θ. $f(\mathcal{Z} \mid \theta)$ is the measurement likelihood and $f(\mathcal{Y} \mid \theta)$ is the complete data likelihood.

The maximum a posterior (MAP) estimate of θ can be derived by

$$\hat{\theta}^{\mathrm{MAP}} = \arg\max_{\theta} \{f(\theta \mid \mathcal{Z})\}. \tag{17}$$

In this subsection, the MCMC algorithm is proposed to derive the MAP estimate of θ given $\mathcal{Z}$. The usual MCMC algorithm comprises the Gibbs sampler and Metropolis-Hastings (M-H) sampler [44]. Here $f(\theta \mid \mathcal{Z})$ is approximated through the use of the Gibbs sampler. Then, the MAP estimate can be derived by the approximated posterior density.

A proper prior density $f(\theta)$ should be first provided for MCMC methods. In this paper, the conjugate priors for the mixing proportions, the component means, and the inverse of the component covariance matrices are Dirichlet distribution, normal distribution, and Wishart distribution [42], respectively. Let $\mathcal{D}(\cdot \mid m^1, \ldots, m^n)$ denote the probability density of the Dirichlet distribution with the parameters $m^1, \ldots, m^n$. Let $\mathcal{W}(\cdot \mid a, \mathbf{A})$ denote the probability density of the Wishart distribution with the parameters $a, \mathbf{A}$.

A reversible-jump MCMC (RJMCMC) algorithm [54, 55] could be used to handle the case where the component number of the mixture is unknown. However, the algorithm is perhaps too computationally demanding since it needs to sample the parameter set of the mixture in a high-dimensional space. Therefore, the strategy of pruning and merging is still proposed to manage the component number of the mixture.

TABLE I
Pseudo Code for the EM Algorithm

Input: The set of the cumulative measurements $\mathcal{Z}$

Initialization-step: Set $t := 0$. Initialize the parameter set $\theta(0) = \{\pi^j(0), \theta^j(0)\}_{j=1}^{N(0)}$.

for $j = 2,\ldots,N(0)$, do

$\pi^j(0) = \dfrac{1}{N(0)}$, $m^* \sim \mathcal{U}(\cdot \mid 1,m)$, $\boldsymbol{\mu}^j(0) = z^{\text{round}(m^*)}$, $\boldsymbol{\Sigma}^j(0) = \dfrac{1}{10d}\text{tr}\left(\dfrac{1}{M}\sum_{i=1}^{M}(\mathbf{z}^i - \bar{\mathbf{z}})(\mathbf{z}^i - \bar{\mathbf{z}})^T\right) \cdot \mathbf{I}_d$;

where $\bar{\mathbf{z}} = \sum_{i=1}^{M}\mathbf{z}^i/M$ is the global data mean; $\mathcal{U}(\cdot \mid 1,m)$ denotes the probability density of the uniform distribution with the parameters $1,m$;

round$(\cdot)$ denotes the nearest integer; tr$(\cdot)$ denotes the trace of a matrix;

d is the dimension of the measurement. $\mathbf{I}_d$ is $d \times d$ dimensional identity matrix.

end for j; $\pi^1(0) = 1 - \sum_{j=2}^{N(0)}\pi^j(0)$.

Repeat:

Expectation-step: Calculate the conditional expectation of the missing-data $\hat{\mathcal{E}}(t) = \{\hat{\mathbf{e}}^1(t),\ldots,\hat{\mathbf{e}}^M(t)\}$.

for $j = 1,\ldots,N(t)$ do, for $i = 1,\ldots,M$ do,

$$\hat{e}^{ij}(t) = \text{E}[e^{ij} \mid \mathcal{Z},\theta(t)] = \frac{\pi^j(t)f^j(\mathbf{z}^i \mid \theta^j(t))}{\sum_{j=1}^{N(t)}\pi^j(t)f^j(\mathbf{z}^i \mid \theta^j(t))}$$

end for i; end for j;

Calculate the conditional expectation of the complete data log likelihood given $\mathcal{Z}$ and $\theta(t)$.

$J(\theta,\theta(t)) = \text{E}\{\log f(\mathcal{Y} \mid \theta) \mid \mathcal{Z},\theta(t)\} = \sum_{i=1}^{M}\sum_{j=1}^{N(t)}\hat{e}^{ij}(t)\log\pi^j(t)f^j(\mathbf{z}^i \mid \theta^j(t))$

Maximization-step: Require the global maximization of $J(\theta,\theta(t))$ with respect to θ over the parameter space to give the updated estimate $\hat{\theta}(t)$. $\hat{\theta}(t) = \arg\max_\theta J(\theta,\theta(t))$.

The updates of the mixing proportions, the means, and the covariance matrices are obtained as an appropriate root of $\partial J(\theta,\theta(t))/\partial\theta = 0$.

for $j = 2,\ldots,N(t)$ do,

$$\hat{\pi}^j(t) = \frac{1}{M}\sum_{i=1}^{M}\hat{e}^{ij}(t),\ \hat{\boldsymbol{\mu}}^j(t) = \frac{\sum_{i=1}^{M}\hat{e}^{ij}(t)\mathbf{z}^i}{\sum_{i=1}^{M}\hat{e}^{ij}(t)},\ \hat{\boldsymbol{\Sigma}}^j(t) = \frac{\sum_{i=1}^{M}\hat{e}^{ij}(t)(\mathbf{z}^i - \hat{\boldsymbol{\mu}}^j(t))(\mathbf{z}^i - \hat{\boldsymbol{\mu}}^j(t))^T}{\sum_{i=1}^{M}\hat{e}^{ij}(t)}.$$

end for j; $\hat{\pi}^1(t) = 1 - \sum_{j=2}^{N(t)}\hat{\pi}^j(t)$.

Pruning and merging step: (The details can be found in [17].)

$\theta(t+1) = \{\pi^j(t+1),\theta^j(t+1)\}_{j=1}^{N(t+1)} \underset{\text{merge}}{\overset{\text{prune}}{\longleftarrow}} \hat{\theta}(t) = \{\hat{\pi}^j(t),\hat{\theta}^j(t)\}_{j=1}^{\hat{N}(t)}$. Set $t := t+1$.

Until $|J(\theta,\theta(t+1)) - J(\theta,\theta(t))| < \varepsilon J(\theta,\theta(t))$, where ε is the convergence threshold.

Output: The set of the estimated parameters $\hat{\theta} = \theta(t+1)$.

