Random set tracking and entropy based control applied to distributed sensor networks

David Stein, James Witkoskie, Stephen Theophanis, Walter Kuklinski

The MITRE Corporation, 202 Bedford Road, Burlington, MA 01730

ABSTRACT

This paper describes an integrated approach to sensor fusion and resource management applicable to sensor networks. The sensor fusion and tracking algorithm is based on the theory of random sets. Tracking is herein considered to be the estimation of parameters in a state space such that for a given target certain components, e.g., position and velocity, are time varying and other components, e.g., identifying features, are stationary. The fusion algorithm provides at each time step the posterior probability density function, known as the global density, on the state space, and the control algorithm identifies the set of sensors that should be used at the next time step in order to minimize, subject to constraints, an approximation of the expected entropy of the global density. The random set approach to target tracking models association ambiguity by statistically weighing all possible hypotheses and associations. Computational complexity is managed by approximating the posterior Global Density using a Gaussian mixture density and using an approach based on the Kulbach-Leibler metric to limit the number of components in the Gaussian mixture representation. A closed form approximation of the expected entropy of the global density, expressed as a Gaussian mixture density, at the next time step for a given set of proposed measurements is developed. Optimal sensor selection involves a search over subsets of sensors, and the computational complexity of this search is managed by employing the Mobius transformation. Field and simulated data from a sensor network comprised of multiple range radars, and acoustic arrays, that measure angle of arrival, are used to demonstrate the approach to sensor fusion and resource management.

Keywords: Tracking, data association, random sets, sensor network control, distributed sensors.

1. INTRODUCTION

Sensor networks are utilized when a single sensor can not adequately provide all desired information, e.g., target identification and localization may require multiple sensing modalities. Furthermore, kinematic or other variables may be better estimated using multiple sensors. Efficiently employing sensor networks requires methods to jointly utilize the available information to achieve the sensing objectives and methods to select sensor subsets to optimize the information content subject to constraints on energy usage, communications bandwidth, computational capability, or other limited resources[1]. This paper describes an integrated approach to sensor fusion and sensor-network resource management. The fusion algorithm produces probability density functions, conditioned on available measurements, on the state space of parameters of interest, and the resource manager selects the sensors used to obtain the next set of measurements so that the expected entropy of the state space density conditioned on past and next measurements is minimized subject to resource constraints.

Tracking is difficult in environments in which the probability of target-measurement association errors is large or where the motion model must allow for several possible target trajectories. Monitoring vehicles traversing a dense road network presents both challenges. Multiple and varying numbers of vehicles may be in view of a given sensor, and a measurement may be plausibly associated with several targets. Furthermore, at road intersections a target may pursue one of several paths.

Random sets are used to develop a tracking algorithm that allows for both target-measurement association ambiguities and target motion uncertainties as well as false alarms and missed detections. The random set approach, as developed herein, provides estimates of parameter values at each time step; it does not explicitly define tracks. It avoids the association ambiguity by statistically weighing all possible hypotheses and associations[2,3,4]. It also incorporates all

possible target trajectories. By Kronecker producting the kinematic space with a feature space, the RST incorporates ID into tracking problems without modifying the framework. To control computational complexity, probability density functions on measurement and state space are approximated using Gaussian mixture models. The number of components of the mixture densities are limited and mixture components are selected using a goodness-of-fit criterion.

The present paper equates the resource optimization problem with minimizing the average entropy of the state space probability density function subject to constraints. Earlier work[2] used a restless bandit approach to solve the sensor selection problem. To apply this technique, a Markov decision process is defined: monitoring road segments are the tasks; states are defined as knowledge of vehicular occupancy of task segments; actions are probing the task segments with sensors; rewards are generated for transitioning from one state to another and transition probabilities depend upon whether sensors were or were not employed. The optimal action given the current state provides the maximum expected infinite time horizon discounted reward, and it is found by solving a linear programming problem[2] . Resource optimization depends only on the current state, and as there are finitely many, the optimal action for a given state can be precomputed and stored in a lookup table. The present approach based on minimizing expected entropy obviates the need for a Markov structure and action dependent transition probabilities. Expected reduction in approximate entropy is calculated directly, but cannot be precomputed. Mutual information is the criterion used to select a single sensor at each time instant in [5,6,7], which consider a single target moving through a two dimensional sensor field. Mutual information is computed in these papers[5,6,7] by discretizing the state and measurement spaces. This approach is not applicable to the present problem of tracking a varying number of targets using potentially multiple simultaneous measurements. Information theoretic control using particle filters is described[8].

