# Information Measures in Distributed Multitarget Tracking

Murat Uney, Daniel E. Clark
Department of Electrical, Electronic & Computer Eng.
Heriot–Watt University
Riccarton, EH14 4AS, Edinburgh, UK
Email: {m.uney, d.e.clark}@hw.ac.uk

Simon J. Julier
Department of Computer Science
University College London
Malet Place, WC1E 6BT, London, UK
Email: S.Julier@cs.ucl.ac.uk

***Abstract*—In this paper, we consider the role that different information measures play in the problem of decentralised multi–target tracking. In many sensor networks, it is not possible to maintain the full joint probability distribution and so suboptimal algorithms must be used. We use a distributed form of the Probability Hypothesis Density (PHD) filter based on a generalisation of covariance intersection known as exponential mixture densities (EMDs). However, EMD-based fusion must be actively controlled to optimise the relative weights placed on different information sources.**

**We explore the performance consequences of using different information measures to optimise the update. By considering approaches that minimise absolute information (entropy and Rényi entropy) or equalise divergence (Kullback-Leibler Divergence and Rényi Divergence), we show that the divergence measures are both simpler and easier to work with. Furthermore, in our simulation scenario, the performance is very similar with all the information measures considered, suggesting that the simpler measures can be used.**

**Keywords: Multi–sensor multi–target tracking, PHD filtering, exponential mixture densities, generalized covariance intersection, decentralised fusion.**

## I. Introduction

Given their advantages in scalability, reconfigurability and robustness, sensor networks are becoming very important for battlefield situation assessment. These networks are composed of a large number of processing nodes, each with limited computational, communication and sensing capabilities. Observations collected locally are combined and fused with one another and, periodically, fused estimates are distributed to other nodes in the network. Because the networks can be large, time varying, and have communication and computational constraints, fusion must occur in a distributed manner throughout the network. These difficulties are compounded when multiple targets are to be tracked and the number and locations of those targets are unknown.

Broadly, there are two main challenges. First, the tracks in each node must be associated with one another. Track-to-track association problem is considered a special case of the data association problem, and can be addressed within the context of multiple hypothesis testing [1]. An alterative approach that unifies the multi-object tracking problem within a Bayesian paradigm is to use random sets and Finite Set Statistics (FISST) [2]. The second problem is that, once the track-to-track associations have been made, the estimates from the different tracks will be fused together. However, the estimates in different nodes are not independent of one another due to common process noise and previously shared estimates between the nodes [3]. The optimal solution to this problem is to divide out common information. However, this can only be carried out for special classes of probability distributions (Gaussians) in special network topologies (tree or fully connected). As a result, practical disributed fusion can only occur using suboptimal algorithms such as Covariance Intersection (CI) [4].

Mahler considered the problem of suboptimal distributed multi-object tracking. He proposed the generalisation of CI is to use the exponential mixture density (or weighted geometric mean) [5]. Theoretical [6], [7] and practical analyses [8], [9] of this generalised form to single target tracking problems have demonstrated that this generalisation appears sound. However, the first attempt to apply this generalisation was only carried out in [10] and [11], where it was shown that several widely-used multi-object distributions could be readily generalised to support recursive forms that could be implemented using a Monte Carlo realisation. However, a fixed value for the optimisation parameter was used.

In this paper, we continue our investigations into the development of tractable distributed algorithms. In particular, we consider the strategies used to determine the weights assigned to each distribution when they are fused. Furthermore, we experiment with the use of the algorithm directly with feedback.

The structure of this paper is as follows. The problem statement and basic notation is introduced in Section II. The relevance and derivation of different measures is discussed in Section III. The performance of the algorithm is presented and assessed in Section IV, and conclusions are drawn in Section V.

## II. Problem Statement

An environment $\mathcal{E}$ is observed by a set of sensing platforms in the sensor network $\mathcal{S}$. Each platform observes a subregion of $\mathcal{E}$. The physical location and sensing capabilities can differ from platform-to-platform. Furthermore, network links can vary. To maximise flexibility and robustness, we assume a *strict locality assumption:* each node only requires local

knowledge of the identity of the neighbours it communicates with. It is not necessary to store information about the global configuration of the environment.

$\mathcal{E}$ is populated by a set of targets. Both the number and states of the targets are unknown. More formally, the state of the environment can be written as the *set* $X$, where $X = \{x_1, ..., x_n\}$. Both the number of targets (cardinality, $n$) and the state of each target (the $i$ target is $x_i$) are random. Therefore, $X$ is a random finite set and the estimation problem is addressed through the use of Finite Set Statistics (FISST).

### A. Centralised Multi-Target Bayes Tracking

In principle, FISST provides a single, unified Bayesian mechanism to handle all aspects of multi-target tracking including track birth and death, clutter, and data association ambiguity [2]. We assume that the environment can be modelled as an independent identically distributed (i.i.d) cluster process. Therefore, the probability density has the form

$$f(X) = p(n)n! \prod_{x \in X} s(x), \tag{1}$$

where $p(n)$, $n = 0, 1, 2, \ldots$ is the *cardinality distribution* (distribution over the number of targets) and $s(x)$ is the *localisation distribution* (the locations of the targets).

Although this density function has a very convenient form, the computational costs associated with applying Bayes rule to it means that it is only practical for a very small number of targets [2]. On the other hand, recursive forms for the propagation of the *first–order moment* of $f(X)$, $D(X)$, can be computed in linear time from the expression

$$\int_{\mathcal{S}} D(x)\mathrm{d}x = E\{|X \cap \mathcal{S}|\}. \tag{2}$$

In other words, the integral of the PHD over a region $\mathcal{S}$ is the *expected number of targets in* $\mathcal{S}$ [2]. The expectation, which we use later to compute the information measures, is defined using the integral

$$\int f(X)\delta X := f(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int f(\{x_1, ..., x_n\})\mathrm{d}x_1 ... \mathrm{d}x_n. \tag{3}$$

For the i.i.d. cluster process, the PHD has the simple form

$$D(x) = s(x) \cdot \sum_{n=1}^{\infty} n \cdot p(n). \tag{4}$$

This is known as the Cardinalised Probability Hypothesis Density (CPHD) filter, and closed form solutions exist for its prediction and update in a centralised system [2].