TABLE II
Pseudo Code for the Gibbs Sampler

Input: The set of the cumulative measurements $\mathcal{Z}$

Initialization-step: Initialize the parameter set $\theta(0) = \{\pi^j(0),\theta^j(0)\}_{j=1}^{N(0)}$. The initialization of the Gibbs sampler is given the same as that of the EM algorithm in Table I except $N(0)$.

for $t = 0,\ldots,t_m$ do, for $j = 2,\ldots,N(t)$ do,

Sampling-step: Sample the mixing proportions, the means and the covariance matrices from the following conditional distributions.

$$(\tilde{\pi}^1(t),\ldots,\tilde{\pi}^{N(t)}(t)) \sim \mathcal{D}(\cdot \mid M^1(t),\ldots,M^{N(t)}(t)),\ \tilde{\boldsymbol{\Sigma}}^j(t)^{-1} \sim \mathcal{W}\left(\cdot \,\middle|\, M^j(t) + r^j, \frac{\boldsymbol{\Sigma}^j(t)^{-1}}{M^j(t) + r^j}\right),\ \tilde{\boldsymbol{\mu}}^j(t) \sim \mathcal{N}(\cdot \mid \boldsymbol{\mu}^j(t), \frac{\tilde{\boldsymbol{\Sigma}}^j(t)}{M^j(t) + \kappa^j}).$$

where $M^j(t) = \sum_{i=1}^{M}e^{ij}(t)$ is the number of the measurements that arise from the jth component of the mixture.

Calculate the condition expectation of the missing data $\hat{\mathcal{E}}(t) = \{\hat{\mathbf{e}}^1(t),\ldots,\hat{\mathbf{e}}^M(t)\}$.

for $i = 1,\ldots,M$ do, $\hat{e}^{ij}(t) = \dfrac{\tilde{\pi}^j(t)f^j(\mathbf{z}^i \mid \tilde{\theta}^j(t))}{\sum_{j=1}^{N(t)}\tilde{\pi}^j(t)f^j(\mathbf{z}^i \mid \tilde{\theta}^j(t))}$. end for i;

Sample the missing data $\tilde{\mathcal{E}}(t) = \{\tilde{\mathbf{e}}^1(t),\ldots,\tilde{\mathbf{e}}^M(t)\}$ according to $\hat{\mathcal{E}}(t) = \{\hat{\mathbf{e}}^1(t),\ldots,\hat{\mathbf{e}}^M(t)\}$.

for $i = 1,\ldots,M$ do,

$\tau \sim \text{Mult}_{N(t)}(1,\hat{\mathbf{e}}^i(t))$, where $\hat{\mathbf{e}}^i(t) = [\hat{e}^{i1}(t),\ldots,\hat{e}^{iN(t)}(t)]$.

Set $\tilde{\mathbf{e}}^i(t) := [\underbrace{0,\ldots,1}_{\tau\text{th}},0,\ldots,0]$.

end for i;

Update-step: Update the mixing proportions, the means and the covariance matrices based on $\tilde{\mathcal{E}}(t)$.

$$\hat{\pi}^j(t) = \frac{1}{M}\sum_{i=1}^{M}\tilde{e}^{ij}(t),\ \hat{\boldsymbol{\mu}}^j(t) = \frac{\sum_{i=1}^{M}\tilde{e}^{ij}(t)\mathbf{z}^i}{\sum_{i=1}^{M}\tilde{e}^{ij}(t)},\ \hat{\boldsymbol{\Sigma}}^j(t) = \frac{\sum_{i=1}^{M}\tilde{e}^{ij}(t)(\mathbf{z}^i - \hat{\boldsymbol{\mu}}^j(t))(\mathbf{z}^i - \hat{\boldsymbol{\mu}}^j(t))^T}{\sum_{i=1}^{M}\tilde{e}^{ij}(t)}.$$

Pruning and merging step: (The details can be found in [17].)

$\theta(t+1) = \{\pi^j(t+1),\theta^j(t+1)\}_{j=1}^{N(t+1)} \underset{\text{merge}}{\overset{\text{prune}}{\longleftarrow}} \hat{\theta}(t) = \{\hat{\pi}^j(t),\hat{\theta}^j(t)\}_{j=1}^{\hat{N}(t)}$.

end for j; end for t;

Output: The set of the estimated parameters $\hat{\theta} = \theta(t+1)$.

The Gibbs sampler proceeds iteratively until the Markov chain converges to stationarity. See Table II for a complete description of the Gibbs sampler. The superscript "$\sim$" denotes the sample.

Note:

1) The number of iterations t_m could be given according to the rate of convergence of the Markov chain to stationarity. Reference [56] gave an upper bound on the number of iterations needed for ensuring that the estimate $\hat{\theta}$ approximates the true value sufficiently well. More details about choosing the iteration number of the Gibbs sampler can be found in [44, 56].

2) r^j and κ^j ($j = 2,\ldots,N(t)$) are the auxiliary parameters of the Wishart distribution and the normal distribution. They are chosen for ensuring that the Bayesian estimates are relatively insensitive to reasonable changes. In this paper, they are chosen empirically as $r^j = 5$, $\kappa^j = 1$ ($j = 2,\ldots,N(t)$).

3) In opposition to the EM algorithm, the MCMC algorithm pertains to the stochastic and sampling methods. In the MCMC algorithm, Bayesian sampling produces an ergodic Markov chain with stationary distribution of the interesting parameters by the ergodic theorem. If only the Markov chain has enough length, the MCMC algorithm could be expected to derive the sufficient statistical information of the posterior distribution of the parameters. Therefore, comparing with the EM algorithm, the parameter estimates of the MCMC algorithm are more likely to jump out of the local extreme and converge to a global maximum (minimum) point for the complex multimodality distribution. And as a consequence, the MCMC algorithm is insensitive to the initialization. The initial component number of the mixture $N(0)$ only needs to be a little bigger than the ground truth for the MCMC algorithm. For the sake of comparison, the initialization of the Gibbs sampler for the other parameters is given the same as that of the EM algorithm, which has been illustrated in Table I.

Finally, the estimated clutter intensity can be derived by the product of the estimated average clutter number per scan and the estimated clutter density.

$$\hat{\kappa}_k(\mathbf{z}) = \hat{\lambda}\hat{c}(\mathbf{z}) = \hat{\lambda}\left(\hat{\pi}^1\mathcal{U}(\mathbf{z}\mid\theta^1) + \sum_{j=2}^{\hat{N}}\hat{\pi}^j\mathcal{N}(\mathbf{z}\mid\hat{\theta}^j)\right). \tag{18}$$

IV. SIMULATION RESULTS

A. Simulation Scenario Description

Consider a two-dimensional scenario with an unknown and time-varying number of targets observed over the surveillance region $[-600,900]\times[-600,900]$ (in meters) for a period of 100 time steps. The state vector of a single target is $\mathbf{x}_k = [p_{x,k},\dot{p}_{x,k},\ddot{p}_{x,k},p_{y,k},\dot{p}_{y,k},\ddot{p}_{y,k}]^T$. $[p_{x,k},p_{y,k}]^T$, $[\dot{p}_{x,k},\dot{p}_{y,k}]^T$, and $[\ddot{p}_{x,k},\ddot{p}_{y,k}]^T$ represent the position, velocity, and acceleration of the target at time k, respectively. Each target has survival probability $p_{S,k|k-1}(\mathbf{x}_{k-1}) = 0.95$ and follows the linear dynamics

$$\mathbf{x}_k = \mathbf{F}_k\mathbf{x}_{k-1} + \mathbf{w}_k \tag{19}$$

where $\mathbf{F}_k$ is the state transition matrix, and $\mathbf{w}_k$ is the process noise. In this simulation, let $\mathbf{F}_k$ be the constant acceleration (CA) model [57] and $\mathbf{w}_k$ be the independent and identically distributed (IID) zero-mean Gaussian white noise with the covariance matrix $\mathbf{Q}_k$. $\mathbf{w}_k \sim \mathcal{N}(\cdot\mid\mathbf{0},\mathbf{Q}_k)$.

