Fast Weighted Exponential Product Rules for Robust General Multi-Robot Data Fusion

Nisar Ahmed*   Jonathan Schoenberg*   Mark Campbell

Autonomous Systems Laboratory, Cornell University

Ithaca, New York 14853

e-mail:[nra6,jrs55,mc288]@cornell.edu

*both authors contributed equally to this work

Abstract—This paper considers the distributed data fusion (DDF) problem for general multi-agent robotic sensor networks in applications such as 3D mapping and target search. In particular, this paper focuses on the use of conservative fusion via the weighted exponential product (WEP) rule to combat inconsistencies that arise from double-counting common information between fusion agents. WEP fusion is ideal for fusing arbitrarily distributed estimates in ad-hoc communication network topologies, but current WEP rule variants have limited applicability to general multi-robot DDF. To address these issues, new information-theoretic WEP metrics are presented along with novel optimization algorithms for efficiently performing DDF within a recursive Bayesian estimation framework. While the proposed WEP fusion methods are generalizable to arbitrary probability distribution functions (pdfs), emphasis is placed here on widely-used Bernoulli and Gaussian mixture pdfs. Experimental results for multi-robot 3D mapping and target search applications show the effectiveness of the proposed methods.

I. INTRODUCTION

The problem of fusing information from an ensemble of noisy data streams is critical to many existing and soon-to-be-realized robotic systems operating in uncertain dynamic environments. This is particularly true for distributed multi-robot systems requiring distributed perception for applications such as collaborative mapping for exploration, target search/tracking for surveillance, and futuristic unmanned urban transport systems. For sensor agents in a network to share information and perform *distributed data fusion (DDF)*, it is most desirable to establish a scalable, flexible and robust network over which the robots can transmit and receive information. An ad-hoc and arbitrary connected network provides scalability for fusion agents to join and drop off the network, flexibility to allow agents to join at any point and robustness to ensure multiple links or agents must fail before the network becomes unconnected [5].

Implementation of DDF for general robot sensor networks thus requires conservative data fusion techniques to maintain estimates that avoid inconsistencies due to rumor propagation [3]. A common conservative fusion rule for estimates with Gaussian probability distribution functions (pdfs) with unknown correlation is Covariance Intersection (CI) [11]. This rule is appropriate for certain types of problems, but it is inadequate for handling non-Gaussian distributions that arise in applications such as target search and 3D mapping. A suitable conservative fusion rule for *arbitrary* pdfs with unknown

correlation is the weighted exponential product (WEP) [1]. The WEP is a generalization of CI to non-Gaussian distributions [9] and different cost metrics to determine an optimal WEP fusion weight have been proposed [3, 9]. However, these existing WEP fusion approaches have drawbacks that can limit their usefulness in robotic DDF, including the theoretical nature of their cost metrics and difficult implementation for arbitrary pdfs.

This paper makes the following contributions: (1) it proposes novel information-theoretic metrics for performing WEP fusion that address these issues and are suitable for fusing arbitrary state estimates shared via ad-hoc network communication topologies in a wide range of robotic DDF applications; (2) it presents new formally consistent algorithms for online implementation of our proposed WEP fusion metrics that can be used to quickly and robustly combine information in a recursive Bayesian estimation framework, with emphasis on the Bernoulli and Gaussian mixture distribution functions used widely in robotics; (3) it demonstrates the effectiveness of our proposed WEP fusion methods for performing online collaborative 3D mapping and 2D target search with multi-robot networks in loopy communication topologies.

A. DDF Preliminaries

Formally, let $x^k \in \mathbb{R}^{n_x}$ be an n_x-dimensional state to be estimated at discrete time steps $k = 0, 1, ...$ by N independent robotic sensor agents. Assume that each robot $i \in \{1, ..., N\}$ obtains n_y local sensor measurements of x^k in the vector y_i^k with likelihood function $p_i(y_i^k|x^k)$. Let $p_i(x^k|Z_i^k)$ be the local posterior pdf for robot i given the set of all information $Z_i^k = \{Z_i^{k-1}, y_i^k\}$ available to i *before* new information Z_j^k arrives from robot $j \neq i$, such that local fusion of y_i^k is given by the posterior pdf from Bayes' rule,

$$p_i(x^k|Z_i^k) \propto p_i(x^{k-1}|Z_i^{k-1})p_i(y_i^k|x^k). \tag{1}$$

For generality, the robot networks considered here are assumed to have arbitrary dynamic node-to-node communication topologies, such that: (i) N may vary; (ii) each robot is only aware of nodes it connected to; (iii) no robot ever knows the complete global network topology; (iv) no robot knows the receipt status of messages it has sent.

Centralized fusion or raw data sharing can fully recover new information in $\bigcup_{i=1}^{N} Z_i^k$, but such methods scale poorly with

N and are vulnerable to node failures. DDF overcomes such bottlenecks and can recover the centralized Bayesian fusion result for a given pair of robots (i, j) [8], since

$$p(x^k|Z_i^k \cup Z_j^k) \propto \frac{p_i(x^k|Z_i^k)p_j(x^k|Z_j^k)}{p(x^k|Z_i^k \cap Z_j^k)}, \quad (2)$$

where the denominator is the *common information pdf* for the local pdf $p_i(x^k|Z_i^k) \propto p(x^k|Z_i^k \cap Z_j^k) \cdot p(x^k|Z_{i\setminus j}^k)$, where $Z_{i\setminus j}^k$ is the exclusive local information set for node i and similarly for node j. Hence, the common information pdf must be explicitly known for DDF to optimally extract only new information in $Z_i^k \cup Z_j^k$. Note that if division by $p(x^k|Z_i^k \cap Z_j^k)$ were ignored in (2), the resulting 'Naive Bayes' fusion rule would always double-count $p(x^k|Z_i^k \cap Z_j^k)$ and induce inconsistencies via rumor propagation. However, tracking $p(x^k|Z_i^k \cap Z_j^k)$ in an arbitrary ad-hoc network requires pedigree-sharing about each distribution to fuse; this becomes computationally prohibitive for non-tree communication topologies with loops/cycles [3, 8]. The generality of assumptions (i)-(iv) above leads to difficulties for implementing DDF via graphical model techniques.