The paper is organized as follows. Section two describes the random set tracker and its implementation with Gaussian mixture density functions. Relationships between the RST and the multiple hypotheses tracker (MHT) and a joint integrated probabilistic data association tracker (JIPDA) are demonstrated. Section three describes the entropy based resource management algorithm and approximations used in its implementation. Section four describes the application of the integrated tracking and resource management algorithms to the problem of monitoring vehicles moving through a network of roads.

2. RANDOM SET THEORY TRACKING (RST)

The random set tracker is defined. Approximations used to control the computational complexity are described, and the random set tracker is compared to the MHT and the JIPDA.

2.1 Description of the random set tracker

Traditional Kalman filters assume state variables, x , and measurements, y , are fixed length random vectors[3]. For traditional tracking applications, x represents the targets' geokinetic variables, and y represents the measurements related to the geokinetic variables. The state vector motion model is linear with Gaussian white noise, $x_{t+1} = \Omega x_t + d\lambda_t$. Additionally, the measurements depend linearly on the state vectors with additive Gaussian white noise, $y_t = Wx_t + d\delta_t$. The Kalman filter can be extended to non-Gaussian noise, non-linear measurements, and non-linear motion models through the Bayesian filter[3]. The Bayesian filter consists of two steps. 1) Starting with a probability density conditioned on previous measurements, $P(x_{t-1}|y_{1:t-1})$, a prediction step estimates the probability density of the vector at time step t ,

$$P(x_t|y_{1:t-1}) = \int dx_{t-1} P(x_t|x_{t-1}) P(x_{t-1}|y_{1:t-1}). \quad (1)$$

This new probability density is combined with a new set of measurements, y_t , during the update step to determine the new estimate of the probability density,

$$P(x_t|y_{1:t}) = N^{-1} P(y_t|x_t) f(x_t|y_{1:t-1}), \quad (2)$$

where N is a normalization factor. The number of state variables and measurements is fixed and the mapping between the state variables and measurements is explicit. The state variables, x, can be expanded to include ID variables, FVs, such as length or color. In this case, the variables associated with target 1 are $\mathbf{x}^i = \{x^i, v^i, f^i\}$, where x^i, v^i, f^i denote the positions, velocities, and FV components of target i. These feature vectors correspond to characteristics of the vehicles being tracked, which we will assume, in this paper, do not vary with time. These FV variables allow us to discriminate targets that cannot be separated by kinematics alone.

Ignoring the computational difficulties, multiple target tracking presents additional difficulties that cannot be accounted for in the normal Bayesian framework because the association of measurements with existing targets is often ambiguous[3,4]. Even if the state variables are linearly related to measurements with Gaussian errors, the ambiguity in associations produces non-Gaussian effects[9]. Missed detections, false alarms or clutter, and the birth and death of targets complicate the scenario. In addition to their numerical values, the number of targets and measurements are also random variables, and the vectors *x* and *y* must be replaced with random sets $\{x\}$ and $\{y\}$. The state variables, $\{x\}$, may take on the values $\{x\} = \{\phi\}$, $\{x^1, v^1, f^1\}$, $\{x^1, v^1, f^1, x^2, v^2, f^2\}, \ldots$ where x^i, v^i, f^i denote the positions, velocities, and FV components of target i, respectively. The set of measurements, $\{y\}$, are estimates of the geokinetic and feature variables recorded from various sensors and include clutter returns and missed detections.