### B. Optimal and Suboptimal Distributed Fusion

Consider two nodes, arbitrarily labelled $i$ and $j$, in the sensor network $\mathcal{S}$. The multi-object density function at each node at time step $k$ is $f_i(X|Z_i^{1:k})$ and $f_j(X|Z_j^{1:k})$ respectively, where $Z_m^{1:k}$ is the set of all observation information which has become available to node $m = \{i, j\}$. This includes both the information collected locally and the information propagated from other nodes.

Using Bayes Rule, the multi-object posterior is [3], [5]

$$f(X|Z_i^{1:k} \cup Z_j^{1:k}) = \frac{f(X|Z_i^{1:k}) f(X|Z_j^{1:k})}{f(X|Z_i^{1:k} \cap Z_j^{1:k})}. \tag{5}$$

The optimal update is computed from the ratio of the product of marginal distributions at each node, divided by the distribution conditioned on the information common to both nodes. The difficulty of applying this equation lies in computing $f(X|Z_i^{1:k} \cap Z_j^{1:k})$. If locality of the sensor network is to be maintained, the common information can only be computed for a small number of special cases. Furthermore, it rarely admits a closed form solution [9].

To overcome the first challenge, Mahler proposed to generalise Covariance Intersection (CI) to the multi-modal distributions generated by FISST [5]. Specifically, he proposed that a suitable suboptimal fusion rule is $f(X|Z_i^{1:k} \cup Z_j^{1:k}) \approx f_\omega(X|Z_i^{1:k}, Z_j^{1:k})$, where

$$f_\omega(X|Z_i^{1:k}, Z_j^{1:k}) = \frac{f(X|Z_i^{1:k})^{(1-\omega)} f(X|Z_j^{1:k})^{\omega}}{\int f(X|Z_i^{1:k})^{(1-\omega)} f(X|Z_j^{1:k})^{\omega} \delta X}. \tag{6}$$

This rule replaces the product of the marginal distributions with their weighted geometric mean, or Exponential Mixture Density (EMD) [12]. The EMD rule automatically prevents *double counting* in arbitrary network topologies [6] and can accumulate information [7].

When the states in both filters can be modelled as i.i.d. cluster processes, it can be shown that by applying (4) to (6), $f_\omega(X)$ is an i.i.d. cluster process and its form is [10]

$$f_\omega(X|Z_i^{1:k}, Z_j^{1:k}) = p_\omega(n|Z_i^{1:k}, Z_j^{1:k})n! \prod_{x \in X} s_\omega(x|Z_i^{1:k}, Z_j^{1:k}), \tag{7}$$

where the cardinality and localisation probability functions are

$$p_\omega(n) = \frac{p_i(n)^{(1-\omega)} p_j(n)^{\omega} \left(\int_{\mathcal{X}} s_i(x)^{(1-\omega)} s_j(x)^{\omega} \mathrm{d}x\right)^n}{\sum_{m=0}^{\infty} p_i(m)^{(1-\omega)} p_j(m)^{\omega} \left(\int_{\mathcal{X}} s_i(x)^{(1-\omega)} s_j(x)^{\omega} \mathrm{d}x\right)^m} \tag{8}$$

$$s_\omega(x) = \frac{s_i(x)^{(1-\omega)} s_j(x)^{\omega}}{\int_{\mathcal{X}} s_i(x)^{(1-\omega)} s_j(x)^{\omega} \mathrm{d}x} \tag{9}$$

There are two main challenges to implement these equations. The first is that they rarely admit a closed form solution. For example, if $s_i(x)$ and $s_j(x)$ are Gaussian mixture models, the weighted geometric mean computed in (9) will not, in general, be another Gaussian mixture model. Although Newton series expansion can be used to approximate the terms, these can become numerically unstable unless an extremely high number of components are used [13]. Therefore, a Monte Carlo realisation must be used instead [11]. Second, the EMD update introduces the free parameter $\omega \in [0, 1]$ which governs the relative weight applied to the distributions from $i$ and $j$. An appropriate choice of $\omega$ is important if good performance is to be achieved. We consider the strategies by which $\omega$ can be selected.

## III. Choosing the EMD Weighting Parameter

Unlike regular Bayes rule, the EMD update rule is controlled through the choice of $\omega$. The question of choosing $\omega$ has been considered within the context of the CI algorithm. Several approaches exist. Two broad approaches have been proposed: those based on maximisation or minimisation of cost measures, or those based on equalising divergence metrics.

### A. Maximisation or Minimisation of a Cost Measure

Most implementations of CI have assumed that $\omega$ should be chosen to maximise or minimise some cost metric of the update. If the cost metric can be written as $J(\omega)$ which is some function of the update, then the parameter value is given by

$$\omega^* = \arg \min_{\omega \in [0,1]} J(\omega). \tag{10}$$

Choices for $J(\omega)$ include the determinant of the covariance matrix, the trace of the covariance matrix, Shannon entropy [14], or maximise the "peakiness" (the mode) [5]. However, the only analysis which has been carried out has compared the determinant with the trace, where it is shown that the determinant generally produces better results.

*1) Shannon Differential Entropy:* The Shannon Differential Entropy is perhaps the best-known measure of uncertainty of a distribution. The larger its value, the greater the uncertainty in the state estimate. In this case, $J(\omega) = \mathcal{H}(f_\omega(X))$, where

$$\mathcal{H}(f_\omega(X)) = -\int_{\mathcal{X}} f_\omega(X) \ln \left(f_\omega(X)\right) \delta X. \tag{11}$$

Substituting (7) into (11) and expanding the set integral term as (3), it can be shown that $\mathcal{H}(f_\omega(X))$ is equivalent to [16]

$$\begin{aligned}\mathcal{H}(f_\omega(X)) = &-\mathcal{H}(p_\omega(n)) - \sum_{n=1}^{\infty} \log n! p_\omega(n) \\ &- \sum_{n=1}^{\infty} p_\omega(n) \mathcal{H}(s_\omega(x))^n,\end{aligned} \tag{12}$$

where

$$\mathcal{H}(s_\omega(x)) = -\int_{\mathcal{X}} s_\omega(x) \log s_\omega(x) dx. \tag{13}$$

*2) Rényi Entropy:* The Rényi entropy is a generalisation of Shannon entropy which makes it possible to emphasise certain aspects of a distribution, such as its peak or tails [17]. It is used extensively in active sensor management [18].

