INFORMATION FUSION USING PARTICLES INTERSECTION

Or Tslil and Avishy Carmi

Department of Mechanical Engineering
Ben-Gurion University of the Negev, Beer Sheva, Israel

ABSTRACT

A technique is presented for combining arbitrary empirical probability density estimates whose interdependencies are unspecified. The underlying estimates may be, for example, the particle approximations of a pair of particle filters. In this respect, our approach provides a way to obtain a new particle approximation, which is better in a precise information-theoretic sense than that of any of the particle filters alone. The viability of the proposed approach is demonstrated in a multiple object tracking scenario.

1. INTRODUCTION

One of the challenges in multiagent/networked systems is to integrate data from several intelligent platforms for enhancing and coordinating their actions. Networked systems are expected to be resilient to varying environmental parameters and uncertainties. This will usually be manifested in a design with the capacity to alter the underlying network topology in response to external variations and possible malfunctions [1]. The network topology is the communication infrastructure by which the entities "talk" with one another and hence governs the emergence of intelligent behaviour of the network as a whole [2, 3].

Adaptive network systems require equally flexible information fusion paradigms [4, 5, 6, 7, 8, 9]. In a large-scale sensor network the nodes' energy consumption and the capacity of the network as a whole to adapt to variations in its environment are particularly important [4]. Such a network may consist of a large number of sensors each having a modest information processing capability. Typically, a sensor will run its own estimation algorithm such as a Kalman filter [6, 9] or a particle filter (PF) [10]. This approach leads to decentralized or distributed information fusion.

Distributed estimation schemes rely on the dissemination of measurements and other statistics across the network. The distributed PFs schemes in [10] employ approximate likelihoods to alleviate the communication overhead when exchanging raw data between nodes. In some consensus-based approaches, global estimates are gradually approached in nodes through an iterative procedure in which they exchange their local estimates [6, 11]. Distributed PF schemes may communicate particles and weights. Alternatively, they may communicate the parameters of a Gaussian mixture model (GMM) which approximates the underlying filtering density. The first approach may be inefficient when the number of particles in each PF is large. The latter approach similarly becomes impractical for high-dimensional states and many Gaussian components. [12].

In networks with hundreds and thousands of sensors, measurements and other statistics from different nodes may occasionally be integrated in some principle nodes where their interdependencies are ignored [9]. This sometimes leads to statistically inconsistent estimators which are prone to diverge. The reason for this is known as double counting [13]; unspecified interdependencies may allow an old piece of information (one which has already been processed) to be reprocessed when it recurs in the data collected from another sensor.

A simple approach that addresses this problem is known as covariance intersection (CI) [9, 14]. CI allows fusing a pair of statistically consistent estimates of the same quantity such that the resulting estimate is similarly consistent. This technique, however, accounts only for the estimators' first and second statistical moments (mean and covariance) and may therefore be inadequate in the case of arbitrary non-Gaussian probability density functions (pdf). For example, CI cannot be used straightforwardly in such scenarios where several sensors track a number of objects or even a single object whose dynamical and observation models vary randomly in time [15]. Moreover, in PF-based multiple object tracking algorithms the system's state is represented by a random finite set rather than a random vector, and so an underlying covariance is not well-defined [16].

In the past years, several extensions of CI have been devised for handling non-Gaussian pdfs. An approximate GMM fusion scheme based on *Chernoff information* is described in [17]. As mentioned there and in [18], this measure of discrepancy between two probability densities has much to do with CI, and in fact recovers its update equations in the Gaussian case. A similar approach is applied for multiple object tracking using multi-Bernoulli filters in [19].

Here, we devise a technique, *particles intersection*, for combining arbitrary empirical probability density estimates

This research is supported by Israel Science Foundation Grant No. 1723/16

whose interdependencies are unspecified. The underlying density estimates may be, for example, the particle approximations of a pair of PFs. In this respect, particles intersection provides a way to obtain a new particle approximation, which is better in a precise information-theoretic sense than that of any of the PFs alone.

This paper is organized as follows. The next section presents a general approach for fusing arbitrary probability density estimators. Particles intersection, the application of this fusion technique to the particle approximations of a pair of PFs, is presented in Section 3. The utility of particles intersection is demonstrated in Section 4 where a pair of PFs are employed for tracking of a group of objects. Some concluding remarks are offered in the last section.