An alternative method for consistent DDF considered here is the conservative weighted exponential product (WEP) fusion rule,

$$p_{\text{WEP}}(x^k|Z_i^k \cup Z_j^k) = \frac{1}{c}p_i^\omega(x^k|Z_i^k) \cdot p_j^{1-\omega}(x^k|Z_j^k), \quad (3)$$
$$\equiv p_{\text{WEP}}(x^k)$$

where $\omega \in [0, 1]$ and c is a normalizing constant. The WEP rule has two properties that make it well-suited to general multi-robot DDF problems: (1) it is applicable to arbitrary non-Gaussian pdfs, and (2) for any $\omega \in [0, 1]$, it is guaranteed to avoid double-counting of $p(x^k|Z_i^k \cap Z_j^k)$ [1]. However, this also means some exclusive information from i and j is always discarded. Hence, the WEP fusion weight ω must be picked carefully to ensure useful information exchange between robots i and j.

B. WEP Fusion Metrics

Ref. [9] considers two popular cost metrics on (3) for optimal selection of $\omega \in [0, 1]$. The first cost is the Shannon entropy of $p_{\text{WEP}}(x^k)$, which is motivated by the idea that the absolute uncertainty of (3) should be minimized. Ref. [9] shows that the resulting Shannon fusion rule [3] is a direct non-Gaussian generalization of the Gaussian CI rule. The second cost selects the $\omega = \omega^*$ corresponding to the Chernoff Information between $p_i(x^k|Z_i^k)$ and $p_j(x^k|Z_j^k)$ [4], and is thus called Chernoff fusion [6, 10]. It can be shown that this equates the Kullback-Leibler divergences (KLDs) between (3) and the individual robot posteriors, which guarantees that the Chernoff fusion result is 'half-way' between the original local i and j posteriors in an information gain/loss sense. Both the Shannon and Chernoff fusion rules provide nice information theoretical reasons for choosing ω, but each rule has potential limitations in the context of general multi-robot DDF, as explained in Section II.

C. WEP Fusion Algorithms

General WEP fusion can present challenging implementation issues when dealing with arbitrary non-Gaussian pdfs. For instance, both Shannon and Chernoff fusion require minimization of integral costs that are in general analytically intractable, making WEP fusion potentially unsuitable for online robotic DDF problems. Furthermore, (3) is generally not available in closed-form. Heuristic WEP fusion approximations have been proposed to address these issues [6, 10], but these have no formal guarantees of either correctly minimizing the desired WEP cost or accurately approximating (3). Such heuristic approximations can thus lose information that should be preserved between robots or introduce spurious information.

II. New Information-based WEP Metrics

A. Limitations of Existing Metrics

Shannon fusion is not always justified from a general Bayesian estimation perspective, since acquisition of new information can, in many cases, *increase* the entropy of the exact/centralized Bayesian posterior for non-Gaussian pdfs [1]. Chernoff fusion ignores the possibility that an imbalance in information gain may exist between robots i and j prior to fusion. To address such scenarios, ref. [9] suggests introducing weighting factors into Chernoff fusion that are equal to the number of measurements taken by each robot in order to bias ω towards the better informed pdf. However, the number of measurements is not a complete reflection of information gain for each robot, as local differences in sensor quality/characteristics and information obtained from other neighbors in the robot network must also be taken into account.

B. Generalized Information Weighted Chernoff Fusion

The first new fusion rule modifies the measurement weighted Chernoff fusion [9] to apply weighting factors $c_i, c_j \geq 0$ that reflect actual information gain by each robot i and j, instead of the number of measurements. This biases ω towards the constituent pdf with greater information content without explicitly counting measurements, so that ω satisfies

$$c_i \cdot D_{\text{KL}}\left[p_{\text{WEP}}(x^k)||p_i(x^k|Z_i^k)\right]$$
$$= c_j \cdot D_{\text{KL}}\left[p_{\text{WEP}}(x^k)||p_j(x^k|Z_j^k)\right], \quad (4)$$

where $D_{\text{KL}}[\cdot||\cdot]$ is the KLD. For problems y_i^k are always guaranteed to reduce uncertainty (e.g. static linear Gaussian estimation), c_i and c_j can be conveniently estimated as the reciprocal of the entropy of the constituent distributions,

$$c_i = \frac{1}{H[p_i(x^k|Z_i^k)]}, \ c_j = \frac{1}{H[p_i(x^k|Z_i^k)]}, \quad (5)$$

where $H[p(x)]$ is the Shannon entropy of $p(x)$. This leads to an Entropy Weighted Chernoff Fusion solution that uses a weight $\omega^* \in [0, 1]$ to equate the entropy weighted KLDs between the WEP and original distributions. This results in a bias towards the distribution with the lower entropy along the chord that connects the two distributions according to the KLDs.

However, as discussed in Section II-A, the Shannon entropy is not always a suitable indicator of information gain. It is also unsuitable for fusing distributions over continuous states x^k, since the differential Shannon entropy of probability density functions can be negative (whereas c_i, c_j should be non-negative). One possible alternative is to select c_i and c_j to be the information gained in the KLD sense with respect to a common reference distribution $p_r(x^k)$ for robots i and j. If a common prior pdf $p(x^0)$ is available to all agents at $k = 0$ for initializing Bayesian estimation, then this forms a convenient choice for $p_r(x^k)$, so that

$$c_i = D_{\text{KL}} \left[p_i(x^k|Z_i^k)||p(x^0) \right], \tag{6}$$

and similarly for c_j. These weights reflect the total information gain of each agent over time with respect to $p(x^0)$ via the combined effects of local Bayesian fusion and DDF. Note that a fixed reference distribution must be used with the KLD, since it is not a true distance metric and does not obey the triangle inequality. The KLD could be replaced by another divergence that is a true metric (e.g. the Hellinger divergence [2] or the symmetric KLD), which removes the need for a common prior and permits c_i, c_j to be computed incrementally, but that is not demonstrated here.