Several methods address variable numbers of targets and detections with ambiguous associations, including joint integrated probabilistic data association (JIPDA) and jump Markov models (JMM)[10,11]. In this paper, we explore applications of finite set statistics (FISST) to road constrained multiple target tracking. FISST is a generalization of the Bayesian equations to sets, equations (1) and (2). The probability density, called the global density, is defined on the possible number and locations of targets. For road networks, the global density has the form

$$P_g = \begin{cases} P_0 \\ P_r(x,v,f) \overset{\Delta}{=} P_r(\{x^{(1)}\}) \\ P_{r_1 r_2}(x^1, v^1, f^1, x^2, v^2, f^2) \overset{\Delta}{=} P_{r_1 r_2}(\{x^{(2)}\}) \\ \vdots \end{cases}, \tag{3}$$

where P_0 is the probability of no targets, $P_{r_1 \ldots r_n}$ is the probability density for n targets on roads $r_1 \ldots r_n$ at $\{x^1, v^1, f^1, \ldots, x^n, v^n, f^n\} = \{x^{(n)}\}$, respectively. The density is normalized with respect to a set integral

$$\sum_{n=0}^{\infty} \frac{1}{n!} \int P_{r_1 \ldots r_n}(x^1, v^1, f^1 \ldots, x^n, v^n, f^n) dx^1 dv^1 df^1 \ldots dx^n dv^n df^n$$
$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int P_{r_1 \ldots r_n}(\{x^{(n)}\}) d\{x^{(n)}\} = 1, \tag{4}$$

where the summation over n refers to the number and identity of possible roads. Similar to the Bayesian filter one can define conditional probabilities, motion models, and measurement models to develop a set of recursive update equations[4,9]. The prediction step includes propagation and birth and death processes so that the conditional expectation of the global density on previous measurements has the form

$$P_{r_1 \ldots r_n}(\{x_t^{(n)}\} \mid \{y_{1:t-1}\}) =$$
$$\sum_{m=0}^{\infty} \frac{1}{m!} \int d\{x_{t-1}\} P_{r_1 \ldots r_n | r_1' \ldots r_m'}(\{x_t^{(n)}\} \mid \{x_{t-1}^{(m)}\}) \, P_{r_1' \ldots r_m'}(\{x_{t-1}^{(m)}\} \mid \{y_{1:t-1}\}) \tag{5}$$

where the sum is over all possible roads, $r_1' \ldots r_m'$. The predicted density is then updated with the measurements at time step t,

$$P_{r_1 \ldots r_n}(\{x_t^{(n)}\} \mid \{y_{1:t}\}) = N^{-1} P_{r_1 \ldots r_n}(\{y_t^{(n)}\} \mid \{x_t^{(n)}\}) \times P_{r_1 \ldots r_n}(\{x_{t-1}^{(n)}\} \mid \{y_{1:t-1}\}). \quad (6)$$

Summation over the associations makes these expressions more complicated than equations (1) and (2). Similar to the Bayesian filter, the FISST tracker can be easily extended to applications with target ID by taking the Kronecker product of the FV space with the kinematic space. The FV extension of this previously proposed tracker handles the ambiguities in FV space in a manner that is analogous to the kinematic trackers handling of ambiguities. The prediction and measurement models used in the calculation of the update equations (6)-(7) are described in detail in[12].

2.2 Numerical implementation

A Gaussian Mixture approximation is used to make the random set implementation numerically feasible. The Gaussian mixture model for this road network scenario also allows a scalable Gaussian sum particle filter (GSPF) representation[13]. The GSPF is similar to particle filter sampling methods, but the delta function kernel associated with each particle is replaced by a variable dimensional Gaussian. The covariance and mean of each particle is propagated instead of simply the position of each particle. Each term in the global density is represented by a finite number of Gaussians,

$$P_{r_1 \ldots r_n}(\{x^{(n)}\}) = \sum_i a_i N(x^{(n)} - \mu_i^{(n)}, \Gamma_i^{(n)}) \quad (7)$$