2. FUSION OF PROBABILITY DENSITY ESTIMATORS

Let $f(x)$ and $g(x)$ be two estimators of the probability density function (pdf) $h(x)$ of some real-valued random vector X whose realization is x. We wish to obtain another estimator, $q(x)$, which accounts for the information in both these estimators. Here, we construct this $q(x)$ as

$$q(x) = \frac{f(x) \triangleright_\alpha g(x)}{\int f(y) \triangleright_\alpha g(y) dy} \tag{1}$$

where the operation, $f \triangleright_\alpha g \stackrel{\text{def}}{=} f^\alpha g^{1-\alpha}$, and $\alpha \in [0,1]$. Some properties of (1) have been pointed out in several works. For example, it is known that for Gaussian $f(x)$ and $g(x)$, the obtained $q(x)$ is also a Gaussian whose mean and covariance are given by the CI method [20, 17, 18]. CI has been shown to maintain statistical consistency in the sense that if the covariances P_f and P_g with respect to the pdfs $f(x)$ and $g(x)$ are such that $P_f - P_h$ and $P_g - P_h$ are positive semidefinite then $P_q - P_h$ is similarly positive semidefinite. In [18] the operation $\triangleright_\alpha$ was shown to generate an algebraic structure which facilitates the application of (1) for large networks with many different pdf estimators. The geometric interpretation of (1) as a point along a geodesic in a statistical manifold was discussed in [20].

Nevertheless, other statistical properties of (1) or why such an approach should be used in the first place remained an open issue. The next results establish the significance of (1) from an information-theoretic perspective.

The three density estimators satisfy,

$$\mathrm{KL}\,(q \parallel h) = \alpha \mathrm{KL}\,(f \parallel h) + (1-\alpha)\mathrm{KL}\,(g \parallel h) - I_\alpha(f,g) \tag{2}$$

where $\mathrm{KL}\,(q \parallel h) \stackrel{\text{def}}{=} \int q(x) \log(q(x)/h(x)) dx$ is the Kullback-Leibler (KL) divergence of $q(x)$ and $h(x)$, and

$$I_\alpha(f,g) \stackrel{\text{def}}{=} \log \int f(x) \triangleright_\alpha g(x) dx \geq 0 \tag{3}$$

is known as the *Chernoff information* of $f(x)$ and $g(x)$. Much like the KL-divergence, $I_\alpha(f,g)$ quantifies the discrepancy between $f(x)$ and $g(x)$. The larger this measure is, the closer $q(x)$ gets to $h(x)$ for $\mathrm{KL}\,(q \parallel h)$ becomes smaller. Moreover, the next results shows that there is always such α for which $q(x)$ is not farther from $h(x)$ than any of the estimators alone.

Theorem 1 (Consistency of KL) *There exists an $\alpha \in [0,1]$ for which*

$$\mathrm{KL}\,(q \parallel h) \leq \min\left\{\mathrm{KL}\,(f \parallel h),\, \mathrm{KL}\,(g \parallel h)\right\} \tag{4}$$

Proof. From (2) it follows that whenever

$$I_\alpha(f,g) \geq (1-\alpha)\left[\mathrm{KL}\,(g \parallel h) - \mathrm{KL}\,(f \parallel h)\right] \tag{5}$$

the KL-divergence, $\mathrm{KL}\,(q \parallel h) \leq \mathrm{KL}\,(f \parallel h)$. Similarly, if

$$I_\alpha(f,g) \geq \alpha\left[\mathrm{KL}\,(f \parallel h) - \mathrm{KL}\,(g \parallel h)\right] \tag{6}$$

then $\mathrm{KL}\,(q \parallel h) \leq \mathrm{KL}\,(g \parallel h)$. Because one of the right-hand sides, either in (5) or (6), is nonpositive, one of these inequalities always holds. The remaining inequality can always be satisfied for some $\alpha \in [0,1]$. QED.