C. Minimum Information Loss Weight Fusion

A second cost is now developed to minimize an approximation on the information lost as a result of WEP fusion. Eq. (3) discounts exclusive information at i and j to ensure that the unknown common information between the two is counted only once. Therefore, there is an information loss between the optimal fusion pdf (2) and the WEP fusion pdf (3), which is defined

$$\mathcal{I}_{\text{LOSS}} \triangleq D_{\text{KL}} \left[p_{\text{OPTIMAL}}(x|Z_i^K \bigcup Z_j^K)||p_{\text{WEP}}(x|Z_i^K \bigcup Z_j^K) \right] \tag{7}$$

which cannot be computed without knowing the optimally fused distribution. However, note that if i and j are known to have no information in common, then the optimal fusion rule simplifies to the Naive Bayes (NB) fusion,

$$p_{\text{NB}}(x|Z_i^K \bigcup Z_j^K) = \frac{1}{c}p(x|Z_i^K)p(x|Z_j^K). \tag{8}$$

In the case of an unknown dependence between the i and j pdfs, the discounting of exclusive information resultant from WEP fusion leads to information loss if the true dependence between the distribution is zero. Therefore, an approximation to the information loss is the KLD between the NB and WEP fusion pdfs,

$$\begin{aligned} \mathcal{I}_{\text{LOSS}} &\approx \bar{\mathcal{I}}_{\text{LOSS}}(\omega) \\ &\triangleq D_{\text{KL}} \left[p_{\text{NB}}(x|Z_i \bigcup Z_j)||p_{\text{WEP}}(x|Z_i \bigcup Z_j) \right] \end{aligned} \tag{9}$$

It is now possible to compute the fusion weight $\omega^* \in [0, 1]$ to minimize the approximate information loss,

$$\omega^* = \arg\min_{\omega \in [0,1]} \bar{\mathcal{I}}_{\text{LOSS}}(\omega). \tag{10}$$

The advantage of this Minimum Information Loss Fusion (10) scheme is that ω is selected to minimize the possible information loss (7) should the two distributions be truly uncorrelated. In practice, Minimum Information Loss Fusion drives the solution towards Naive Bayes fusion in the case when the distributions to fuse are significantly different (in a Kullback-Leibler sense), which could indicate the two do not share a significant amount of common information. The Minimum Information Loss Fusion scheme provides an automatic method for trading between the Naive Bayes and conservative fusion rules, without resorting to a heuristic decision.

III. WEP FUSION FOR MULTI-ROBOT OCCUPANCY GRIDS AND GAUSSIAN MIXTURE BELIEFS

The proofs have been omitted due to limited space, but it can be shown that the newly proposed metrics Section II are all convex in ω for arbitrary pdfs $p_i(x^k|Z_i^k)$ and $p_j(x^k|Z_j^k)$. In this section, the problem of implementing WEP fusion using the (un)weighted Chernoff and Minimum Information Loss cost metrics is addressed. The focus is on cases where both $p_i(x^k|Z_i^k)$ and $p_j(x^k|Z_j^k)$ are described by either discrete Bernoulli distributions or continuous finite Gaussian mixtures, which are commonly used in robotics, e.g. for occupancy grid mapping and object tracking/localization, respectively. With some minor modifications, the methods described here can be readily applied in other contexts and to other arbitrary discrete/continuous pdfs.

A. Fast WEP Fusion for Bernoulli Distributions

Applying the general WEP fusion rule (3) to two Bernoulli distributions with unknown correlation results in a fused distribution that is also Bernoulli,

$$p_i(x^k|Z_i^k) = p_i, \quad p_j(x^k|Z_j^k) = p_j \tag{11}$$

$$p_{\text{WEP}}(x^k) = \frac{p_i^\omega p_j^{1-\omega}}{p_i^\omega p_j^{1-\omega} + (1-p_i)^\omega (1-p_j)^{1-\omega}}. \tag{12}$$

In the case of Chernoff fusion, the optimal weight ω^* is given in closed form,

$$\omega_{\text{CF}}^* = \frac{\log \frac{[\log(1-p_i)-\log(1-p_j)]}{[\log(p_i)-\log(p_j)]} - \log p_j + \log(1-p_j)}{[\log p_i - \log p_j - \log(1-p_i) + \log(1-p_j)]} \tag{13}$$

For Entropy Weighted Chernoff Fusion, ω^* cannot be found in closed form, and the following equation must be solved numerically,

$$\frac{1}{H[p_i]} \left[p_\omega \log \frac{p_\omega}{p_i} + (1-p_\omega) \log \frac{(1-p_\omega)}{(1-p_i)} \right] - \cdots$$
$$\frac{1}{H[p_j]} \left[p_\omega \log \frac{p_\omega}{p_j} + (1-p_\omega) \log \frac{(1-p_\omega)}{(1-p_j)} \right] = 0 \tag{14}$$

where p_ω is given by (12) and the entropy of a Bernoulli distribution is $H[p] = -p \log p - (1-p) \log(1-p)$. When the entropy of the two distributions is equal, Entropy Weighted Chernoff Fusion collapses to regular (i.e. unweighted) Chernoff fusion.

In the case of the Minimum Information Loss fusion, ω^* is given in closed form,

$$\omega_{\text{M-L}}^* = \frac{\log p_j - \log(1 - p_j) - \log\left[\frac{p_i p_j}{(p_i - 1)(p_j - 1)}\right]}{\log p_j - \log(1 - p_j) - \log p_i + \log(1 - p_i)}. \quad (15)$$

B. Fast WEP Fusion for Finite Gaussian Mixtures

Assume each robot's pdf is a finite Gaussian mixture (GM),

$$p_i(x^k | Z_i^k) = p_i(x^k) = \sum_{m=1}^{M^i} w_m^i \mathcal{N}(\mu_m^i, \Sigma_m^i) \quad (16)$$

where (for agent i and similarly for j) M^i is the number of mixands, μ^i and Σ^i are the mean and covariance matrix of component m, and w^m is the mth mixing weight such that $w_m \in [0, 1]$ and $\sum_{m=1}^{M} w^m = 1$. The WEP cost metric integrals and the resulting fusion pdf (3) are not closed form for GMs. A new fast and consistent two-phase Monte Carlo-based optimization and learning procedure is presented to overcome these issues.