The Rényi entropy $\mathcal{H}_\alpha(f_\omega(X))$ is computed from

$$\mathcal{H}_\alpha(f_\omega(X)) = \frac{1}{1-\alpha} \log \left( \int_{\mathcal{X}} f_\omega(X)^\alpha \delta X \right), \tag{14}$$

where $\alpha > 0$ is the free parameter which controls which part of the distribution to emphasise. For $\alpha < 1$, the contribution from the tails (which are low probabilty events) are increased. For $\alpha > 1$, the contribution is assigned to the peaks (high probability events) are increased. In the limit as $\alpha \to 1$, the Rényi entropy converges to the Shannon Differential Entropy.

By substituting from (7) into (14) and applying the set integral (3), we find that

$$\mathcal{H}_\alpha(f_\omega(X)) = \frac{1}{1-\alpha} \log \left( \sum_{n=0}^{\infty} \frac{(p_\omega(n) n!)^\alpha}{n!} \left( \int s_\omega(x)^\alpha \, \mathrm{d}x \right)^n \right). \tag{15}$$

### B. Divergence Equality Metrics

Although minimising absolute information measures provides a baseline, there is an issue with the computational cost: specifically, the distribution can be highly multimodal and exhibit significant variation. An alternative approach is to consider the divergence metrics and the equality between them. Specifically, suppose that there exists a divergence measure $\mathcal{D}(f||g)$ which measures the similiarity between two distributions. We use the squared exponential cost function

$$J_{\mathcal{D}}(\omega) = \exp \left( -0.5 \left\{ \mathcal{D}(f_\omega||f_i) - \mathcal{D}(f_\omega||f_j) \right\}^2 \right). \tag{16}$$

The goal is to find

$$\omega^* = \arg \max_{\omega \in [0,1]} J_{\mathcal{D}}(\omega), \tag{17}$$

which occurs when $J_{\mathcal{D}}(\omega^*) = 1$.

*1) Kullback-Leibler Divergence (KLD):* Hurley [14] proposed to use

$$D(f||g) = \int_{\mathcal{X}} f(X) \log \left( \frac{f(X)}{g(X)} \right) \delta X. \tag{18}$$

Hurley justified this choice of metric on the grounds that this can be related to the *Chernoff Information* which is the upper bound of classification error in a two-class classifier. Although the Chernoff Information has a clear justification for a classification problem, its relevance to distributed fusion is less clear. We propose two alternative arguments for its applicability:

1) Performing an update changes the prior distribution to the posterior distribution. Therefore, it is reasonable to choose $\omega$ to maximise this change, and the KLD provides a means of quantifying this change. However, $D(f_\omega||f_n)$ is a non-decreasing function of $\omega$ [12]. Therefore, simply maximising $D(f_\omega||f_n)$ will cause $\omega$ to always take the value 1. Therefore, a strategy which makes the *change* the same for both distributions takes a suitable "middle ground".
2) From an implementation perspective, the cost metric is computed from the difference between two functions whose values monotonically change. Therefore, there is a single unique maximum and so local optimisation algorithms can be used. However, the global cost functions (e.g., $J(\omega) = \mathcal{H}(f_\omega(X))$) can contain multiple local minima.

Applying (3) to (18), it can be shown that [16]

$$D(f_\omega||f_i) = D(p_\omega(n)||p_i(n)) + \sum_{n=1}^{\infty} p_\omega(n) D(s_\omega(x)||s_i(x))^n \tag{19}$$

where $p_\omega(n)$ and $s_\omega(x)$ are given by (8) and (9) respectively.

The first term in the right hand side of this equation is the KLD of the cardinality distribution and its value is directly computed from

$$D(p_\omega(n)||p_i(n)) = \sum_{n=0}^{\infty} p_\omega(n) \log\left(\frac{p_\omega(n)}{p_i(n)}\right).$$

The KLD of the localisation distributions $D(s_\omega(x)||s_i(x))$ is computed from

$$D(s_\omega(x)||s_i(x)) = \int_{x\in\mathcal{X}} s_\omega(x) \log\left(\frac{s_\omega(x)}{s_i(x)}\right) \mathrm{d}x$$

which, after expanding the logarithm, is

$$\begin{aligned} &= \int_{x\in\mathcal{X}} s_\omega(x) \log\left(\frac{s_j(x)^\omega}{s_i(x)^\omega}\right) \mathrm{d}x \\ &\quad - \int_{x\in\mathcal{X}} s_\omega(x) \log\left(\int_{\mathcal{X}} s_i(x)^{(1-\omega)} s_j(x)^\omega \,\mathrm{d}x\right) \mathrm{d}x \\ &= \omega \int_{\mathcal{X}} s_\omega(x) \left(\log s_j(x) - \log s_i(x)\right) \mathrm{d}x \\ &\quad - \log\left(\int_{\mathcal{X}} s_i(x)^{(1-\omega)} s_j(x)^\omega \,\mathrm{d}x\right). \end{aligned}$$

The divergence $D(f_\omega||f_j)$ has a similar form, but with $p_j(n)$ subsituted for $p_i(n)$, $s_j(x)$ substituted for $s_i(x)$, and $(1-\omega)$ substituted for $\omega$.

*2) Rényi Divergence:* The Rényi Divergence generalises the KLD through the use of a free parameter $\alpha$ which can be used to emphasise particular aspects of the differences between the distributions which are of interest. Therefore, $\mathcal{D}(f||g) = R_\alpha(f||g)$, where

$$R_\alpha(f(X), g(X)) \triangleq \frac{1}{\alpha-1} \log\left(\int_{\mathcal{X}} f(X)^\alpha g(X)^{(1-\alpha)} \,\delta X\right). \tag{20}$$

For an i.i.d. cluster process, Ristić *et al.* proved that [18]

$$\begin{aligned} R_\alpha(f_\omega(X)||f_i(X)) &= \frac{1}{\alpha-1} \\ \log\left(\sum_{n=0}^{\infty} p_\omega(n)^\alpha p_i(n)^{(1-\alpha)} \left[\int_{\mathcal{X}} \left(s_\omega(x)\right)^\alpha s_i(x)^{1-\alpha} \,\mathrm{d}x\right]^n\right). \end{aligned} \tag{21}$$

In the next section, we evaluate the information measures in some simulated scenarios.