where N is a multidimensional Gaussian distribution, $N(\mu, \Gamma; x) = \frac{1}{|2\pi\Gamma|^{0.5}} \exp\left(-\frac{1}{2}(x-\mu)^t \Gamma^{-1}(x-\mu)\right)$ which may include the FV. The Gaussian mixture representation requires several approximations: 1) After combining detections with the Gaussian, the probability of detect is based on the mean value of each Gaussian component. 2) The non-linear maps between the coordinates and the measurements are Taylor expanded around the mean values, $g \approx g(x(s_0)) + (\partial g(x(s_0))/\partial s)(s - s_0)$ in directions parallel and transverse to the road network—each road is parameterized by arclength, s, from an initial point s_0. 3) The motion model is still linear. 4) Instead of truncating overhanging density, road switching and death processes correspond to the mean of the mixture component overhanging the end of the road. The entire Gaussian component either switches roads or is marginalized. 5) Birth processes correspond to adding mixture components to the ends of roads with a mean velocity pointing onto the roads and a fixed initial width.

All of these approximations maintain the Gaussian mixture representation of the global density. The validity of these approximations depends on the Gaussian mixture components' variances being much smaller than variations in the detection probabilities, the variations in h, and the distance from the ends of the roads. We ensure the validity of these approximations by replacing components with large variances with several Gaussian components of smaller variances through a Kullback-Leibler measure. Similarly, if a target is on the boundary of a classification, it is possible to split up a Gaussian into multiple Gaussian mixtures that do not straddle the boundary, but we have not implemented this approach[13].

Although these approximations ensure that the global density maintains a Gaussian mixture functional form, the associations of the measurements result in a geometric explosion in the number of mixture components. To avoid the geometric growth, the mixture components are recombined so that the original target distribution $P_{r_1 \ldots r_n}$ is replaces with a new Gaussian mixture distribution, $\tilde{P}_{r_1 \ldots r_n}$, that contains two components with the same mean, variance[14],

$$\tilde{P}_{r_1 \ldots r_n}(\{x^{(n)}\}) = \sum_k b_k N(x^{(n)} - m_k^{(n)}, M_k^{(n)}) = \sum_{k=i,j} a_k N(x^{(n)} - \tilde{\mu}_k^{(n)}, \tilde{C}_k^{(n)}) + \sum_{k \neq i,j} a_k N(x^{(n)} - \mu_k^{(n)}, C_k^{(n)}). \quad (8)$$

where $\tilde{\mu}^{(n)}$ and $\tilde{C}^{(n)}$ are the optimal mean and covariance, respectively. The optimal mean and covariance corresponds to the optimal mean and covariance of the distribution $(a_i + a_j)^{-1} \sum_{k=i,j} a_k N(x^{(n)} - \tilde{\mu}_k^{(n)}, \tilde{C}_k^{(n)})$. The recombination is

based on the sensor system's ability to distinguish the true distribution, $P_{r_1 \ldots r_n}$, from the proposed distribution, $\tilde{P}_{r_1 \ldots r_n}$, given that the sensor measurements are derived from the true distribution. If the original and proposed distributions cannot be statistically distinguished by the possible sensor measurements sampled from the true distribution, the proposed distribution is accepted since it has less parameters (Occam's razor).

The KL metric is approximated by a pairwise K-L metric

$$D_{P|\tilde{P}} \approx \sum_k \int d\{x^{(n)}\} a_k N(x^{(n)} - \mu_k^{(n)}, C_k^{(n)}) \ln\left(\frac{a_k N(x^{(n)} - \mu_k^{(n)}, C_k^{(n)})}{b_k N(x^{(n)} - m_k^{(n)}, M_k^{(n)})}\right)$$
$$= \sum_{k=i,j} \int d\{x^{(n)}\} a_k \left[N(x^{(n)} - \mu_k^{(n)}, C_k^{(n)}) \ln\left(\frac{N(x^{(n)} - \mu_k^{(n)}, C_k^{(n)})}{N(x^{(n)} - \tilde{\mu}_k^{(n)}, \tilde{C}_k^{(n)})}\right)\right]. \quad (9)$$

The pair of mixture components that result in the smallest KL metric are combined until either the number of components is smaller than a set maximum number or the KL metric for combining two mixtures is greater than a tolerance[14]. Other criteria for combining classes have also been explored, including the application of a threshold test to the probability that the log-likelihood ratio of the affected classes exceeds a value[12]. The procedure results in a reduction in the number of mixture components, as demonstrated in Figure 1. This technique may be extended to include feature vectors as well[12].