$$\begin{aligned}
\mathbf{F}_k &= \begin{bmatrix}\mathbf{F}_{\mathrm{CA}} & \\ & \mathbf{F}_{\mathrm{CA}}\end{bmatrix}\\
\mathbf{Q}_k &= \begin{bmatrix}\mathbf{Q}_{\mathrm{CA}} & \\ & \mathbf{Q}_{\mathrm{CA}}\end{bmatrix}\\
\mathbf{F}_{\mathrm{CA}} &= \begin{bmatrix}1 & \Delta & \frac{\Delta^2}{2}\\ & 1 & \Delta\\ & & 1\end{bmatrix}\\
\mathbf{Q}_{\mathrm{CA}} &= \sigma_w^2\cdot\begin{bmatrix}\frac{\Delta^4}{4} & \frac{\Delta^3}{2} & \frac{\Delta^2}{2}\\ \frac{\Delta^3}{2} & \frac{\Delta^2}{2} & \Delta\\ \frac{\Delta^2}{2} & \Delta & 1\end{bmatrix}
\end{aligned} \tag{20}$$

where σ_w is the standard deviation of $\mathbf{w}_k$, $\sigma_w = 0.01$ m/s^2, Δ is the sampling period, $\Delta = 1$ s. Other models such as the constant velocity (CV) model, the Singer model, the "current" model, and so on [57], can be also accommodated within our framework for describing the target motion.

The sensor is located on $[0,0]^T$ and its detection probability is $p_{D,k}(\mathbf{x}_k) = 0.95$. The single-target measurement follows the linear model.

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \tag{21}$$

where $\mathbf{H}_k$ is the observation matrix, and $\mathbf{v}_k$ is the measurement noise. Let $\mathbf{v}_k$ be the IID zero-mean Gaussian white noise with the covariance matrix $\mathbf{R}_k$. $\mathbf{v}_k \sim \mathcal{N}(\cdot\mid\mathbf{0},\mathbf{R}_k)$.

$$\begin{aligned}
\mathbf{H}_k &= \begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\end{bmatrix}\\
\mathbf{R}_k &= \sigma_v^2\cdot\begin{bmatrix}1 & \\ & 1\end{bmatrix}
\end{aligned} \tag{22}$$

where σ_v is the standard deviation of $\mathbf{v}_k$. In this simulation, it is given as $\sigma_v = 10$ m.

The clutter is modeled as a Poisson RFS with the intensity $\kappa_k(\mathbf{z}_k) = \lambda c(\mathbf{z}_k)$. In this example, the actual average clutter number per scan is $\lambda = 50$. The actual

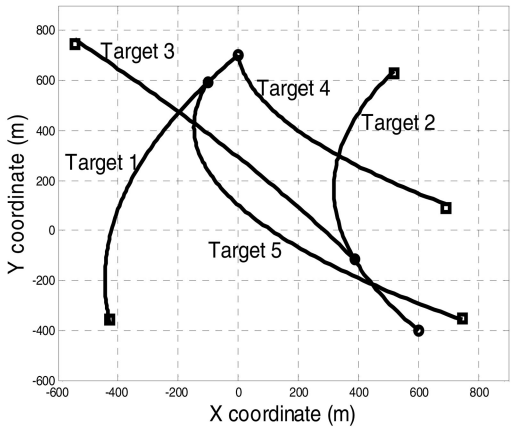

Fig. 1. True target trajectories. "○": locations at which targets are born; "□": locations at which targets die.

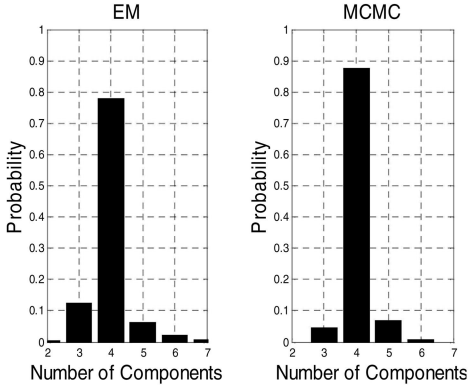

Fig. 2. Histogram for number of components of estimated clutter density based on 300 MC runs.

clutter density is

$$c(\mathbf{z}_k) = 0.2\mathcal{U}(\mathbf{z}_k \mid \theta^1) + 0.3\mathcal{N}(\mathbf{z}_k \mid \boldsymbol{\mu}^2, \boldsymbol{\Sigma}^2) + 0.25\mathcal{N}(\mathbf{z}_k \mid \boldsymbol{\mu}^3, \boldsymbol{\Sigma}^3) + 0.25\mathcal{N}(\mathbf{z}_k \mid \boldsymbol{\mu}^4, \boldsymbol{\Sigma}^4) \tag{23}$$

where

$$\theta^1 = [-600, 900] \times [-600, 900], \qquad \boldsymbol{\mu}^2 = [400, 300]^T$$

$$\boldsymbol{\mu}^3 = [300, -100]^T, \qquad \boldsymbol{\mu}^4 = [-100, 500]^T;$$

$$\boldsymbol{\Sigma}^2 = \begin{bmatrix} 1.44 & \\ & 1 \end{bmatrix} \times 10^4, \qquad \boldsymbol{\Sigma}^3 = \begin{bmatrix} 1 & \\ & 0.64 \end{bmatrix} \times 10^4$$

$$\boldsymbol{\Sigma}^4 = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix} \times 10^4.$$

Fig. 1 shows the true target trajectories. Target 1 is born at 1 s and dies at 90 s; target 2 is born at 1 s and dies at 70 s; target 3 is spawned from target 2 at 20 s and dies at 80 s; target 4 is born at 40 s and dies at 100 s; target 5 is spawned from target 1 at 10 s and dies at 100 s.

For comparison, 300 MC runs are performed on the same clutter intensity and target trajectories but with independently generated clutter and target-generated measurements for each trial.

B. Simulation Results for Clutter Intensity Estimation

In this simulation example, since the number of clutter per scan is much larger than the number of target-generated measurements per scan, the clutter intensity can be estimated directly according to the cumulative sensor measurements within the window $[1, L]$ in the period for multi-target detecting and tracking. We choose the length of the window $L = 6$ empirically, which is large enough for estimating the clutter intensity sufficiently accurately.

The ML estimate of the average clutter number per scan $\hat{\lambda}$ can be derived by (10) in each trial. The MC average of the mean of $\hat{\lambda}$ is 52.2, while the MC average of the standard deviation of $\hat{\lambda}$ is 8.5.