## IV. Simulation Analysis

### A. Scenario

The scenario is illustrated in Figure 1(a) and consists of four targets with constant velocity motion and two sensor platforms ($i=0$, $j=1$) making range–bearing measurements.

The velocity vectors of the targets are given by $[-176.7,\ 0.0]^T$, $[-220.9,\ 88.0]^T$, $[176.7,\ -176.7]^T$ and $[176.7,\ 0.0]^T$ for targets $1, 2, 3$ and $4$ respectively. The motion model does not incorporate any random perturbations in the target trajectories.

The sensors can detect targets up to a maximum range of 7500m. The measurements are noise corrupted by Gaussian-zero mean noise sources. The standard deviation of the range is 3m and bearing is $2^\circ$. In each scan, the number of clutter returns are Poisson distributed with an intensity $\lambda = 5$. Clutter is uniformly distributed within the detection region. The union of the time history of measurements is shown in Figure 1(b).

### B. Implementation

The sensor platforms employ Sequential Monte Carlo (SMC) CPHD filters as described in [19]. The PHD is represented by a set of equally–weighted particles generated from the posterior localisation distribution $s_i(x|Z_i)$. The multi–object state estimate from the set of particles is obtained through a clustering scheme based on the observations [19].

At every time step, each sensor receives the posterior PHD and cardinality distribution of its neighbour for fusion. To mimic the effects of a finite bandwidth, the transmitted particle distribution is approximated by a Gaussian Mixture Models (GMM) [20]. In our implementation, we use the clustering scheme from [19] to partition the PHD particles. Each cluster is replaced by a mean and covariance matching Gaussian component with the same relative weight.

The fused posterior is regarded as an approximation to the posterior conditioned on the observations from both of the sensors and is the best description of the multi–object scene the fusion platform maintains. Since the EMD rule prevents double–counting of any recurring information, it can be fed back to the filter before the prediction step.

After receiving GMM representation of a PHD and a cardinality distribution from the neighbour, the fusion platform produces samples from this PHD and performs the relevant Monte Carlo computations.

*1) Monte Carlo methods for estimating the information measures:* Consider equally weighted sets of $N_i$ and $N_j$ particles $P_i$ and $P_j$ generated from $s_i(x)$ and $s_j(x)$ respectively. Note that the union of $P_i$ and $P_j$, i.e.,

$$P_U \triangleq P_i \cup P_j \tag{22}$$

is a particle set generated by the mixture density given by

$$s_U(x) = \frac{N_i s_i(x) + N_j s_j(x)}{N_i + N_j} \tag{23}$$

Therefore, utilisation of $P_U$ for Importance Sampling (IS) estimates corresponds to selecting the proposal distribution as $s_U(x)$ given by (23).

The information measures presented in Section III requires evaluation of integrals involving $s_i(x)$ and $s_j(x)$. We employ Importance Sampling methods using $s_U(x)$ given by (23) as well as conventional Monte Carlo integration (Chp. 3 of [21]) for estimating these values. In order to evaluate the distributions $s_i(x)$ and $s_j(x)$ at $x \in P_i$ or $x \in P_j$ during these computations, we employ observation–based Kernel Density Estimates [11].

*2) Choosing the Value of $\omega$:* Since the aim of this paper is to assess the impact of different choices of cost function rather than develop highly efficient schemes for choosing $\omega^*$, we use a simple brute force search strategy. We evaluate the MC approximations over a fixed set of values for $\omega \in [0, 1]$. The nature of the information measures means that these can be computed in terms of $p_i(n)$, $s_i(x)$, $p_j(n)$ and $s_j(x)$. Therefore it is not necessary to explicitly compute $s_\omega(x)$ given in (9).

*3) Sampling from the EMD:* After $\omega^*$ is found, the cardinality distribution $p_{\omega^*}(n)$ is constructed using (8). We approximate $s_{\omega^*}(x)$ by using Importance Sampling and resampling methods to approximate $P_{\omega^*} = \{x | x \sim s_{\omega^*}(x)\}$. Using

Importance Sampling and resampling methods based on the particle set $P_U$ given by (22) and the corresponding proposal distribution $s_U(x)$ given by (23). Note that the importance sampling weights of particles $x^{(k)}$ generated from a proposal distribution $q(x)$ in order to represent a distribution $p(x)$ are given by $\zeta^{(k)} = p(x^{(k)})/q(x^{(k)})$. We substitute $s_\omega(x)$ given by (9) and $s_U(x)$ given by (23) in place of $p(x)$ and $q(x)$ respectively, and obtain the importance sampling weights for particles $x_U^{(k)} \in P_U$ to represent $s_\omega(x)$ as

$$\zeta_U^{(k)} \propto \frac{s_i\left(x_U^{(k)}\right)^{(1-\omega)} s_j\left(x_U^{(k)}\right)^{\omega}}{N_i s_i\left(x_U^{(k)}\right) + N_j s_j\left(x_U^{(k)}\right)} \tag{24}$$

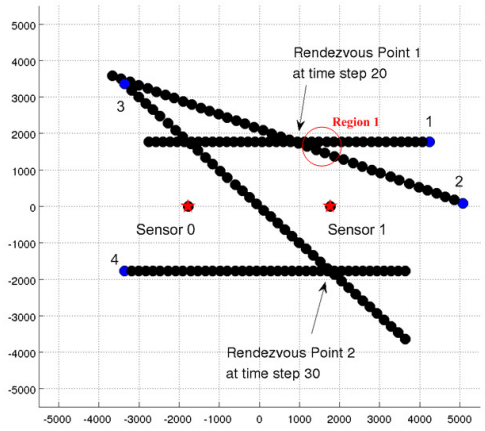
After resampling $P_U$ with $\{\zeta_U^{(k)}\}$, samples generated approximately from $s_\omega(x)$ are obtained.