Large scale applications can take advantage of the global density's structure to further reduce the computational burden. For example, the targets might be segregated into groups such that targets from different groups are not confused by any of the sensors and the estimates for each group of targets could be done independently of the other groups. These ideas are discussed in more details in[12].

2.3 Comparison with other techniques

The random set tracker proposed in this paper may be compared with other techniques. The linearization in measurement models is identical to the extended Kalman filter. In fact, if there were no association ambiguities, (i.e. we knew which target creates each detection), the resulting algorithm would be an extended Kalman filter. Similarly, if we always reduced the distribution to a single Gaussian, a tracker similar to the JIPDA filter would result[10]. However, the random set tracker sums over associations, which avoids hard decisions made in many traditional Kalman Filter based approaches. A MHT attempts to track several possible hard associations, but introduces a pruning procedure to reduce the number of possible hypotheses. Unlike MHT, the random set tracker avoids the exponential branching of hypotheses by reducing information (not the explicit number of hypotheses). The number of hypotheses is allowed to vary as needed to maintain the appropriate level of information about the targets. Figure 1 gives an illustration of the differences between MHT and RST. Figure 1 (a) shows a path where a confuser vehicle will become confused with the VOI track at two points in time (1 and 3). The multiple hypothesis tracker will have two branching processes, resulting in four hypotheses. If the number of hypotheses maintained by the tracker is less than four, some of these hypotheses will be eliminated despite that the probability of each hypothesis may be close to equally likely. The RST tracker would prevent this branching and pruning by determining that the hypotheses at time point 3 contain similar information and can be approximated by a single Gaussian with little loss of information. This reduces the number of hypotheses handled by the tracker. When the second branch occurs, only two hypotheses result.
The RST fusion algorithm lacks the traditional concept of a track, which is a contiguous set of position estimates over time believed to be the result of a single vehicle. Internally, the RST fusion algorithm simply reports the probability of a set of target with specific FVs and geokinetic variables at a specific time point. A continuous track can be inferred from a MAP estimate performed on the global density, but this track is a refined data product that cannot be used internally by the fusion algorithm.

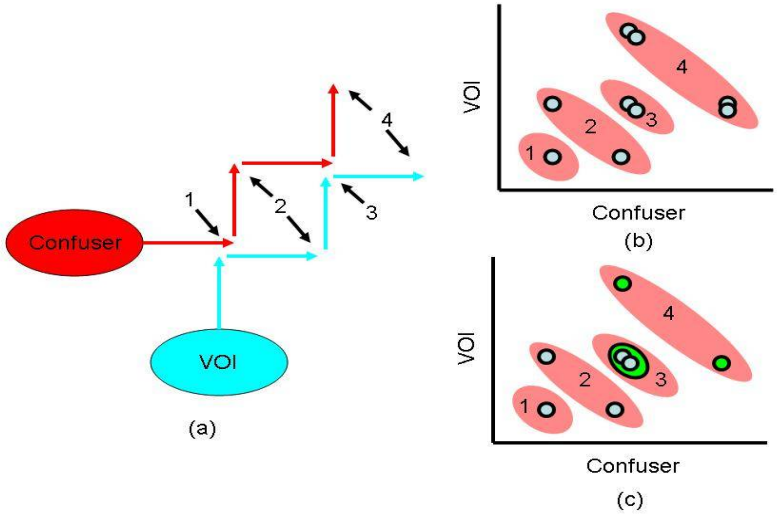


Figure 1: (a) A path where the VOI and confuser become entangled at points 1 and 3. If no ID measurements can be performed at point 2, a MHT will give 4 possible hypotheses at point 4 as shown in (b). The random set tracker recombines the two hypotheses at point 3 so that only two hypotheses exist at point 4 (c).