1) Particle-based optimization: The Information Weighted Chernoff fusion cost can be written as

$$f_{\text{WEP}}^k(\omega) = \int h_{ij}(x^k) \cdot p_{\text{WEP}}(x^k) dx^k, \quad (17)$$

$$\text{where } h_{ij}(x^k) = \frac{c_i \log p_i(x^k) - c_j \log p_j(x^k)}{\log p_i(x^k) - \log p_j(x^k)},$$

(where unweighted Chernoff fusion uses $c_i = c_j = 1$); the Minimum Information Loss cost (4) can be written as

$$f_{\text{WEP}}^k(\omega) = \omega \cdot \kappa + \log \int h_{ij}(x^k) \cdot p_{\text{WEP}}(x^k) dx^k, \quad (18)$$

$$h_{ij}(x^k) = 1, \quad \kappa = \int p_i(x^k) p_j(x^k) \log\left(\frac{p_j(x^k)}{p_i(x^k)}\right) dx^k$$

The integral terms with respect to $p_{\text{WEP}}(x^k)$ in each of these analytically intractable costs are approximated via importance sampling,

$$\int h_{ij}(x^k) \cdot p_{\text{WEP}}(x^k) dx^k \approx \frac{1}{N_s} \sum_{s=1}^{N_s} \theta_s(\omega) \cdot h_{ij}(x_s), \quad (19)$$

$$\theta_s(\omega) = \frac{p_i^\omega(x_s) p_j^{1-\omega}(x_s)}{g(x_s; \omega)} \quad (20)$$

where $h_{ij}(x_s)$ are sample values of $h_{ij}(x^k)$ and $\theta_s(\omega)$ are unnormalized non-negative weights for N_s i.i.d. samples $\{x_s\}_{s=1}^{N_s}$ drawn from a fixed importance pdf $g(x^k; \omega)$. The constants c_i and c_j in (6) and κ in (18) are generally not available in closed form for GMs, but can be well-approximated estimated with fast deterministic methods [7].

To ensure that (19) is well-behaved and easy to compute, the importance pdf $g(x^k; \omega)$ should be easy to sample from and evaluate, have non-zero support over the entire domain of integration and resemble (3) closely enough to control the variance of (19). Many choices are possible for $g(x^k; \omega)$, but

---

Algorithm 1 GM WEP Fusion via Importance Sampling

Inputs: robot GM pdfs $p_i(x^k)$ and $p_j(x^k)$; number of samples N_s; initial guess ω_0; importance pdf exponent $\bar{\omega}$; update function $[\omega_{\text{old}}, \omega_{\text{curr}}] \leftarrow R_{1D}[\tilde{f}(\omega_{\text{curr}}), \omega_{\text{curr}}, \omega_{\text{old}}]$ for desired 1D minimization rule; effective sample size threshold, N_t
Output: $\tilde{\omega}^* \in [0, 1]$; samples $\{x_s\}_{s=1}^{N_s}$ with normalized weights $\{\theta_s\}_{s=1}^{N_s}$

1. Initialize $\omega_{\text{curr}} \leftarrow \omega_0$ and ω_{old} according to R_{1D}
2. construct GM importance sampling pdf $g(x; \bar{\omega})$ via (21)
3. draw N_s samples $\{x_s\}_{s=1}^{N_s} \sim g(x; \omega_{\text{curr}})$
4. store pdf values $p_i(x^k)$, $p_j(x^k)$, $g(x; \omega_{\text{curr}})$ for $\{x_s\}_{s=1}^{N_s}$
while ω_{curr} not converged **do**
    5. compute unnormalized importance weights $\{\theta_s\}_{s=1}^{N_s}$ via (20)
    6. using (19), compute $\tilde{f}(\omega_{\text{curr}})$ via (17) or (18)
    7. update $[\omega_{\text{old}}, \omega_{\text{curr}}] \leftarrow R_{1D}[\tilde{f}(\omega_{\text{curr}}), \omega_{\text{curr}}, \omega_{\text{old}}]$
end while
8. if effective sample size $< N_t$, set $\bar{\omega} = \omega_{\text{curr}}$ and go to step 1.
9. normalize $\{\theta_s\}_{s=1}^{N_s}$ s.t. $\sum_{s=1}^{N_s} \theta_s = 1$ and return $\tilde{\omega}^* = \omega_{\text{curr}}$.

---

the following 'first-order' GM approximation to (3) proposed by [10] is particularly effective,

$$g(x; \omega) = \sum_{m=1}^{M^{\text{F}}} w_m^{\text{F}} \mathcal{N}(\mu_m^{\text{F}}, \Sigma_m^{\text{F}}) \quad (21)$$

$$\Sigma_m^{\text{F}} = \left(\omega(\Sigma_q^i)^{-1} + (1 - \omega)(\Sigma_r^j)^{-1}\right)^{-1} \quad (22)$$

$$\mu_m^{\text{F}} = \Sigma_m^{\text{F}}\left(\omega(\Sigma_q^i)^{-1}\mu_q^i + (1 - \omega)(\Sigma_r^j)^{-1}\mu_r^j\right) \quad (23)$$

$$w_m^{\text{F}} = \frac{(w_q^i)^\omega (w_r^j)^{1-\omega}}{\sum_{q', r'} (w_{q'}^i)^\omega (w_{r'}^j)^{1-\omega}}, \quad (24)$$

which is a component-wise covariance intersection operation on the constituent GMs (with M^{F} resulting mixands).

Eq. (19) is unbiased and converges to the true value of the integral as $N_s \to \infty$. Hence, a numerical 1D optimization routine can obtain an unbiased estimate $\tilde{\omega}^*$ of ω using (19) to evaluate (17) or (18). The 1D search is accelerated by using a fixed importance pdf $g(x; \omega) = g(x; \bar{\omega})$, where $\bar{\omega}$ is the same in all optimization steps. As such, the samples $\{x_s\}_{s=1}^{N_s}$ are drawn once and the pdf values $\{p_i(x_s), p_j(x_s), g(x_s; \bar{\omega})\}$ are stored as constants for recalculating (20) at new ω values. The resulting optimization procedure is shown in Algorithm 1, where $R_{1D}[\tilde{f}(\omega_{\text{curr}}), \omega_{\text{curr}}, \omega_{\text{old}}]$ denotes an ω-update rule for any desired 1D optimization algorithm (e.g. bisection search or golden section search) and $\bar{\omega} = 0.5$ is typically used in our applications as an initial estimate. Algorithm 1 typically converges quite quickly and therefore may be run again if it terminates with an effective sample size less than some prescribed threshold N_t. In this case, $\bar{\omega}$ is reset to the current best estimate for ω to restart the optimization.