From (2) it is apparent that maximizing $I_\alpha(f,g)$ with respect to α may in some cases lead to smaller $\mathrm{KL}\,(q \parallel h)$. The Chernoff information, as distinct from other terms in (2), is independent of the unknown $h(x)$, and so such an optimization can be carried out in practice,

$$\alpha_{opt} = \arg\max_\alpha I_\alpha(f,g) \tag{7}$$

3. PARTICLES INTERSECTION

We name particles intersection the application of (1) using the empirical pdf representations obtained by a pair of PFs. A PF is a numerical approximation technique for solving the Bayesian filtering problem. Here, one is commonly interested in obtaining $p_{X_k|Z_{0:k}}(x_k \mid z_{0:k})$, the conditional pdf of the system's state at time k, represented by the random vector X_k, given all observations up to that time, the set of random vectors, $Z_{0:k} \stackrel{\text{def}}{=} \{Z_1, \ldots, Z_k\}$. The output of a conventional PF is a set of N samples (particles), $\{x_k^{(i)}\}_{i=1}^N$, and corresponding weights, $\{w_k^{(i)}\}_{i=1}^N$, $\sum_i w_k^{(i)} = 1$, which together provide an empirical estimate of the underlying filtering pdf,

$$\pi_{X_k|Z_{0:k}}(x_k \mid z_{0:k}) = \sum_{i=1}^N w_k^{(i)} \delta(x_k - x_k^{(i)}) \tag{8}$$

where $\delta(\cdot)$ is the Dirac delta function. The minimum mean square error (MMSE) estimator, for example, can be approximated this way as the weighted sum of all particles. For brevity, random variables subscript as in (8) are omitted for the rest of this paper.

Consider a pair of PFs, each with its own approximation of the filtering pdf of the same system state. For example, the two PFs may be tracking the same objects while being executed in different nodes within a sensor network. In a standard distributed scheme the observation set at time k in one node, z_k^1, may be shared with another node whose own observation set is z_k^2, and vice versa. In each node, a PF employs the likelihoods $p(z_k^1 \mid x_k)$ and $p(z_k^2 \mid x_k, z_k^1)$. When there are many sensors, however, it may be practically infeasible to disseminate the observations of one node to all other nodes in the network. Moreover, the interdependencies between observations, described by likelihoods of the form $p(z_k^j \mid x_k, \{z_k^i\}_{i \neq j})$, where z_k^j is the observation set of the jth node, may not be known. On the other hand, the particle approximations of the filtering pdfs at neighboring nodes can always be combined in a (KL) consistent manner so as to improve one another, as well as the PFs in other nodes.

Suppose that the pair of PFs exchange their particle approximations or communicate them to a third location, where this information may be fused. The difficulty in applying (1) to the approximated pdfs, $\pi(x_k \mid z_{0:k}^1)$ and $\pi(x_k \mid z_{0:k}^2)$, may be appreciated by noting that in this case (1) involves a product of two discrete pdfs. It now follows that $q(x_k)$ vanishes unless x_k coincides with the same particle in both sets, $\{x_k^{(i),1}\}_i$ and $\{x_k^{(i),2}\}_i$, where $\{x_k^{(i),j}\}_i$ is the particles approximation of the jth PF. This issue may be alleviated by using instead

$$\pi_r(x_k \mid z_{0:k}^j) \propto \sum_{i=1}^{N} w_k^{(i),j} \exp\left\{-\beta \parallel x_k - x_k^{(i),j} \parallel_2^2\right\} \quad (9)$$

for $j = 1, 2$, where $\beta > 0$ is a regularization parameter, and $\| \cdot \|_2$ is the Euclidean norm.

Substituting (9) into (1) yields the approximation $q(x_k) \propto \pi_r(x_k \mid z_{0:k}^1) \triangleright_\alpha \pi_r(x_k \mid z_{0:k}^2)$. There are two ways how to use this pdf estimate. We may employ a sampling technique such as the Metropolis-Hastings to produce a new set of particles from $\hat{q}(x_k)$, which would then substitute the particle approximations in both PFs. Alternatively, we may substitute the weights $\{w_k^{(i),j}\}_i$ of the jth PF for new (unnormalized) weights $\{q(x_k^{(i),j})\}_i$. The latter approach, which is employed in this work, saves the burden of producing a new set of particles per fusion operation. This *interaction* procedure is summarized in Algorithm 1.