3. ENTROPY BASED RESOURCE MANAGEMENT

Resource management is essential to the efficient utilization of surveillance assets when all areas of interest cannot be surveyed simultaneously either due to limitations of the available sensing assets or limitations on energy, computational capability, communications bandwidth, or other resources. Dynamic resource management determines sensor utilization based on the current representation of an evolving situation. At each time instant, the global density function, defined above, represents current knowledge, and as described below, it provides the foundation for an entropy based resource management algorithm that minimizes average entropy of the global density subject to resource constraints.

3.1 Derivation of the entropy based resource manager

Assume that sensors are employed and that y_t denotes the measurement vector at time *t*. The state space model includes a measurement model, $g_t(y_t \mid x_t)$,where g_t is the probability density on the measurements conditioned on the state space variate. Note that the measurement density further depends on the target-sensor relations, $\alpha \in \mathrm{A}$, i.e.,

$$g_t(y_t \mid x_t) = \sum_{\alpha \in \mathrm{A}} g_t(y_t \mid x_t, \alpha) p(\alpha) \tag{10}$$

The state space model is completed by specifying a state space propagation model, i.e. the probability distribution function on state space at time *t+1* conditioned on the variate at time *t:* $f_{t+1}(x_{t+1} \mid x_t)$. In this context optimal control is the selection of sensors to optimize an objective function subject to resource constraints. For the present work sensors are chosen to minimize the expected entropy of the state space distribution subject to a resource constraint on each sensor or a total resource constraint.

Let $$H(x_t \mid y_{0:t}) = \int f(x_t \mid y_{0:t}) \ln\left(f(x_t \mid y_{0:t})\right) dx_t \tag{11}$$

be the entropy in the state space distribution conditioned on the measurements up to time t. The increment in entropy over one time step is

$$A\left(X^t \mid Y^{0:t-1}\right) = \int f\left(X^t \mid Y^{0:t-1}\right) \ln\left(f\left(X^t \mid Y^{0:t-1}\right)\right) dX^t - H\left(X^{t-1} \mid Y^{0:t-1}\right), \tag{12}$$

And the reduction in entropy due to the measurements is

$$R\left(x_t \mid y_{0:t}\right) = \int f\left(x_t \mid y_{0:t-1}\right) \ln\left(f\left(x_t \mid y_{0:t-1}\right)\right) dx_t - H\left(x_t \mid y_{0:t}\right). \tag{13}$$

Then

$$H\left(x_t\right) = H\left(x_{t-1}\right) + A\left(x_t\right) - R\left(x_t\right) \tag{14}$$

Total entropy over n time steps is

$$C = \sum_{j=0}^{N} H\left(X^{t_j}\right)$$

Assume that the system objective is the minimization of total expected entropy subject to a resource constraint. Further assume that the amount of entropy removed at t_j is a function $R\left(x_{t_j}\right) = R_j\left(l_1^{\,j}, \ldots, l_s^{\,j}\right)$, where $l_i^{\,j}$ is the resource committed to sensor i at time j. Lagrange multipliers may be used[16] to show that the optimum strategy, assuming that the increment in entropy $A\left(x_t\right)$ is independent of past measurements is the solution of

$$\left(N - j + 1\right) \sum_{i=1}^{s} \frac{\partial ER_j}{\partial l_i^{\,j}}\left(l_1^{\,j}, \ldots, l_s^{\,j}\right) = \lambda, \tag{15}$$

Note that *E(R)* is the mutual information, between the state variable conditioned on past measurements and current measurements[16].

3.2 Implementation with the RST

The PDf in state-space is represented as a Gaussian mixture

$$f\left(x_{t+1} \mid y_{0:t}\right) = \sum_{j=1}^{n_{t+1}} w_j^{\,t+1} N\left(\mu_j^{\,t+1|t}, \Gamma_j^{\,t+1|t}; x_{t+1}\right). \tag{16}$$