2) WEM Learning: Immediately following optimization, density estimation procedures are applied to the importance samples to recover an accurate GM to approximate (3) at $\tilde{\omega}^*$. The Weighted EM (WEM) algorithm is used for GMs to exploit the *normalized* importance weights as part of the learning problem. Details for implementing WEM learning of GMs from weighted samples are given in [7]. The initial GM parameters for WEM are found via compressing (21) to a desired maximum number of GM terms, M^{MAX}, using [13].
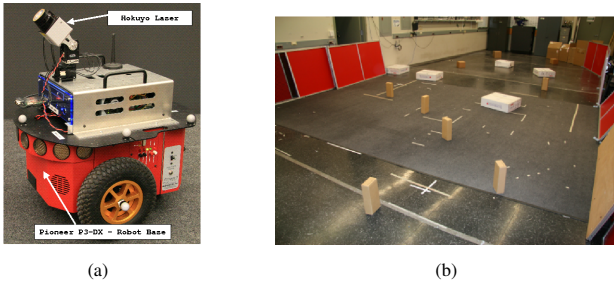
(a)   (b)

Fig. 1. (a) Mobile robot set up with Hokuyo lidar and on-board computer for distributed occupancy grid map fusion experiment. (b) Laboratory environment for the occupancy grid mapping experiment is 15×8 meters and contains boxes of different sizes.
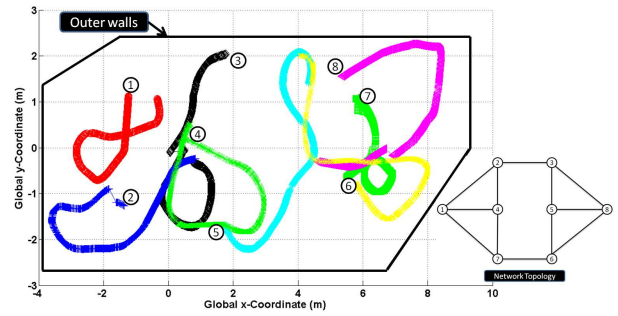

Fig. 2. The overhead paths of the eight mobile robots exploring the environment for distributed occupancy grid mapping experiment. The data collection interval is 120 seconds and different robots explore different regions of the environment with some overlap to demonstrate the benefit of the generalized fusion rules to distributed mapping. The distributed mapping optimal network topology is shown as an inset.

IV. EXPERIMENTAL RESULTS

A. Multi-Robot Occupancy Grid Map Fusion

The generalized fusion of the Bernoulli distribution (Section III-A) is directly applicable to fusion of multi-robot occupancy grid maps. The probability of a voxel of space being occupied is represented as a Bernoulli probability and each voxel is assumed independent. As a result, the fusion of occupancy grid maps from multiple robots is accomplished by sequentially performing fusion across each of the grid cells in the map. A laboratory experiment demonstrates the fusion of 3D occupancy grid maps based on the different fusion rules derived, including Chernoff Fusion, Entropy Weighted Chernoff Fusion and Minimum Information Loss fusion and compares the resulting approximate information loss as a function of map location for the different techniques.

Data was collected using the Pioneer P3-DX differential drive mobile robot from Mobile Robots Inc. shown in Fig. 1(a). The primary sensor for the occupancy grid mapping is the Hokuyo URG-04X laser scanner which features a 240° field-of-view and angular resolution of 0.36^o and maximum range of ≈ 5 m. The laser is pitched downward 45° and scans along the ground as the robot moves forward.

The test environment is 15×8 meters and is instrumented with a Vicon MX+ precision tracking system for robot localization and 3D object position/attitude determination. The features in the environment consist of boxes between 10 and 25 cm tall that are meant to simulate traffic cones or other similarly sized obstacles for a full-size traffic vehicle. Eight robots are run in different paths around the environment for a 120 second data collection (Fig. 2). The robots are run sequentially to avoid sensing each other during map construction.

The robots each construct a 3D occupancy grid map using the Octomap [14] implementation. The occupancy grid resolution is 0.05 m. To establish a baseline occupancy grid map, the data from all the mobile robots is processed to make the *centralized* solution that is equivalent to all agents sending each laser scan and position report to a central server for processing. The resulting centralized map is shown in Fig. 3(a). The map is rendered displaying only the voxels that are occupied or have a $p_{occ} > 0.8$, the empty and unknown voxels are not shown and the voxels are falsely colored according to
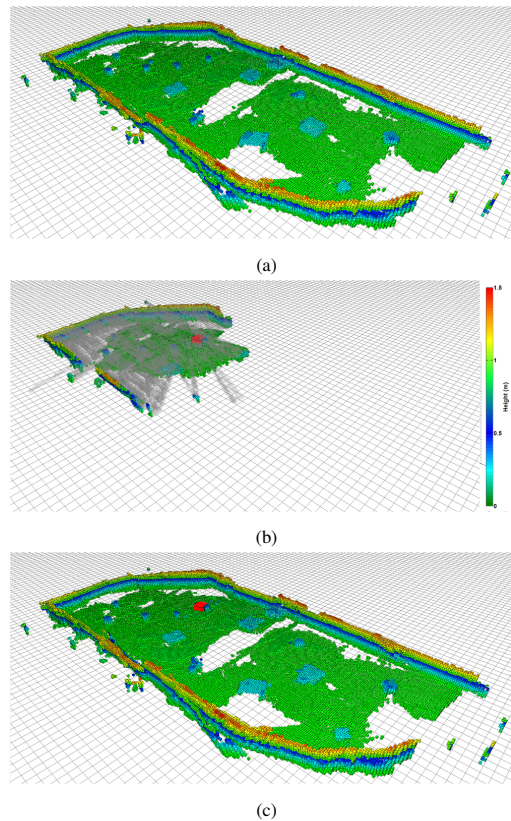

(a)

(b)

(c)

Fig. 3. (a) 3D centralized occupancy grid map displaying the occupied voxels with false coloring based on height. (b) 3D Occupancy grid map from Agent 1 using only local updates. The occupied and empty cells (ghost gray) are shown along with the final pose of the robot (red box). (c) 3D Occupancy grid map from Agent 1 after Minimum Information Loss Fusion; this map is qualitatively similar to the centralized solution.

height according to the scale in Fig. 3(b); a 0.25 m grid is shown in gray for reference. The line of small skinny boxes to the left of the environment are clearly visible along with the larger boxes towards the right of the environment and the outer walls. There are gaps in the center of the map where no robot explored. The centralized solution establishes a baseline map for use in the distributed occupancy grid mapping approaches.