### C. Results

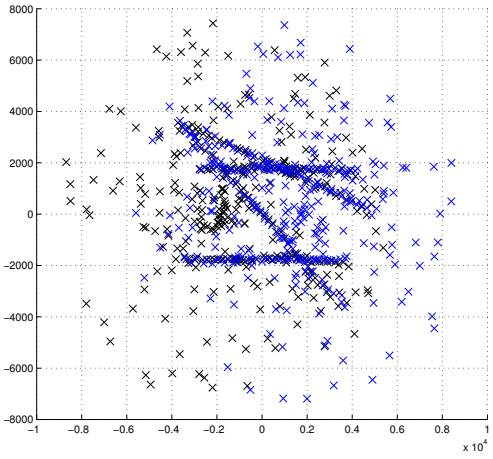
*1) Choosing the EMD Weighting Parameter:* In order to select the best $\omega$ that specifies the EMD of the fusion rule (the EMD weighting parameter), we use the minimum Rényi Entropy and Rényi divergence equality criteria ( Section III). Let us consider the Rényi divergence $R_\alpha$ for $\alpha = 0.25, 0.5, ..., 1.75$. For each $\alpha$, the average $\omega^*$ specifying the EMD at platform 0 is given in Figure 2(a). It is apparent that if $\alpha_1 < \alpha_2 < \alpha_3$ holds, the corresponding best $\omega$ satisfies $\min(\omega_1^*, \omega_3^*) < \omega_2^* < \max(\omega_1^*, \omega_3^*)$. As $\alpha$ increases, the variation in $\omega^*$ decreases and the average shows a steady behaviour. The average $\omega^*$ at platform 1 for the case in which sensor 0 transmits its results to this platform is given in Figure 2(b). The complementary behaviour of $\omega^*$ in the sense that $\omega^*$s at a particular time step in Figure 2(a) and Figure 2(b) respectively adds to 1 demonstrates the consistency of the underlying computations.

The average $\omega^*$ for the Rényi Entropy minimisation criterion is given in Figure 3(a) and (b). First, we observe that this criterion exhibits a high deviation in comparison to the divergence equality results. Second, the $\omega^*$ values do not show the complementary behaviour and do not sum to one. This suggests that the nodes do not operate in a symmetric manner with respect to one another, suggesting potential issues with numerical robustness.

To illustrate the behaviour of the divergence equality criterion, Figures 4(a) and (b), shows the Rényi divergences from $f_\omega$ to $f_0$ $(R_\alpha(f_\omega(X)||f_0(X)))$ and $f_1$ $(R_\alpha(f_\omega(X)||f_1(X)))$




(a) Example scenario.

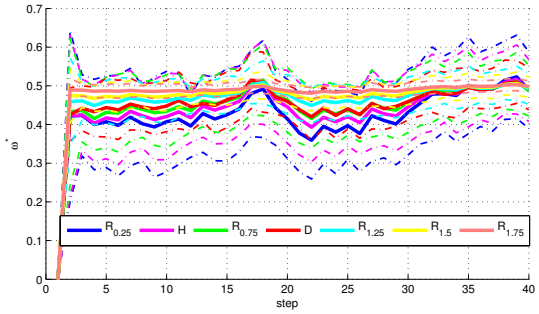


(b) Observations collected in a typical run.

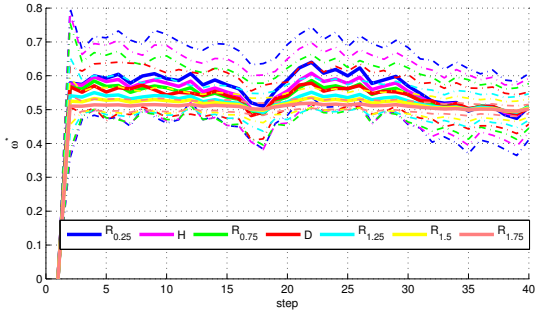
Figure 1. (a) Example scenario: Two range–bearing sensors (red stars) observe four targets (black dots) for 40 time steps. The blue dots indicate the initial target positions. At time step 20, the tracks of targets 1 and 2 cross at rendezvous point 1, and, targets 3 and 4 meet at rendezvous point 2 at time step 30. (b) Example observations of sensors 0 and 1 in Cartesian coordinates superpositioned for 40 steps (indicated by black and blue crosses, respectively).

for different values of $\alpha$. Note the non–decreasing behaviour of $R_\alpha(f_\omega(X)||f_0(X))$ with respect to $\omega$ (and, similarly, the non–decreasing behaviour of $R_\alpha(f_\omega(X)||f_1(X))$ with $1-\omega$) as discussed in Section III-B. As $\alpha$ is increased, the divergence exhibits a higher rate of increase. For a given value of $\alpha$, the EMD parameter $\omega^*$ is the intersection of the corresponding curves in Figure 4(a) and (b) and it is unique.

Figures 5(a)–(c), plot typical performance functions constructed for the Rényi divergence equality criterion for $\alpha = 0.5$ (Hellinger distance), $\alpha = 1.0$ (KLD) and $\alpha = 1.5$ respectively, in a way similar to (16) at both platforms 0 and 1. The complementary nature of the EMD parameters for platforms 0 and 1 manifest itself with the corresponding extrema located at values adding to 1. In Figure 5(d), we present the normalised (Shannon) Entropy of the EMD versus $\omega$ which presents the lack of this complementary nature with the minimum Entropy criterion. In addition, note that there is a local extremum approximately at $\omega = 0.8$ for fusion at platform 0

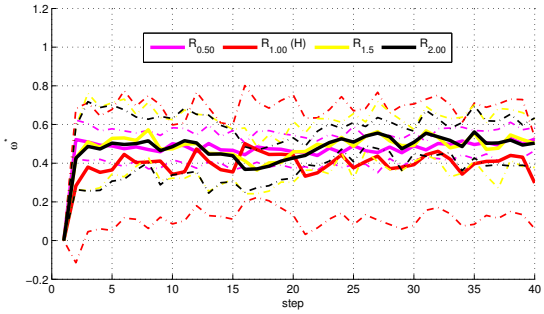



(a) Average $\omega^*$ for fusion at platform 0

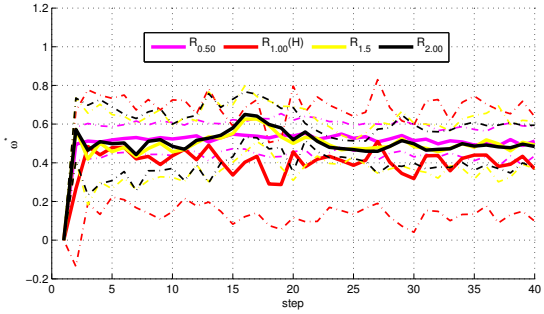



(b) Average $\omega^*$ for fusion at platform 1.