Algorithm 1 Interaction via particles intersection

Input: $\{x_k^{(i),1}, w_k^{(i),1}\}_{i=1}^N$ and $\{x_k^{(i),2}, w_k^{(i),2}\}_{i=1}^N$

Output: $\{x_k^{(i),1}, v_k^{(i),1}\}_{i=1}^N$ and $\{x_k^{(i),2}, v_k^{(i),2}\}_{i=1}^N$

for $j = 1 : N$ **do**

$\bar{v}_k^{(j),1} = \pi_r(x_k^{(j),1} \mid z_{0:k}^1) \triangleright_\alpha \pi_r(x_k^{(j),1} \mid z_{0:k}^2)$

$\bar{v}_k^{(j),2} = \pi_r(x_k^{(j),2} \mid z_{0:k}^1) \triangleright_\alpha \pi_r(x_k^{(j),2} \mid z_{0:k}^2)$

end for

for $j = 1 : N$ **do**

$v_k^{(j),1} = \bar{v}_k^{(j),1} / \left(\sum_{l=1}^N \bar{v}_k^{(l),1}\right)$

$v_k^{(j),2} = \bar{v}_k^{(j),2} / \left(\sum_{l=1}^N \bar{v}_k^{(l),2}\right)$

end for

4. NUMERICAL STUDY

4.1. KL consistency

The application of (1) in the case of Gaussian mixtures has already been discussed in [17]. Here, a similar example is used for demonstrating the consequences of Theorem 1. We pick a particular Gaussian mixture, $h(x) = \frac{1}{4} \sum_{i=1}^4 \mathcal{N}(x \mid \mu_i, \sigma_i^2)$, and draw an increasing number, m, of samples from it. Here, the means and variances are $\mu_i \in [-1, 1]$ and $\sigma_i^2 \in [0.01, 0.1]$. An expectation maximization (EM) algorithm is then employed for estimating the parameters of $h(x)$ based on these samples. Per m we repeat this procedure twice for generating two set of different samples which are then used by the EM to yield the density estimates $f(x)$ and $g(x)$. In this example, α is numerically optimized such that $I_\alpha(f, g)$ is maximized (7). The respective KL-divergences averaged over 70 Monte Carlo runs, where in each run new sets of samples are drawn for $m = 100$ to $m = 1000$, are shown in Figure 1.

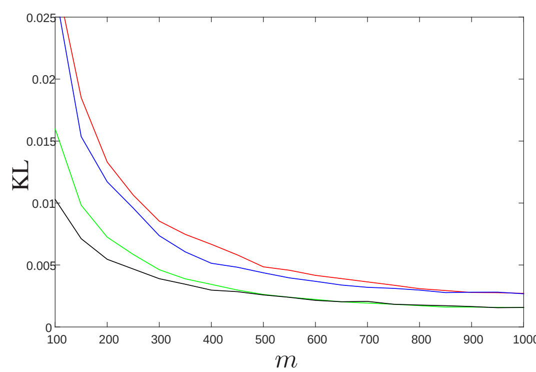

Fig. 1. The average KL-divergence KL $(q \parallel h)$ (green), where $q(x)$ is given by (1), is smaller than the corresponding KL-divergences of any of the density estimators $f(x)$ and $g(x)$ (red and blue). The black line shows the corresponding KL-divergence of an ideal density estimator that uses the data sets of both estimators.

4.2. Particles intersection for multiple object tracking

In what follows we demonstrate the utility of particles intersection in a distributed tracking scenario. A pair of sequential importance resampling (SIR) PFs are employed for tracking a variable number of objects. Like in [21], the ith extended object in the scene is represented by a two-dimensional Gaussian $\mathcal{N}(\cdot \mid \mu_k^i, C_k^i)$, whose mean μ_k^i and 2×2 covariance C_k^i correspond to the location and spatial extent of the object at time k. When a number of objects, not greater than n, appear together, this representation amounts to a Gaussian mixture, $L(z \mid x_k) \propto \sum_{i=1}^n e_k^i \mathcal{N}(z \mid \mu_k^i, C_k^i)$, where

$x_k \stackrel{\text{def}}{=} \{\{\mu_k^i\}_{i=1}^n, \{C_k^i\}_{i=1}^n, \{e_k^i\}_{i=1}^n\}$. Here, $e_k^i \in \{0, 1\}$ is a realization of a Bernoulli random variable indicating whether or not the ith object is present at time k. The likelihood function used by the jth PF is

$$p(\mathcal{Z}_k^j \mid x_k) = \prod_{i=1}^{m_k^j} L(z_k^j(i) \mid x_k) \tag{10}$$

where $\mathcal{Z}_k^j \stackrel{\text{def}}{=} \{z_k^j(i)\}_{i=1}^{m_k^j}$ denotes a set of m_k^j statistically independent observations at time k.