Distributed Occupancy Grid Fusion on Optimally Connected Network: To evaluate the application of the distributed generalized fusion rules to occupancy grid mapping, the mobile robots are connected in a network. The network is connected according to the topology with maximum robustness [5]. The topology is symmetric, has equal node and link connectivity equal to the minimum degree, and each node is a central vertex of the graph. The topology is shown as an inset to Fig. 2. This makes it difficult to track common information during fusion without data tagging and generalized data fusion for unknown correlations is used.

The individual sensor nodes collect and process local scans to build a local occupancy grid map that will be updated with map data passed along the network from other agents. An example of map constructed via local updates only is in Fig. 3(b), which shows the map from Agent 1 (Fig. 2 Red Path). The map is rendered showing the occupied cells falsely colored according to height and the empty cells are shown in a ghost gray; the unknown cells are not shown. The final pose of the robot is shown using a solid red box. The map shows Agent 1 explores only a portion of the map. To enable a full representation of the environment, the distributed data fusion techniques are used over the maximally robust sensor network.

The robots share map information across the bi-directional links connecting the robots. The agents are required to share the following information for each voxel: the center coordinate and the log-odds probability. This implies each voxel requires 32 bytes of data (if all numbers are in double precision) transmitted for each communication. The nodes communicate aperiodically as they collect information. The final occupancy grid map after Minimum Information Loss Fusion is shown in Fig. 3(c). The distributed data fusion via Minimum Information Loss fusion does not have access to the centralized solution nor the optimal fusion distribution, but is successfully utilized to build a map that is qualitatively similar to the centralized solution (Fig. 3(a)). The maps for Chernoff Fusion and Entropy Weighted Chernoff Fusion are similar (not shown).

Approximate Information Loss Maps: The approximate information loss $\bar{\mathcal{I}}_{\text{LOSS}}$ (9) is computed after fusion for each of the different fusion rules. The resulting information loss maps can be used for planning purposes to balance exploration vs verification of cells that may have contained substantial information loss as a result of fusion. The information loss map resulting from Chernoff Fusion is shown in Fig. 4(a). The map is falsely colored according to the approximate information loss metric $\bar{\mathcal{I}}_{\text{LOSS}}$ and the scale is shown in Fig. 4(b). This color range reflects the maximum range of values according to the empirical CDF computed over all possible combinations of two Bernoulli distributions for the different fusion rules (not shown). The areas where the largest information loss occurs when information coming from remote nodes clashes with the content estimated locally at Agent 1 and are in the left hand portion of the map. The areas of information loss to the right hand portion of the map are resulting from the sequential application of the fusion rule as information is received on the
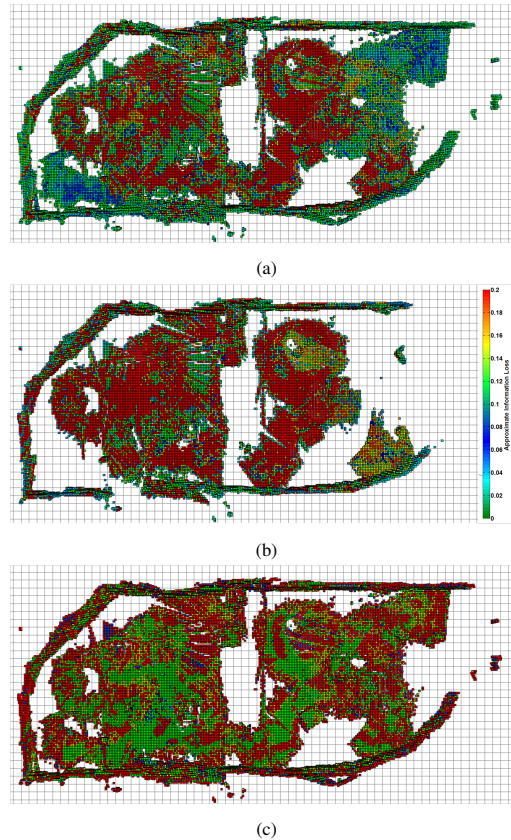

Fig. 4. (a) Approximate information loss $\bar{\mathcal{I}}_{\text{LOSS}}$ on the occupancy grid after Chernoff Fusion for Agent 1. (b) Approximate information loss $\bar{\mathcal{I}}_{\text{LOSS}}$ on the occupancy grid after Entropy Weighted Chernoff Fusion for Agent 1. (c) Approximate information loss $\bar{\mathcal{I}}_{\text{LOSS}}$ on the occupancy grid after Minimum Information Loss Fusion for Agent 1.

three links connected to Agent 1.

The resulting information loss map for Entropy Weighted Chernoff Fusion at Agent 1 is shown in Fig. 4(b). The results shows an increase in some areas and a decrease in the approximate information loss in other areas as a result of Entropy Weighted Chernoff Fusion. The resulting fused map (not shown) is similar to Fig. 3(b). The Minimum Information Loss Fusion rule is significantly better than either approach in terms of reducing the need for verification due to potential information loss, especially along the ground.

The resulting information loss map for Minimum Information Loss Fusion at Agent 1 is shown in Fig. 4(c). The results show a dramatic improvement in the potential information loss, because the fusion rule has zero loss for 50% of the possible combinations of Bernoulli distributions. The Minimum Information Loss Fusion rule generates consistent and quality occupancy grid maps and has the lowest information loss. Therefore, this fusion rule is the best for distributed data fusion of occupancy grid maps.