The aforementioned EM and MCMC algorithms, with the pruning threshold $\eta_T = 0.05$ and the merging threshold $\eta_M = 4$, are applied for estimating the clutter density. In the EM method, the initial component number of the mixture is $N(0) = 50$ and the convergence threshold is $\varepsilon = 10^{-5}$, while $N(0) = 7$ and the number of iterations is $t_m = 20$ in the MCMC method. 300 MC runs are performed for EM and MCMC, respectively. The simulation results of the estimated clutter density are as follows.

Fig. 2 plots the histogram for the number of components of the estimated clutter density via either the EM or MCMC algorithm based on 300 MC runs. The figure shows that the EM algorithm correctly identifies the actual component number in 78% of the trials, while the MCMC algorithm does this in 88% of the trials (here the actual component number is $N = 4$).

The estimated clutter densities from the EM and MCMC algorithms in several trials are presented along with the actual clutter density by plotting the 95% ellipses of each component. For clarity, the cumulative sensor measurements within the window $[1, L]$ are also plotted in these figures.

Fig. 3(a) and (b) show the case where the estimated component number is consistent with the actual component number. It can be clearly seen from the two figures that the estimated clutter density approximates the actual clutter density comparably well in such a case. Furthermore, the estimation accuracy of the MCMC algorithm seems to be better than that of the EM algorithm. Fig. 3(c) and (d) show the case where the estimated component number is inconsistent with the actual component number. In this case, the estimated and actual clutter densities have obvious deviation.

Figs. 2 and 3 implicitly demonstrate that the MCMC algorithm outperforms the EM algorithm slightly for estimating the clutter density in this simulation example. A possible reason is that the

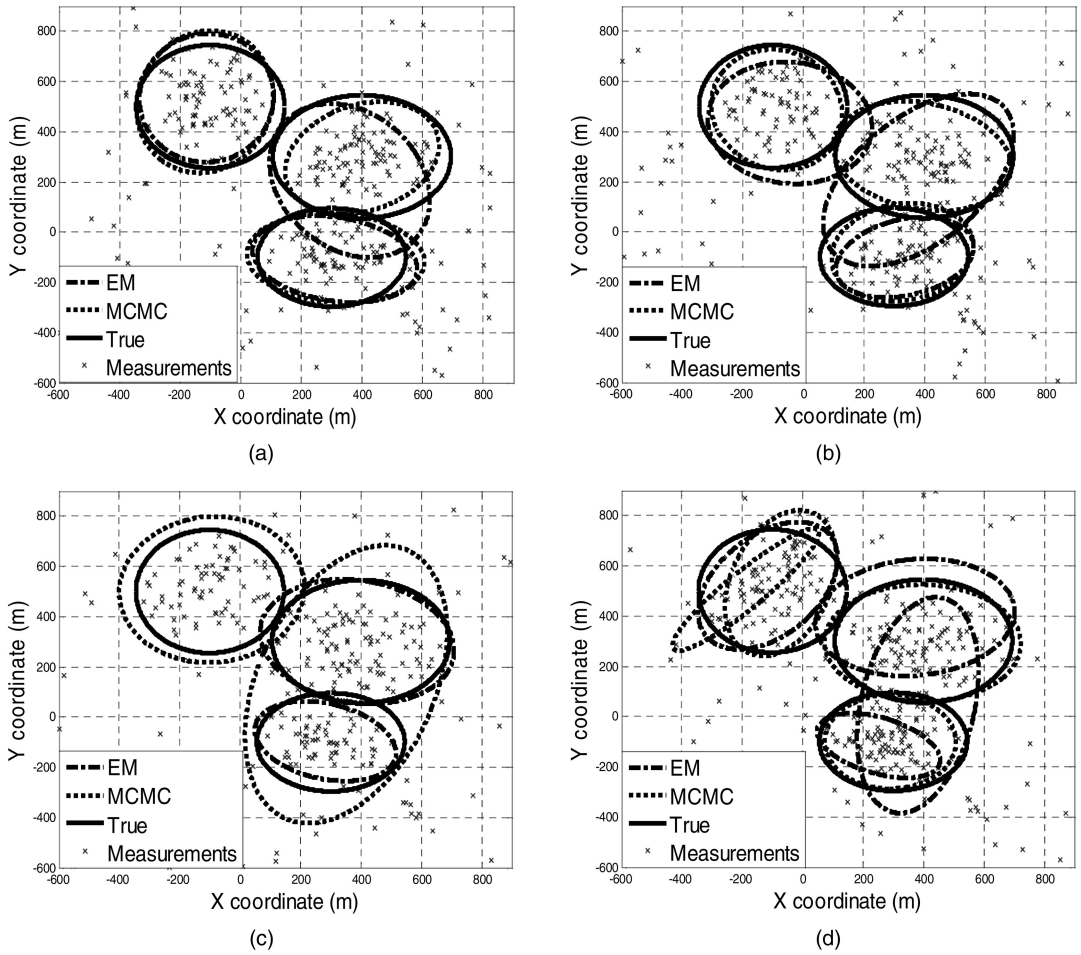

Fig. 3. Actual and estimated clutter density in several trials. (a, b) Number of components of estimated clutter density is consistent with the actual number. (c, d) Number of components of estimated clutter density is inconsistent with the actual number.

MCMC algorithm is more suitable than the EM algorithm for fitting the complex multimodality distribution.

The computational requirement of the EM and MCMC algorithms for estimating the clutter density is compared via an indication of CPU processing time. Based on 300 MC runs, the average computational times of a fairly optimal Matlab implementation for the EM and MCMC methods on 1.83 GHz AMD Athlon (tm) XP 2500+ processor and 512MB RAM are, respectively, 10.2 s and 30.4 s. Obviously MCMC consumes much more time than EM. As a result, although the MCMC-based techniques are perhaps more effective than the EM algorithm for fitting the clutter distribution, we think that they are still too computationally demanding to be applied in MTT.

C. Simulation Results for Multi-Target Tracking

The GM-PHD filter [17] is used to perform multi-target detecting and tracking in this simulation. In the GM-PHD filter, the intensity of target birth is

$$\gamma_k(\mathbf{x}_k) = 0.1\mathcal{N}(\mathbf{x}_k \mid \mathbf{m}_\gamma^1, \mathbf{P}_\gamma^1) + 0.1\mathcal{N}(\mathbf{x}_k \mid \mathbf{m}_\gamma^2, \mathbf{P}_\gamma^2) \tag{24}$$

where $\mathbf{m}_\gamma^1 = [0,0,0,700,0,0]^T$, $\mathbf{m}_\gamma^2 = [600,0,0,-400,0,0]^T$, $\mathbf{P}_\gamma^1 = \mathbf{P}_\gamma^2 = \mathrm{diag}(100,400,1,100,400,1)$ ($\mathrm{diag}(\cdot)$ denotes the diagonal matrix).

The intensity of target spawn is

$$\beta_{k|k-1}(\mathbf{x}_k \mid \mathbf{x}_{k-1}) = 0.05\mathcal{N}(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{Q}_\beta) \tag{25}$$

where $\mathbf{Q}_\beta = \mathrm{diag}(100,100,1,100,100,1)$.

We compare the performance of the following three PHD filters in the simulation example.