Figure 2. (a) The average $\omega^*$ specifying the EMD obtained using the Rényi divergence equality criterion for $\alpha = 0.25, 0.5, 0.75, 1.00, 1.25, 1.50$ and $1.75$ (denoted by $R_{0.25}$, $H$, $R_{0.75}$, $D$, $R_{1.25}$, $R_{1.50}$ and $R_{1.75}$ respectively noting that $R_{0.5}$ is the Hellinger distance and $R_{1.00}$ is the KLD). The solid lines indicate the average whereas the dash–dotted lines are the $\pm 1$ standard deviation boundaries. (b) Similar results for fusion at sensor platform 1. The results are averaged over 30 runs.




(a) Average $\omega^*$ for fusion at platform 0




(b) Average $\omega^*$ for fusion at platform 1.

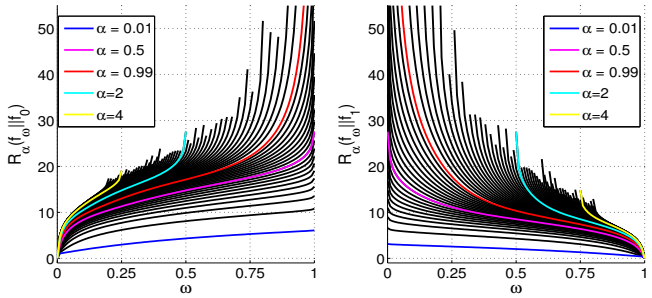
Figure 3. (a) The average $\omega^*$ specifying the EMD obtained using the minimum Rényi Entropy criterion for $\alpha = 0.5, 1.0, 1.5$ and $2.0$ (note that $R_{1.0}$ is the Shannon Entropy denoted here by $H$). The solid lines indicate the average whereas the dash–dotted lines are the $\pm 1$ standard deviation boundaries. (b) Similar results for fusion at sensor platform 1. The results are averaged over 30 runs.

demonstrating non–unimodality of this criterion as mentioned in Section III-B1 for motivating the use of divergence equality.

*2) Performance comparison:* The performance of the individual filter outputs and the fusion results are compared using the Optimal Subpattern Assignment (OSPA) distance (over the space of finite sets) [22] between the multi–object state estimates obtained with an observation based clustering scheme from the PHDs [19], and the true multi–target state. The OSPA error is composed of a penalty for the cardinality difference, i.e., the difference in the number of elements of the two sets, and a localisation distance.

The improvement in the target localisation is assessed using the OSPA localisation distance. In Figure 6(a), we present the localisation error for the CPHD filter of sensor 0 together with EMD fusion performed by this platform using divergence equality for $\alpha = 0.5, 1.0, 1.5$. The fused cluster distribution $f_\omega$ exhibits a significant improvement in the localisation performance which is consistent for different values of $\alpha$. Similar results for fusion at platform 1 (using the posterior transmitted from platform 0) is presented in Figure 6(b). Let us compare Figure 6(a) and Figure 6(b) for time steps between 14 and 23 corresponding to targets 1 and 2 first approaching and then diverging (over Region 1 in Figure 1(a)). Sensor 1 resolves the target locations better than sensor 0. One reason for the relatively lower localisation performance with high variance at sensor 0 is the effect of multi–object state estimation through clustering based on the observations, for which sensor 1 is more advantegous.

In Figure 7(a) and (b), we present the average localisation error for the minimum Rényi Entropy criterion. Note that the performance is similar to that of the divergence equality criterion. Comparing Figure 6(b) and Figure 7(b) around time steps 20 and 30 (when target crossings occur), the entropy minimisation yields a higher variation in performance.




(a) (b)

Figure 4. Rényi divergence curves (computed at time step 28 by platform 0): The divergence estimates (a) from $f_\omega(X)$ to $f_0(X)$, and (b) from $f_\omega(X)$ to $f_1(X)$ for $\alpha = 0.05, 0.1, ..., 0.95, 1.05, 1.1, ..., 5$ (black curves). Note that for $\alpha = 1$, the Rényi divergence is singular with a limit approaching to the Kullback–Leibler Divergence. The $\alpha$ values to which the coloured curves correspond can be seen in the legends.

Table I
TIME AVERAGE OF EMPIRICAL OSPA CARDINALITY ERROR STANDARD DEVIATION.

| | |
|---|---|
| Regular CPHD Filter of platform 0 | 5.39187 |
| Filter w/ feedback from EMD using H | 3.96837 |
| Filter w/ feedback from EMD using KLD | 3.96837 |
| Fusion of the feedback filter and sensor 1 using H | 0.858884 |
| Fusion of the feedback filter and sensor 1 using KLD | 1.44549 |
| Fusion of the usual CPHD filter and sensor 1 using H | 5.52667 |
| Fusion of the usual CPHD filter and sensor 1 using KLD | 5.52186 |

We also consider feeding $f_\omega$ back to the filter at platform 0 (Section IV-B). We run separate CPHD filters at platform 0 each receiving feedback from $f_\omega$ obtained by using the divergence equality criterion and one of $\alpha = 0.5, 1.0$. In Figure 6(c), the OSPA localisation errors of filters receiving feedback are given in comparison with the regular CPHD filter and the fusion results. The filters with feedback exhibit an improved localisation performance as expected and the fusion results are on a par with those in Figure 6(a).