B. GM Fusion for Multi-robot Target Search

In another experiment, three Pioneer 3D-X robots are tasked to search for a static target with location x_{targ} in an environment similar to the one shown in Fig. 1. Each robot equipped with single forward facing Unibrain Fire-I camera
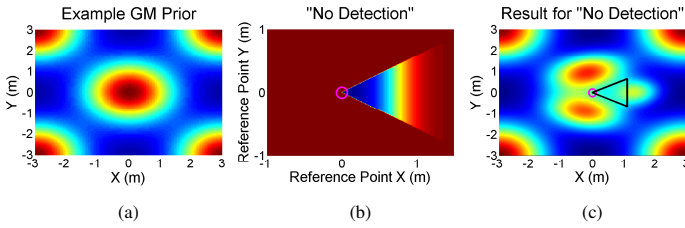
Fig. 5.   (a) Example GM prior over x_{TARG} (red indicates higher density); (b) model for camera-based binary 'Detection'/'No Detection' likelihood (P('No Detection'$|x_{\text{TARG}}$) shown zoomed in; blue indicates probability near 0); (c) posterior GM after fusing GM prior in (a) with 'No Detection' report in (b) via eq.(1) using Gaussian sum filter approximation.

with a 42.5 deg field of view, which feeds onboard visual object detection software that produces binary 'detection'/'no detection' outputs at 1 Hz with 1 m maximum detection range. The likelihood model for the visual object detectors is shown in Fig. 5(b) and the common GM prior pdf for x_{TARG} at $k = 0$ is shown in Fig. 6(a). The sensor likelihood model for each robot is highly non-Gaussian and induces non-Gaussian posterior via (1). As shown in Fig. 5(c), this local Bayesian posterior is well-approximated at each time step k by a finite GM using a variant of the Gaussian sum particle filter with 1000 particles per mixand [12] and a standard GM reduction algorithm to keep $M^{\text{MAX}} = 15$ for each robot [13].

The robots move along different pre-defined search trajectories for 65 secs, during which they fuse their own camera observations to modify their local copy of the GM pdf for x_{TARG} (none of the agents finds the target). The robots communicate 5 times (once every 13 secs) using the cyclic communication pattern $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ for time steps $k = 26$ and $k = 39$, and $1 \rightarrow 3 \rightarrow 2 \rightarrow 1$ for the remaining instances. The unweighted Chernoff, KLD-gain Information Weighted Chernoff, and Minimum Information Loss rules were all evaluated for DDF via Algorithm 1 with $N_s = 1000$. Non-DDF benchmark runs using no sharing and centralized fusion were also evaluated, alongside alternative DDF benchmark runs obtained via Naive Bayes and conservative Bhattacharyya fusion. Bhattacharyya fusion is another WEP fusion rule that always assigns $\omega = 0.5$ [3]. This fusion rule requires no optimization but does require approximation of the WEP fusion pdf, which is accomplished by applying WEM to weighted importance samples, as done for the other WEP rules used here. Fig. 6(b)-(h) show the resulting fusion posterior GM pdfs at $k = 65$ for robot 1 (bottom left of map). Fig. 7 shows the KLD information gains (6) for each robot over time under each fusion scheme.

Fig. 6(b) and Fig. 7(a) show that robot 1 contributes the least amount of new information, as it remains completely still at the bottom left of the map throughout the search . When using DDF to update its local copy of the x_{TARG} GM, robot 1 exchanges GMs with robots 2 and 3 without any of the robots being aware of the common information in the network. The Naive Bayes DDF results in Fig. 6(d) and Fig. 7(c) show that repeated double-counting of common information drives the final DDF posterior pdfs for all robots very far from the centralized fusion posterior in Fig. 6(c) and thus leads to severely inconsistent results.

In contrast, the final WEP DDF posterior GMs in Fig. 6(e)-(h) and information gains in Fig. 7(d)-(f) all bear much closer resemblance to the centralized fusion result (with KLDs to the centralized result roughly the same between 0.4-0.55 nats). The posterior GM plots for Information Weighted Chernoff and Minimum Information Loss fusion show that these rules are better at retaining some of the subtler features of the centralized robot fusion GM pdf, such as the distinctness of the modes around the outskirts of the map and the 'plowed lanes' created by robots 2 and 3. On the other hand, the posterior GMs for unweighted Chernoff and Bhattacharyya fusion are smeared out, indicating that less novel information is retained by WEP fusion at the expense of robot 1's older outdated common prior information (note the frequent information 'dips' for unweighted Chernoff). Interestingly, Fig. 7(f) shows that each robot is more likely to gain (and less likely to lose) information via Information Weighted Chernoff fusion, as the gains for each robot rarely drop from their previous values, as they do near the end of the unweighted Chernoff and Minumum Information Loss results. Further analysis shows that this behavior for Information Weighted Chernoff fusion stems from an aggressive 'winner takes all' strategy for selecting ω in the initial stages of the search, so that values of ω close to 0 or 1 are observed frequently. As the robot information gains become similar after $k = 30$, the ω values selected by Information Weighted Chernoff become less extreme. This suggests different approaches are more appropriate at different times during the search process.

V. Conclusion

The WEP approach for robust fusing two arbitrary probability distributions with unknown common information was addressed for general multi-robot applications. The nominal Chernoff fusion rule does not account for differences in content or quality of information sources used to construct constituent robot distributions; the Information Weighted Chernoff Fusion rule was developed to account for such potential disparities without explicitly counting the number of measurements. To approximately minimize losses incurred by WEP fusion relative to the exact fusion posterior, a second novel Minimum Information Loss Fusion rule was also developed using the KLD between the Naive Bayes and WEP-fused distributions as the cost metric. Novel fast convex implementations of the proposed WEP fusion rules were presented for fusion of Bernoulli and arbitrary Gaussian mixture pdfs. These methods were experimentally demonstrated in loopy multi-robot communication topologies for fusion of 3D occupancy grids and 2D target location beliefs. Future work will investigate the theoretical properties and online performance of the proposed WEP methods across a variety of dynamic ad hoc fusion scenarios, spanning scales from dozens to thousands of robots.