1) EM-PHD: perform multi-target detecting and tracking using the GM-PHD filter with the estimated clutter intensity derived by the EM algorithm.

2) MCMC-PHD: perform multi-target detecting and tracking using the GM-PHD filter with the estimated clutter intensity derived by the MCMC algorithm.

3) Naive-PHD: perform multi-target detecting and tracking using the GM-PHD filter with the nominal clutter intensity which assumes that the clutter is uniformly distributed over the observation region.

In the three PHD filters, the estimates of the average clutter number per scan $\hat{\lambda}$ are all derived by (10).

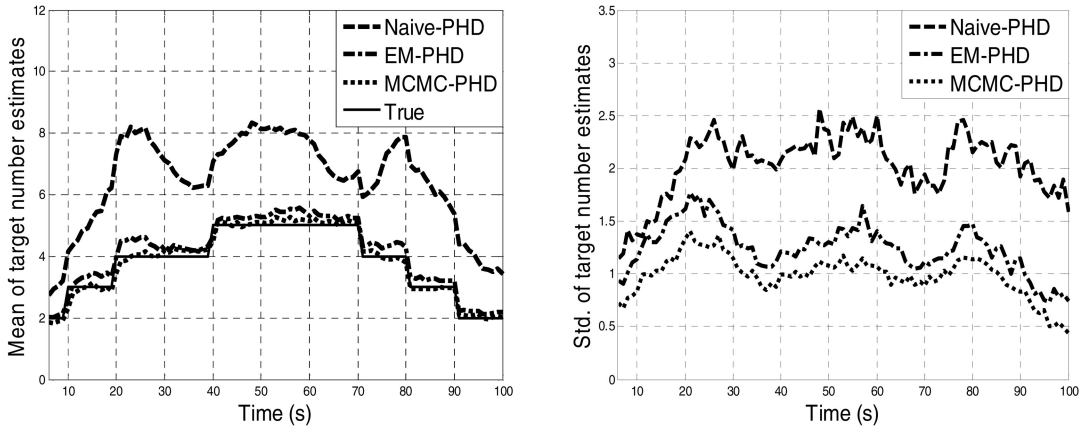

Fig. 4. 300 MC run average of cardinality statistics versus time for the three PHD filters. (a) Means of target number estimates. (b) Standard deviations of target number estimates.

In Fig. 4(a), the true target number at each scan is shown along with the MC average of the means of the target number estimates for the three PHD filters. The plot demonstrates that both the EM-PHD and the MCMC-PHD converge to the true target number relatively well. Furthermore, the MCMC-PHD outperforms the EM-PHD slightly. Nevertheless, the naive-PHD deteriorates badly and generates an obvious bias in the target number estimates. Many false targets estimates are generated in the naive-PHD due to the incorrect assumption that clutter is uniformly distributed. Moreover, as plotted in Fig. 4(a), the target number estimates from the naive-PHD fluctuate obviously versus time. The fluctuation is mainly generated by the nonuniform clutter distribution. When more targets move close to the dense clutter areas, the bias of target number estimates from the naive-PHD becomes the more significant.

In Fig. 4(b), the MC average of the standard deviations of the target number estimates at each scan is shown for the three PHD filters. The plot demonstrates that the standard deviations from the MCMC-PHD are slightly smaller than those from the EM-PHD. Nevertheless, they are both significantly smaller than those from the naive-PHD. The smaller the standard deviations are, the more reliable and accurate the filter's estimates are. Similarly, due to the nonuniform clutter distribution, when more targets move close to the dense clutter areas, all of the curves in Fig. 4(b) become bigger. It indicates that the target number estimates in the dense clutter areas are worse than the estimates in the sparse clutter areas for the three PHD filters.

In Figs. 5(a)–(c) the position estimates of one trial are shown along with the true target trajectories for the three PHD filters. In this trial, the component number of the estimated clutter density via either EM or MCMC is consistent with the ground truth. Fig. 5(a) clearly demonstrates that lots of false target estimates are generated by the naive-PHD when the targets move close to the dense clutter areas. The EM-PHD and the MCMC-PHD, as plotted in Fig. 5(b) and (c), still perform multi-target detecting and tracking correctly. It can be further seen that the target position estimates from the EM-PHD and the MCMC-PHD become appreciably poor when the targets move close to the dense clutter areas.

The Wasserstein distance (WD) [58] and the circular position error probability (CPEP) [59] for a radius of $r = 20$ m are used for performance evaluation of the three PHD filters. Let $\hat{X}_k$ and X_k be the finite sets of the estimated and the true multi-target states at time k. The L^p is $1 \leq p < \infty$. In this simulation example, it is given as $p = 2$. WD between the two sets is defined by

$$d_k^p(\hat{X}_k, X_k) = \min_{\mathbf{C}_k} \sqrt[p]{\sum_{i=1}^{|\hat{X}_k|} \sum_{i=1}^{|X_k|} C_k^{ij} \|\hat{\mathbf{x}}_k^i - \mathbf{x}_k^j\|_2^p} \tag{26}$$

where the minimum is taken over the set of all transportation matrices $\mathbf{C}_k$ (a transportation matrix is one whose entries C_k^{ij} satisfy $C_k^{ij} \geq 0$, $\sum_{j=1}^{|X_k|} C_k^{ij} = 1/|\hat{X}_k|$, $\sum_{i=1}^{|\hat{X}_k|} C_k^{ij} = 1/|X_k|$). $\|\cdot\|_2$ denotes the 2-norm. $|\cdot|$ denotes the cardinality of a set.The WD is not defined if either X_k or $\hat{X}_k$ is empty.

The CPEP at time k is defined by

$$\text{CPEP}(r) = \frac{1}{|X_k|} \sum_{\mathbf{x}_k \in X_k} \Pr\{\|\mathbf{H}_k \hat{\mathbf{x}}_k - \mathbf{H}_k \mathbf{x}_k\|_2 > r,\ \forall \hat{\mathbf{x}}_k \in \hat{X}_k\}. \tag{27}$$

The MC average of the WD of the multi-target positions is shown in Fig. 6(a) for the three PHD filters while the MC average of the CPEP is shown in Fig. 6(b) for the three PHD filters. And for more clarity, their time averages are also given in Table III.