The evolution of objects' positions, spatial extent, and indicators is described similarly to [21] by a discrete-time Markov process,

$$p(x_{k+1} \mid x_k) = p(\mu_{k+1} \mid \mu_k) p(C_{k+1} \mid C_k) p(e_{k+1} \mid e_k) \tag{11}$$

where $p(x_0)$ is known. The PFs approximate the filtering pdfs $p(x_k \mid \mathcal{Z}_{0:k}^j)$, $j = 1, 2$, where $\mathcal{Z}_{0:k}^j \stackrel{\text{def}}{=} \{\mathcal{Z}_0^j, \ldots, \mathcal{Z}_k^j\}$, from which the underlying MMSE estimators are obtained.

A Gazebo environment is used for generating a realistic scenario where the movement of four robotic platforms are recorded by laser scanners at two different locations. One of the laser scanners is used for producing the set $\mathcal{Z}_k^1$ of point observations, and another for producing the set $\mathcal{Z}_k^2$. These observations are contaminated with a zero-mean Gaussian white noise whose standard deviation is 0.03 in both coordinates. There are $m_k^j = 20$ observations in each set of which 10 percent represent clutter uniformly distributed in the sensor field of view. In this scenario a robot may appear or disappear at some predetermined times to which the PFs are oblivious. The parameters underlying the time propagation models in the PFs are the same as those used in the synthetic example in [21]. The regularization parameter β in (9) is taken as 0.7.

The proposed interaction scheme has been analyzed in the above tracking scenario based on 100 Monte Carlo runs. In each run the initial particles of each PF are drawn from a Gaussian mixture whose parameters are obtained by the K-means algorithm using the first set of observations. Several PFs, each with $N = 150$ particles, are employed: a pair of PFs which interact through particles intersection, another pair of PFs that use the same observations but which do not interact, and a centralized PF that uses the observations from both laser sensors. The interaction between the two PFs, one which is fed with $\mathcal{Z}_k^1$ and another with $\mathcal{Z}_k^2$, is carried out as described in Algorithm 1. For comparison, a standard CI is also implemented using corresponding state estimates and covariances of individual objects from both non-interacting PFs.

The mean square estimation errors (MSE) of the various PFs are shown in Figure 2. The advantage of the interacting PFs over their non-interacting counterparts is apparent in this figure. In contrast, the standard CI technique, which is also shown, exhibits MSE errors no better and occasionally worse than those of the non-interacting PFs. The visible peaks in this figure reflect changes in the actual number of objects. As shown in Figure 3, the accuracy of the estimated number of objects by one of the interacting PFs is better than that of any of the non-interacting PFs.

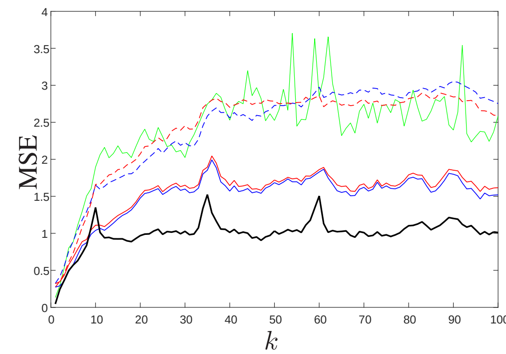

Fig. 2. MSE errors evaluated using 100 Monte Carlo runs. The performance of a centralized PF that uses observations from both laser scanners is shown in black. The estimation errors of the interacting PFs are shown as the solid red and blue lines. The respective estimation errors of the non-interacting PFs are shown by the dashed red and blue lines. The MSE errors of the CI technique is shown in green.

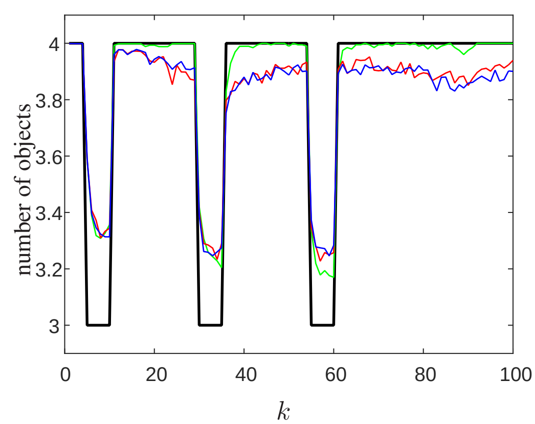

Fig. 3. Estimated (colored lines) and actual (black line) number of objects. The red and blue lines are the estimated number of objects averaged over 100 Monte Carlo runs of the non-interacting PFs. The respective estimate of one of the interacting PFs is shown in green.