References

[1] T. Bailey, S. Julier, and G. Agamennoni. On Conservative Fusion of Information with Unknown Non-Gaussian De-
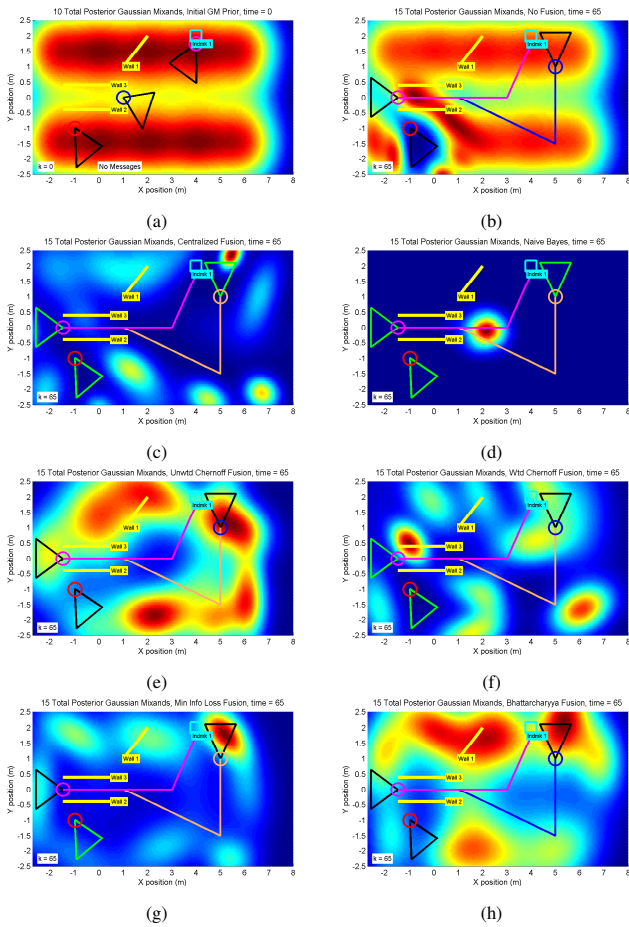
Fig. 6. (a) Prior GM pdf for x_{TARG} at $k = 0$; (b)-(h) Robot 1's (bottom left) GM posterior pdf at $k = 65$ under various fusion rules: (b) no fusion, (c) centralized fusion, (d) Naive Bayes, (e) unweighted Chernoff, (f) Information Weighted Chernoff, (g) Minimum Information Loss, (h) Bhattacharyya fusion.
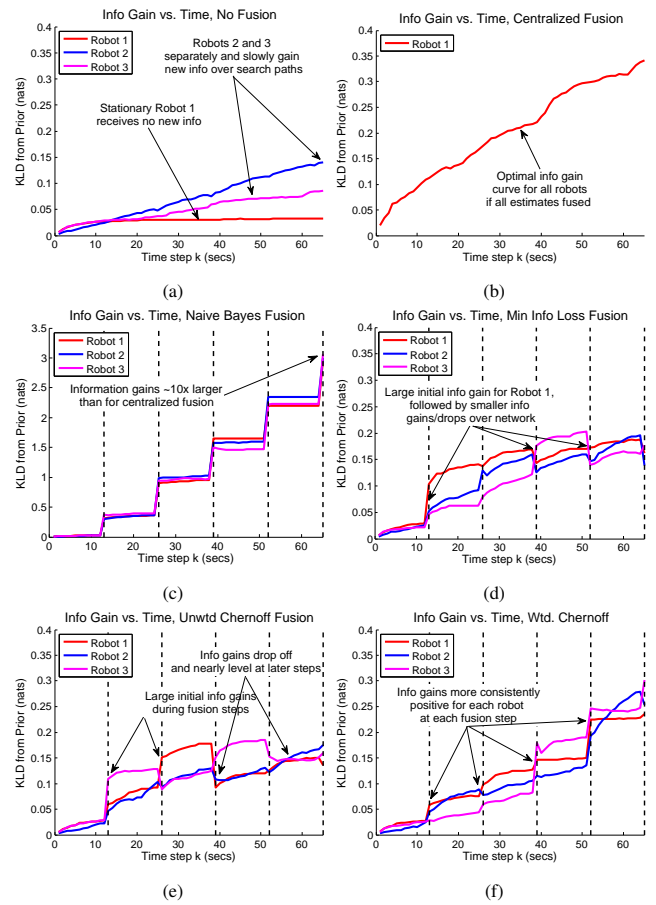

Fig. 7. Robot information gains vs. time relative to $k = 0$ GM prior under different fusion rules: (a) no fusion, (b) centralized, (c) Naive Bayes (note larger vertical scale) (d) Minimum Information Loss, (e) unweighted Chernoff, (f) Information Weighted Chernoff. Results for Bhattacharyya fusion (not shown) are similar to unweighted Chernoff results. Dashed vertical lines denote interagent DDF time steps.

pendence. Technical report, Australian Center for Field Robotics, 2011.

[2] F. Bourgault. *Decentralized control in a Bayesian World*. PhD thesis, University of Sydney, 2005.

[3] K. Chang, Chong Chee-Yee, and S. Mori. Analytical and Computational Evaluation of Scalable Distributed Fusion Algorithms. *IEEE Trans. on Aerospace and Electronic Systems*, 46(4):2022–2034, 2010.

[4] T. M. Cover and J. A. Thomas. *Elements of information theory*, volume 6. Wiley Online Library, 1991.

[5] Anthony H. Dekker and Bernard D. Colbert. Network robustness and graph topology, 2004.

[6] W. J. Farrell and C. Ganesh. Generalized chernoff fusion approximation for practical distributed data fusion. In *12th Int'l Conf. on Information Fusion.*, pages 555–562, 2009.

[7] J. Goldberger, H. K. Greenspan, and J. Dreyfuss. Simplifying Mixture Models Using the Unscented Transform. *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, 30(8):1496–1502, 2008.

[8] S. Grime and H. F. Durrant-Whyte. Data fusion in decentralized sensor networks. *Control Engineering Practice*, 2(5):849–863, 1994.

[9] M. B. Hurley. An Information-Theoretic Justification for Covariance Intersection and Its Generalization. Technical report, MIT Lincoln Lab, 2001.

[10] S. Julier. An Empirical Study into the Use of Chernoff Information for Robust, Distributed Fusion of Gaussian Mixture Models. In *9th Int'l Conf. on Information Fusion.*, 2006.

[11] S. J. Julier and J. K. Uhlmann. A non-divergent estimation algorithm in the presence of unknown correlations. In *Proc. of the American Control Conference.*, volume 4, pages 2369–2373 vol.4, 1997.

[12] J. H. Kotecha and P. M. Djuric. Gaussian sum particle filtering. *Signal Processing, IEEE Trans. on*, 51(10): 2602–2612, 2003. 1053-587X.

[13] Andrew R. Runnalls. Kullback-Leibler Approach to Gaussian Mixture Reduction. *IEEE Trans. on Aerospace and Electronic Systems.*, 43(3):989–999, 2007.

[14] K. M. Wurm, A. Hornung, M. Bennewitz, C. Stachniss, and W. Burgard. OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems. In *Proc. of the ICRA 2010 workshop on best practice in 3D perception and modeling for mobile manipulation*, 2010.