The WD tends to penalize sets of different cardinalities. Fig. 6(a) and Table III show that the

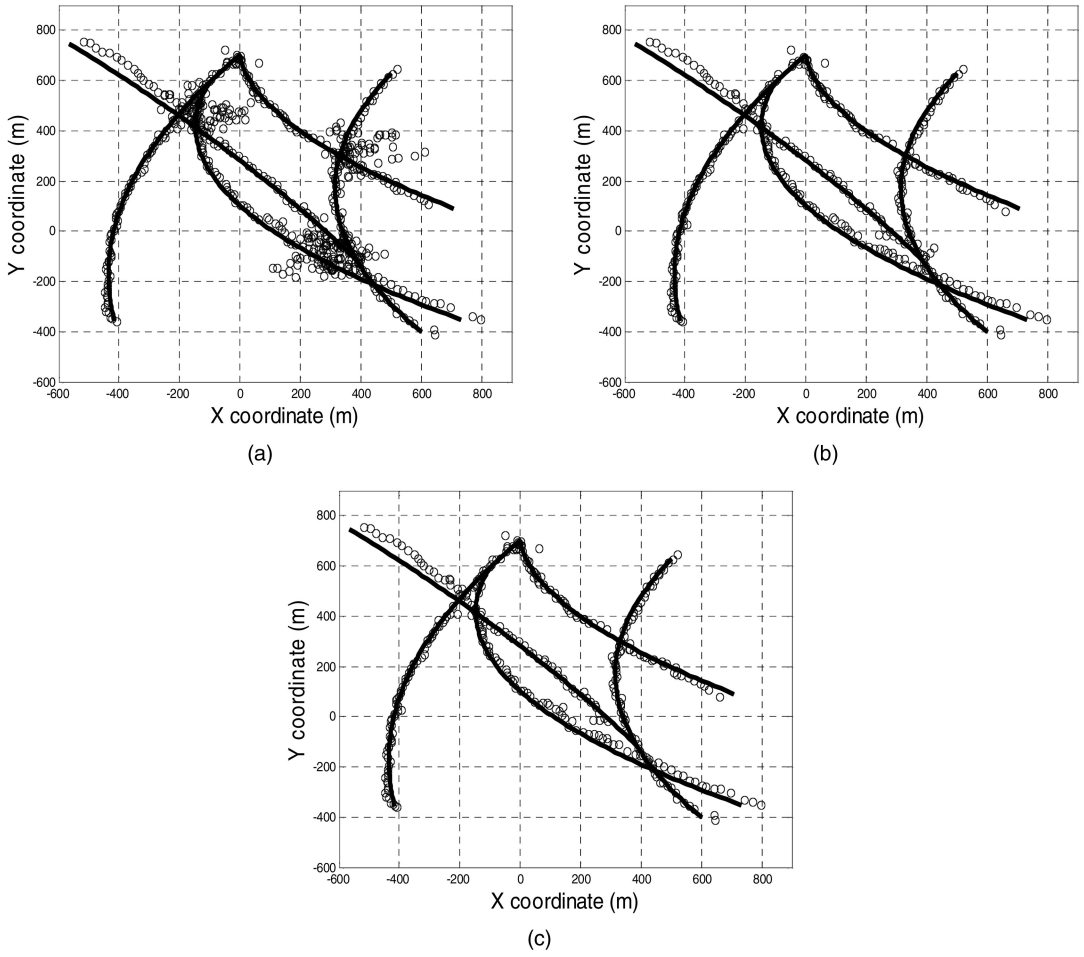

Fig. 5. Position estimates from the three PHD filters. (a) Naive-PHD. (b) EM-PHD. (c) MCMC-PHD. "∘": estimated target positions. Real lines: true target trajectories.

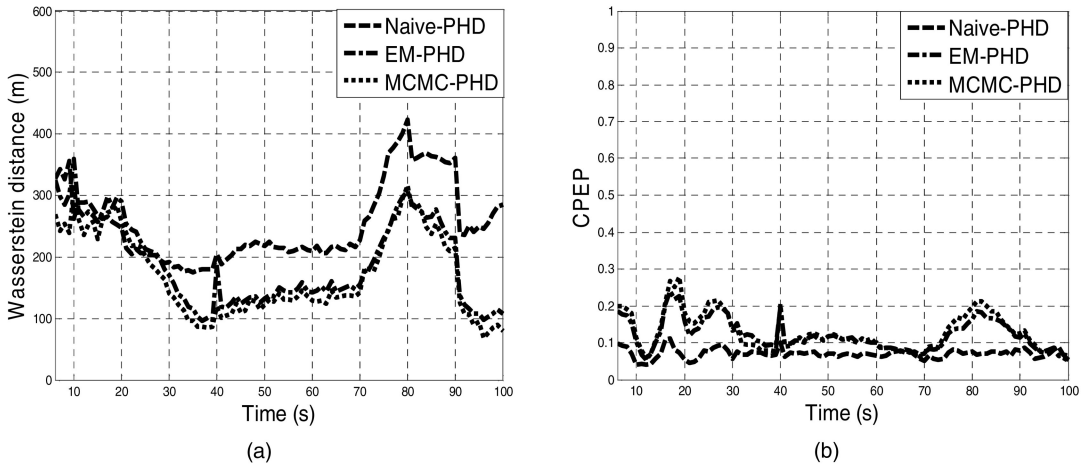

Fig. 6. Comparison of performance measures for the three PHD filters. (a) WD of the multi-target positions versus time. (b) CPEP (with $r = 20$ m) versus time.

WD of the naive-PHD is obviously bigger than that of the EM-PHD and the MCMC-PHD in most of the period for tracking. A reasonable explanation is that the latter have more accurate target number estimates than the former. So it incurs a lower WD on average. For the three PHD filters, the fluctuation of the WD is generated by the change of the actual target number

TABLE III
Time Average of the WD and the CPEP over 100 Time Steps for the Three PHD Filters Based on 300 MC Runs

	Naive-PHD	EM-PHD	MCMC-PHD
WD (in m)	259.7	194.1	173.3
CPEP	0.08	0.12	0.13

and the nonuniform clutter distribution. When more targets move close to the dense clutter area, the WD becomes the bigger.

CPEP tends to penalize the loss of the individual target estimates. It is insensitive to the false target estimates. Fig. 6(b) and Table III show that the CPEP of the naive-PHD is slightly smaller than that of the EM-PHD and the MCMC-PHD. However, the price is that the naive-PHD generates more false target estimates than the EM-PHD and the MCMC-PHD. Furthermore, similar to the WD, when the targets move close to the dense clutter area, the CPEP of the EM-PHD and the MCMC-PHD becomes bigger. A reasonable explanation is that the target estimates in the dense clutter areas are more easily lost than the estimates in sparse clutter areas for the two filters.

V. CONCLUSION

This paper generalizes the PHD filter to the MTT problem in clutter with unknown but temporally unvarying intensity. The proposed method first estimates the clutter intensity using FMM. The parameter set of the mixture is estimated via either EM or MCMC algorithm. And then the estimated clutter intensity is used directly in the PHD filter to perform multi-target detecting and tracking. The simulation results illustrate a significant improvement in performance over a naive PHD filter by assuming uniform clutter intensity.

When the clutter intensity varies slowly with time and the clutter is much more than the target-generated measurements, the proposed method for the clutter intensity estimation can still perform satisfactorily using a sliding window. Contrarily, when the clutter intensity varies quickly with time or the clutter is of the same order of magnitude as the target-generated measurements, the performance of the proposed method might deteriorate badly. The MTT problem in this situation would become much more complicated. It will be further studied in the future work.

FENG LIAN
CHONGZHAO HAN
WEIFENG LIU
SKLMSE Laboratory
Xi'an Jiaotong University
Xi'an, Shannxi
People's Republic of China
710049
E-mail: (lianfeng1981@gmail.com)

REFERENCES

[1] Bar-Shalom, Y. and Fortman, T. E.
Tracking and Data Association.
Burlington, MA: Academic Press, 1988.