5. CONCLUSIONS

In this paper, a technique named particles intersection is proposed for fusing the empirical pdf approximations of a pair of PFs. It relies on a fusion approach of which covariance intersection is a special case. The pdf estimate obtained using particles intersection is shown here to be closer in terms of KL-divergence to the actual pdf than any of the pdf estimators alone. The viability of this technique is demonstrated in a scenario where a pair of PFs track a group of objects.

6. REFERENCES

[1] M. Bahramgiri, M. Hajiaghayi, and Mirrokni, "Fault-tolerant and 3-dimensional distributed topology control algorithms in wireless multi-hop networks," *Wireless Networks*, vol. 12, no. 2, pp. 179–188, 2006.

[2] S. Bornholdt and R. Thimo, "Topological evolution of dynamical networks: Global criticality from local dynamics," *Physical Review Letters*, vol. 84, no. 26, pp. 6114–6117, 2000.

[3] C. Zhou and J. Kurths, "Dynamical weights and enhanced synchronization in adaptive complex networks," *Physical review letters*, vol. 96, no. 16, pp. 164102, 2006.

[4] C. Chong and S.P. Kumar, "Sensor networks: evolution, opportunities, and challenges," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1247–1256, 2003.

[5] R.R. Brooks, P. Ramanathan, and A.M. Sayeed, "Distributed target classification and tracking in sensor networks," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1163–1171, 2003.

[6] R. Olfati-Saber and J.S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*. IEEE, 2005, pp. 6698–6703.

[7] P. Yang, R. A Freeman, and K.M. Lynch, "Multi-agent coordination by decentralized estimation and control," *IEEE Transactions on Automatic Control*, vol. 53, no. 11, pp. 2480–2496, 2008.

[8] N. Xiong and P. Svensson, "Multi-sensor management for information fusion: issues and approaches," *Information fusion*, vol. 3, no. 2, pp. 163–186, 2002.

[9] S.J. Julier and J.K. Uhlmann, "General decentralized data fusion with covariance intersection," *Handbook of multisensor data fusion: theory and practice*, pp. 319–344, 2009.

[10] M. Coates, "Distributed particle filters for sensor networks," in *Proceedings of the 3rd international symposium on Information processing in sensor networks*. ACM, 2004, pp. 99–107.

[11] D. Gu, "Distributed particle filter for target tracking," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, April 2007, pp. 3856–3861.

[12] S. Farahmand, S.I. Roumeliotis, and G.B. Giannakis, "Set-membership constrained particle filter: Distributed adaptation for sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4122–4138, 2011.

[13] A.H. Jazwinski, *Stochastic processes and filtering theory*, Courier Corporation, 2007.

[14] O. Hlinka, O. Sluiak, F. Hlawatsch, and M. Rupp, "Distributed data fusion using iterative covariance intersection," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 1861–1865.

[15] L. Mihaylova, A. Carmi, F. Septier, A. Gnig, S. K. Pang, and S. Godsill, "Overview of bayesian sequential Monte Carlo methods for group and extended object tracking," *Digital Signal Processing*, vol. 25, pp. 1–16, 2014.

[16] M. ney, D.E. Clark, and S.J. Julier, "Distributed fusion of phd filters via exponential mixture densities," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 521–531, 2013.

[17] S.J. Julier, "An empirical study into the use of Chernoff information for robust, distributed fusion of gaussian mixture models," in *Information Fusion, 2006 9th International Conference on*. IEEE, 2006, pp. 1–8.

[18] D. Moskovich and A.Y. Carmi, "Low-dimensional topology of information fusion," in *Proceedings of the 8th ACM/ICST International Conference on Bioinspired Information and Communications Technologies*. ACM/ICST, 2014, pp. 251–258.

[19] B. Wang, W. Yi, S. Li, L. Kong, and X. Yang, "Distributed fusion with multi-bernoulli filter based on generalized covariance intersection," in *2015 IEEE Radar Conference (RadarCon)*, 2015, pp. 0958–0962.

[20] M.B. Hurley, "An information theoretic justification for covariance intersection and its generalization," in *Information Fusion, 2002. Proceedings of the Fifth International Conference on*. IEEE, 2002, vol. 1, pp. 505–511.

[21] A. Carmi, F. Septier, and S. Godsill, "The Gaussian mixture MCMC particle algorithm for dynamic cluster tracking," *Automatica*, vol. 48, pp. 2454–2467, 2012.