The benefits of the feedback scheme is better understood comparing the empirical standard deviation of the OSPA cardinality error averaged over time as an indicator of immunity to the effects of clutter. In Table I, we present this quantity for a regular CPHD filter at platform 0 and filters receiving feedback from the fusion using different divergences together

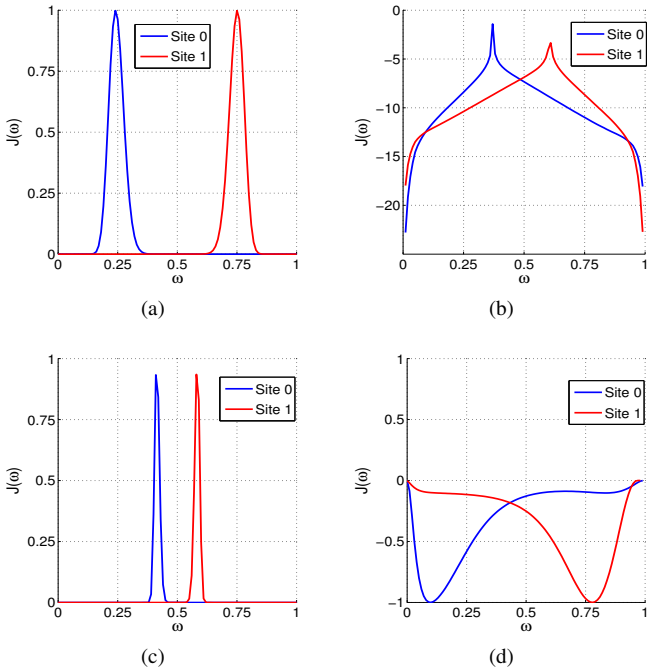
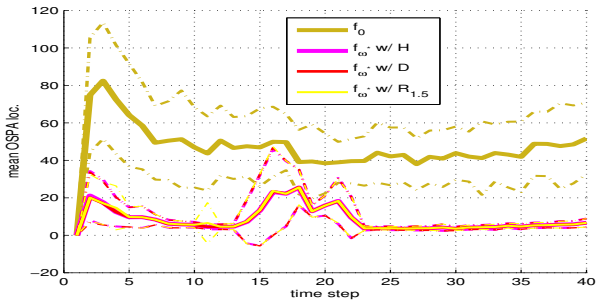
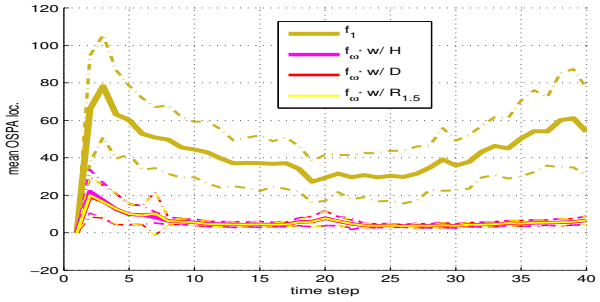


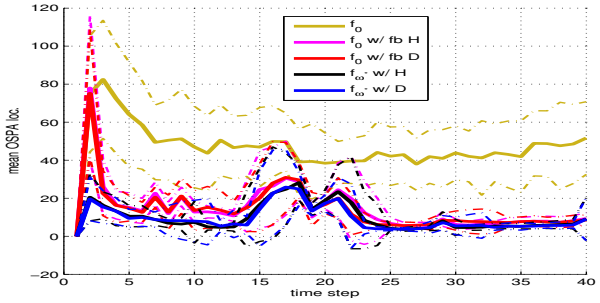

Figure 5. Performance functions $J(\omega)$ based on information measures vs. $\omega$ for a typical step:
(a) $J(\omega) = \exp(-0.5(R_{0.5}(f_\omega(X), f_0(X)) - R_{0.5}(f_\omega(X), f_1(X)))^2)$,
(b) $J(\omega) = -\log(|D(f_\omega(X)||f_0(X)) - D(f_\omega(X)||f_1(X))| + 1)$,
(c) $J(\omega) = \exp(-0.5(R_{1.5}(f_\omega(X), f_0(X)) - R_{1.5}(f_\omega(X), f_1(X)))^2)$,
(d) $J(\omega) = \mathcal{H}(f_\omega(X))/\max_{\omega\in[0,1]}|H(f_\omega(X))|$, computed at platform 0 (blue curves) and platform 1 (red curves).




(a) Fusion at platform 0.




(b) Fusion at platform 1.



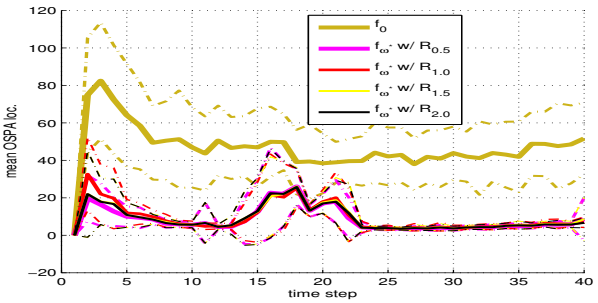

(c) Fusion at platform 0 with feedback.

Figure 6. Average OSPA localisation error between the true multi–target states and the multi–object state estimates obtained from (a) CPHD filter 0 (black solid line), and EMD Fusion performed at platform 0 using the Rényi divergence with $\alpha = 0.5$(H),1(D) and 1.5 ( the dash–dotted lines stand for $\pm 1$ standard deviation and the average is over 30 runs), (b) CPHD filter 1 and Fusion performed at platform 1 using divergences similar to those in (a), (c) CPHD filter 0 with feedback from the Fusion (using the Rényi divergence with $\alpha = 0.5$(H) and 1(D) represented by the magenta and red solid lines respectively) and the EMD result at platform 0 (dashed blue and green lines).