[2] Blackman, S.
Multiple Target Tracking with Radar Applications.
Norwood, MA: Artech House, 1986.

[3] Blackman, S. and Popoli, R.
Design and Analysis of Modern Tracking Systems.
Norwood, MA: Artech House, 1999.

[4] Singer, R. A. and Stein, J. J.
An optimal tracking filter for processing sensor data of imprecisely determined origin in surveillance system.
In *Proceedings of the Tenth IEEE Conference on Decision and Control*, 1971, 171–175.

[5] Fortmann, T. E., Bar-Shalom, Y., and Scheffe, M.
Sonar tracking of multiple targets using joint probabilistic data association.
IEEE Journal of Oceanic Engineering, **OE-8** (1983), 173–184.

[6] Blackman, S.
Multiple hypothesis tracking for multitarget tracking.
IEEE Transactions on Aerospace and Electronic Systems, **40**, 1 (2004), 5–18.

[7] Goodman, I., Mahler, R., and Nguyen, H.
Mathematics of Data Fusion.
Norwell, MA: Kluwer Academic Publishing Co., 1997.

[8] Mahler, R.
An Introduction to Multisource-Multitarget Statistics and Applications.
Lockheed Martin Technical Monograph, 2000.

[9] Mahler, R.
Random set theory for tracking and identification.
In M. E. Liggins, D. Hall, and J. Llinas (Eds.), *The Handbook of Sensor Data Fusion: Theory and Practice* (2nd ed.), Boca Raton, FL: CRC Press, 2008, ch. 16.

[10] Mahler, R.
Multi-target Bayes filtering via first-order multi-target moments.
IEEE Transactions on Aerospace and Electronic Systems, **39**, 4 (2003), 1152–1178.

[11] Mahler, R.
Statistical Multisource Multitarget Information Fusion.
Norwood, MA: Artech House 2007.

[12] Zajic, T. and Mahler, R.
A particle-systems implementation of the PHD multitarget tracking filter.
In I. Kadar (Ed.), *Proceedings of SPIE, Signal Processing, Sensor Fusion, and Target Recognition XII*, vol. 5096, 2003, 291–299.

[13] Sidenbladh, H.
Multi-target particle filtering for the probability hypothesis density.
In *Proceedings of the 6th International Conference on Information Fusion*, Cairns, Australia, 2003, 800–806.

[14] Vo, B-N., Singh, S., and Doucet, A.
Sequential Monte Carlo methods for multi-target filtering with random finite sets.
IEEE Transactions on Aerospace and Electronic Systems, **41**, 4 (2005), 1224–1245.

[15] Clark, D. E. and Bell, J.
Multi-target state estimation and track continuity for the particle PHD filter.
IEEE Transactions on Aerospace and Electronic Systems, **43**, 4 (2007), 1441–1453.

[16] Liu, W. F., Han, C. Z., Lian, F., and Zhu, H. Y.
Multi-target state extraction for the probability hypotheses density using Markov chain Monte Carlo.
IEEE Transactions on Aerospace and Electronic Systems, **46** (Apr. 2010), 864–883.

[17] Vo, B-N. and Ma, W-K.
The Gaussian mixture probability hypothesis density filter.
IEEE Transactions on Signal Processing, **54**, 11 (2006), 4091–4104.

[18] Johansen, A., Singh, S., and Vo, B-N.
Convergence of the SMC-PHD filter.
Methodology and Computing in Applied Probability, **8**, 2 (2006), 265–291.

[19] Clark, D. E. and Bell, J.
Convergence results for the particles PHD filter.
IEEE Transactions on Signal Processing, **54**, 7 (2006), 2652–2661.

[20] Clark, D. E. and Vo, B-N.
Convergence analysis of the Gaussian mixture PHD filter.
IEEE Transactions on Signal Processing, **55**, 4 (2007), 1204–1212.

[21] Clark, D. E., Panta, K., and Vo, B-N.
The GM-PHD filter multiple target tracker.
In *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, 2006, 1–8.

[22] Panta, K., Vo, B-N., and Singh, S.
Novel data association schemes for the probability hypothesis density filter.
IEEE Transactions on Aerospace and Electronic Systems, **43**, 2 (2007), 556–570.

[23] Panta, K., Vo, B-N., and Clark, D. E.
An efficient track management scheme for the Gaussian-mixture probability hypothesis density tracker.
In *Proceeding of the Fourth International Conference on Intelligent Sensing and Information Processing*, 2006, 230–235.

[24] Panta, K., Clark, D. E., and Vo, B-N.
Data association and track management for the Gaussian mixture probability hypothesis density filter.
IEEE Transactions on Aerospace and Electronic Systems, **45**, 3 (July 2009), 1003–1016.

[25] Lin, L. Bar-Shalom, Y., and Kirubarajan, T.
Track labelling and PHD filter for multi-target tracking.
IEEE Transactions on Aerospace and Electronic Systems, **42**, 3 (2006), 778–793.

[26] Punithakumar, K., Kirubarajan, T., and Sinha, A.
Multiple-model probability hypothesis density filter for tracking maneuvering targets.
IEEE Transactions on Aerospace and Electronic Systems, **44**, 1 (2008), 87–98.

[27] Vo, B-N., Pasha, A., and Tuan, H. D.
A Gaussian mixture PHD filter for nonlinear jump Markov models.
In *Proceeding of the 45th IEEE Conference on Decision and Control*, San Diego, CA, 2006, 3162–3167.

[28] Erdinc, O., Willett, P., and Bar-Shalom, Y.
A physical-space approach for the probability hypothesis density and cardinalized probability hypothesis density filters.
In O. E. Drummond (Ed.), *Proceedings of Signal and Data Processing of Small Targets*, vol. 6236, Apr. 2006.

[29] Nandakumaran, N., Punithakumar, K., and Kirubarajan, T.
Improved multi-target tracking using probability hypothesis density smoothing.
In *Proceedings of Signal and Data Processing of Small Targets*, vol. 6699, 2007.

[30] Nandakumaran, N., Tharmarasa, R., Lang, T., and Kirubarajan, T.
Gaussian mixture probability hypothesis density smoothing with multistatic sonar.
In *Proceedings of Signal and Data Processing of Small Targets*, vol. 6968, 2008.

[31] Tobias, M. and Lanterman, A. D.
Probability hypothesis density-based multi-target tracking with bistatic range and Doppler observations.
IEE Proceedings—Radar, Sonar and Navigation, **152**, 3 (2005), 195–205.

[32] Sidenbladh, H. and Wirkander, S-L.
Tracking random sets of vehicles in terrain.
In *Proceedings of the IEEE Workshop Multi-Object Tracking*, Madison, WI, 2003.

[33] Maehlisch, M., Schweiger, R., Ritter, W., and Dietmayer, K.
Multisensor vehicle tracking with the probability hypothesis density filter.
In *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, 1–8, 2006.

[34] Clark, D. E., Ruiz, I. T., Petillot, Y., and Bell, J.
Particle PHD filter multiple target tracking in sonar image.
IEEE Transactions on Aerospace and Electronic Systems, **43**, 1 (Jan. 2007), 409–416.

[35] Maggio, E., Taj, M., and Cavallaro, A.
Efficient multitarget visual tracking using random finite sets.
IEEE Transactions on Circuits and Systems for Video Technology, **18**, 8 (2008), 1016–1027.

[36] Wang, Y. D., Wu, J. K., Kassim, A., and Huang, W.
Data-driven probability hypothesis density filter for visual tracking.
IEEE Transactions on Circuits and Systems for Video Technology, **18**, 8 (2008), 1085–1095.

[37] Mullane, J., Vo, B-N., Adams, M. D., and Wijesoma, W. S.
Random set formulation for Bayesian SLAM.
In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems Acropolis Convention Center*, Nice, France, Sept. 22–26, 2008, 1043–1049.

[38] Musicki, D. and Evans, R.
Clutter map information for data association and track initialization.
IEEE Transactions on Aerospace and Electronic Systems, **40**, 2 (2004), 387–398.

[39] Musicki, D., Morelande, M., and Scala, B. L.
Non parametric target tracking in non uniform clutter.
In *Proceeding of the 8th International Conference on Information Fusion*, Philadelphia, PA, 2005, 48–53.

[40] Rozovsky, B. L.
Stochastic Evolution Systems. Linear Theory with Applications to Non-Linear Filtering (Mathematics and its Applications (Soviet Series)) 35.
Norwell, MA: Kluwer Academic Publishing Co., 1990.

[41] Kligys, S., Rozovsky, B. L., and Tartakovsky, A. G.
Detection algorithms and track before detect architecture based on nonlinear filtering for infrared search and track systems.
Center for Applied Mathematical Sciences, University of Southern California, Technical Report CAMS-98.9.1, 1998. Available at http://www.usc.edu/dept/LAS/CAMS/usr/facmemb/tartakov/preprints.html.

[42] McLachlan, G. J. and Peel, D.
Finite Mixture Models.
Hoboken, NJ: Wiley, 2000.

[43] Dempster, A., Laird, N., and Rubin, D.
Maximum likelihood estimation from incomplete data via the EM algorithm.
Journal of the Royal Statistical Society, Series B, **39** (1977), 1–38.

[44] Robert, C. P. and Casella, G.
Monte Carlo Statistical Methods.
New York: Springer, 1999.

[45] Vo, B-T., Vo, B-N., and Cantoni, A.
Analytic implementations of the cardinalized probability hypothesis density filter.
IEEE Transactions on Signal Processing, Pt. 2, **55**, 7 (2007), 3553–3567.

[46] Mahler, R.
PHD filters of higher order in target number.
IEEE Transactions on Aerospace and Electronic Systems, **43**, 3 (2007), 1523–1543.

[47] Bartfai, P. and Tomko, J.
Point Processes and Queuing Problems.
Amsterdam, Holland: North-Holland Publishing Co., 1981.

[48] Schroeter, P., Vesin, J. M., Langenberger, T., and Meuli, R.
Robust parameter estimation of intensity distribution for brain magnetic resonance images.
IEEE Transactions on Medical Imaging, **17**, 2 (1998), 172–186.

[49] Wallance, C. and Dowe, D.
Minimum message length and Kolmogorov complexity.
The Computer Journal, **42**, 4 (1999), 270–283.

[50] Dasgupta, A. and Raftery, A. E.
Detecting features in spatial point processes with clutter via model-based clustering.
Journal of the American Statistical Association, **93** (1999), 294–302.

[51] Akaike, H.
A new look at the statistical model identification.
IEEE Transactions on Automatic Control, **19**, 6 (1974), 716–723.

[52] Seidel, W., Mosler, K., and Alker, M.
Likelihood ratio tests based on subglobal optimization: A power comparison in exponential mixture models.
Statistical Papers, **41**, 85–98.

[53] Figueiredo, M. A. F. and Jain, A. K.
Unsupervised learning of finite mixture models.
IEEE Transactions on Pattern Analysis and Machine Intelligence, **24**, 3 (2002), 381–396.

[54] Green, P. J.
Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.
Biometrika, **4**, 82 (1997), 711–732.

[55] Matthew, S.
Bayesian analysis of mixture models with an unknown number of components—An alternative to reversible jump methods.
The Annals of Statistics, **28**, 1 (2000), 40–74.

[56] Oh, S., Russell, S., and Sastry, S.
Markov chain Monte Carlo data association for multi-target tracking.
IEEE Transactions on Automatic Control, **54**, 3 (2009), 481–497.

[57] Li, X. R. and Jilkov, V. P.
A survey of maneuvering target tracking: Dynamic models.
In *Proceedings of SPIE International Conference on Signal and Data Processing of Small Targets*, 2000, 212–235.

[58] Hoffman, J. and Mahler, R.
Multitarget miss distance via optimal assignment.
IEEE Transactions on Systems, Man, and Cybernetics, Pt. A, **34**, 3 (2004), 327–336.

[59] Ruan, Y. and Willett, P.
The turbo PMHT.
IEEE Transactions on Aerospace and Electronic Systems, **40**, 4 (2004), 1388–1398.

Correcting for Bias in Mahalanobis and Log-Likelihood Estimates

Mahalanobis and log-likelihood estimates are used extensively in tracking systems, but are affected by residual bias and by uncertainty in covariance matrices. This paper provides a formal framework that, to some extent, justifies covariance inflation techniques and allows existing fudge factor thresholds to be interpreted in terms of residual bias covariance matrices and covariance matrix uncertainty.

I. INTRODUCTION

Mahalanobis distance and log-likelihood estimates are frequently used for computing fine-gates and costs in track matching algorithms, validating the consistency of measurement and track covariance matrices, and for discrimination purposes. However, the theoretical basis for the Mahalanobis distance requires unbiased data and perfectly known covariance matrices. In practice, both of these assumptions are frequently violated. Throughout this paper, it is assumed that an attempt has already been made to remove biases in the available data. The remaining (unknown) bias is referred to as residual bias. The residual bias can be interpreted as a random variable with unknown mean and known covariance matrix. The latter is referred to as the residual-bias covariance matrix. Residual bias typically exists. Furthermore, in reality, provided covariance matrices are estimates of covariance matrices rather than the theoretical underlying (deterministic) covariance matrices. (Throughout this paper, noise or uncertainty in the covariance matrices is referred to as covariance jitter.) When the unknown residual bias is described in terms of a residual-bias covariance matrix, it is shown that an unbiased estimate of the underlying true Mahalanobis distance can be achieved by either covariance inflation or modifying the threshold towards which the original (biased) Mahalanobis distance estimate is

Manuscripts received August 14, 2008; revised March 1 and September 11, 2009; released for publication October 15, 2009.

IEEE Log No. T-AES/46/4/938819.

Refereeing of this contribution was handled by F. Gini.

This work was supported in part by the U.S. Department of Defense, The Missile Defense Agency under Contract HQ0006-05-F-0006.

Approved for Public Release 08-MDA-3680 (July 31, 2008).