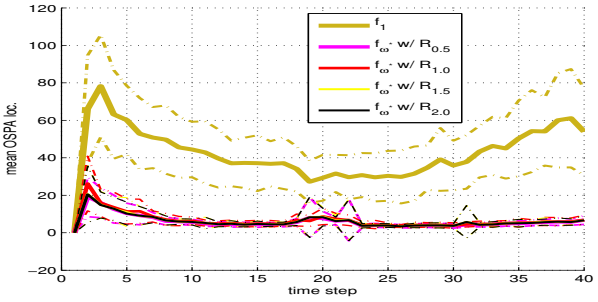
with the fusion results. The persistence of the cardinality output by filter 0 is improved by the feedback scheme and exhibits a potential of better immunity to the effects of clutter. In addition, when the feedback filters are fused, the resulting cluster process exhibit lower cardinality error variance.

## V. CONCLUSION

In this paper, we have considered the problem of choosing the cost function which underlies the optimisation problem in suboptimal distributed fusion. We have presented a series of equations to derive different forms, and we have compared the results on an example. The results suggest that, as predicted by the theory, divergence measures are better-behaved from an optimisation perspective. Furthermore, the overall performance using equality of the divergence measures criteria are very




(a) Fusion at platform 0.




(b) Fusion at platform 1.

Figure 7. Average OSPA localisation error between the true multi–target states and the multi–object state estimates obtained from (a) the CPHD filter 0 (black solid line), and the EMD Fusion at platform 0 using the minimum Rényi Entropy criterion with $\alpha = 0.5$, 1.0(Shannon Entropy) and 1.5 ( the dash–dotted lines stand for $\pm 1$ standard deviation and the average is over 30 runs), (b) CPHD filter 1 and Fusion performed at platform 1 using divergences similar to those in (a).

similar to those when the entropy reduction criterion is used. The results also suggest that picking a fixed value of $\omega = 0.5$ leads to good performance and, in this case, suggests that optimisation is not required.

One possible explanation for these results is that there is a high degree of heterogeneity in the state estimates contained at each node. Future work will explore this further by examining more variations in the types and kinds of sensors used, varying the degree of overlap region, and changing the frequency with which nodes communicate with one another.

## ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) (Grant number EP/H011990/1) and the MOD University Defence Research Centre on Signal Processing (UDRC). Dr Clark is a Royal Academy of Engineering/ EPSRC Research Fellow.

## REFERENCES

[1] M. E. Liggins II, C.-Y. Chong, I. Kadar, M. G. Alford, V. Vannicola, and S. Thomopoulos, "Distributed fusion architectures and algorithms for target tracking," *Proc. of the IEEE*, vol. 85, no. 2, pp. 95–107, 1997.

[2] R. P. S. Mahler, *Statistical Multisource Multitarget Information Fusion*. Springer, 2007.

[3] C.-Y. Chong, S.Mori, and K.-C. Chang, "Distributed multitarget multisensor tracking," in *Multitarget-Multisensor Tracking: Advanced Applications*, Y. Bar-Shalom, Ed. Artech House, 1990, ch. 8.

[4] J. Uhlmann, "Dynamic map building and localisation for autonomous vehicles," Ph.D. dissertation, University of Oxford, 1995.

[5] R. Mahler, "Optimal/robust distributed data fusion: a unified approach," in *Proc. of the SPIE Defense and Security Symposium 2000*. SPIE Defense and Security Symposium, 2000.

[6] S. Julier, "Fusion without independence (keynote abstract)," in *Proc. of the IET Seminar on Tracking and Data Fusion: Algorithms and Applications*, 15–16 April 2008, pp. 1–5.

[7] S. J. Julier, T. Bailey, and J. K. Uhlmann, "Using Exponential Mixture Models for Suboptimal Distributed Data Fusion," in *Proc. of the 2006 IEEE Nonlinear Stat. Signal Proc. Workshop (NSSPW'06)*. NSSPW'06, September 2006, pp. 160–163.

[8] S. J. Julier, "An empirical study into the use of chernoff information for robust, distributed fusion of gaussian mixture models," in *Proc. of the IEEE ICIF*, 2006.

[9] L.-L. Ong, T. Bailey, H. Durrant-Whyte, and B. Upcroft, "Decentralised Particle Filtering for Multiple Target Tracking in Wireless Sensor Networks," in *Proc. of the 11th International Conference on Information Fusion*, Cologne, Germany, July 2008, pp. 1–8.

[10] D. Clark, S. Julier, R. Mahler, and B. Ristić, "Robust Multi-object sensor Fusion with Unknown Correlations," in *Proc. of the 2010 Sensor Signal Processing for Defence 2010 Conference (SSPD 2010)*. SSPD 2010, September 2010.

[11] M. Uney, S. Julier, D. Clark, and B. Ristić, "Monte Carlo Realisation of a Distributed Multi–Object Fusion Algorithm," in *Proc. of the 2010 Sensor Signal Processing for Defence 2010 Conference (SSPD 2010)*. SSPD 2010, September 2010.

[12] A. Dabak, "A Geometry for Detection Theory," Ph.D. dissertation, Rice University, Houston, TX, USA, 1992.

[13] N. Mariam, "Conservative non-gaussian data fusion for decentralized networks," Master's thesis, University of Sydney, 2007.

[14] M. B. Hurley, "An Information-Theoretic Justification for Covariance Intersection and Its Generalization," in *Proc. of the 2002 FUSION Conference*, vol. 1. FUSION 2002, July 2002, pp. 505–511.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second, Ed. John Wiley and Sons, 2006.

[16] D. Daley and D. V. Jones, *An Introduction to the Theory of Point Processes Volume II: General Theory and Structure*. Springer, 2008, ch. 14.

[17] A. Rényi, "On Measures of Entropy and Information," in *Proc. of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, vol. 1, 1960, pp. 547–561.

[18] B. Ristić, B.-N. Vo, and D. Clark, "A Note on the Reward Function for PHD Filters with Sensor Control," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 2, pp. 72–80, April 2011.

[19] B. Ristić, D. E. Clark, and B. N. Vo, "Improved smc implementation of the phd filter," in *Proc. of the ICIF 2010 (FUSION2010)*, July 2010.

[20] A. Ihler, "Inference in Sensor Networks: Graphical Models and Particle Methods," Ph.D. dissertation, MIT, 2005.

[21] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. Springer, 2004.

[22] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Process*, vol. 56, no. 8, pp. 3447–3457, August 2008.