For each set, *S*, of sensors and each component, *j*, of the state space mixture define a set of target–sensor relations, $\mathrm{A}(S, j)$. The probability, $p(j, \alpha)$, of any target-sensor relation, $\alpha \in \mathrm{A}(S, j)$, is determined by the target-sensor geometry and characteristics of the background and sensors. Using a linearization of the measurement model about the mean of each mixture component (see above) , the measurement space density may be expressed as

$$g_S\left(y_{t+1} \mid y_{0:t}\right) = \sum_{j=1}^{n_{t+1}} \sum_{\alpha \in \mathrm{A}(S,j)} p(j, \alpha) N\left(\hat{\mu}_{j\alpha}^{\,t+1|t}, \hat{\Gamma}_{j\alpha}^{\,t+1|t}; y_{t+1}\right), \tag{17}$$

where $\hat{\mu}_{j\alpha}^{\,t+1|t}$ and $\hat{\Gamma}_{j\alpha}^{\,t+1|t}$ are obtained by linearizing the measurement model about $\mu_j^{\,t+1|t}$. The density in state space conditioned on the sensor set *S* and measurements Y^{t+1} is approximated by

$$f_S\left(x_{t+1} \mid y_{0:t}, y_{t+1}\right) = \sum_{j=1}^{n_{t+1}} \sum_{\alpha \in \mathrm{A}(S,j)} p_{j\alpha}\left(y_{t+1}\right) N\left(\tilde{\mu}_{j\alpha}^{\,t+1|t}\left(y_{t+1}\right), \tilde{\Gamma}_{j\alpha}^{\,t+1|t}; x^{t+1}\right) \tag{18}$$

$$\approx \sum_{j=1}^{n_{t+1}} \sum_{\alpha \in (S,j)} p_{j\alpha}\left(y_{t+1}\right) N\left(\mu_j^{\,t+1|t}, \tilde{\Gamma}_{j\alpha}^{\,t+1|t}; x_{t+1}\right) \tag{19}$$

$$\approx \sum_{j=1}^{n_{t+1}} p_j(y_{t+1}) N\left(\mu_j^{t+1|t}, \overline{\Gamma}_j^{t+1|t}; x_{t+1}\right) \qquad (20)$$

$$\equiv \tilde{f}\left(x_{t+1}; y_{0:t}, S, y_{t+1}\right) \qquad (21)$$

where $\tilde{\mu}_{j\alpha}{}^{t+1|t}$ and $\tilde{\Gamma}_{j\alpha}{}^{t+1|t}$ are the Kalman filter updates of $\mu_j^{t+1|t}$, and $\Gamma_j^{t+1|t}$, respectively[17]. Note that $\tilde{\Gamma}_{j\alpha}{}^{t+1|t}$ is independent of the measurements. In going from (19) to (20), $\tilde{\mu}_{j\alpha}{}^{t+1|t}(Y^{t+1})$ is replaced with $\mu_j^{t+1|t}$, and on going from (20) to (21) the mixture terms for a given j are combined and approximated by the given normal.

3.3 Approximations

The entropy of a normal mixture distribution is replaced with the following upper bound on the entropy.

If $f(x) = \sum_{j=1}^{n} \omega_j N(\mu_j, \Gamma_j; x)$ is a normal mixture density function, then

$$H(x) = -\int f(x) \ln(f(x)) dx$$

$$\leq -\sum_{j=1}^{n} \omega_j \ln(\omega_j) - \sum_{j=1}^{n} \int N(\mu_j, \Gamma_j; x) \ln(N(\mu_j, \Gamma_j; x)) dx$$

$$= -\sum_{j=1}^{n} \omega_j \ln(\omega_j) + \frac{n}{2}(1 + \ln(2\pi)) + \sum_{j=1}^{n} \frac{\omega_j}{2} \ln(|\Gamma_j|) \qquad (22)$$

$$= \tilde{H}(x). \qquad (23)$$

$\tilde{H}(x)$ is the entropy of the class weights, ω_j, plus the weighted entropy of the classes. Note that

$$\tilde{H}(x) = H((x, j)), \qquad (24)$$

which might be called the complete entropy.

The mutual information is

$$\tilde{I}\left(x_{t+1}, y_{St+1}\right) = \tilde{H}\left(x_{t+1} \mid y_{0t}\right) - E_{Y_S t+1}\left(\tilde{H}(x_{t+1} \mid y_{0t}, y_{St+1})\right). \qquad (25)$$

The mutual information (23) is shown to be[16]

$$\tilde{I}\left(x_{t+1}, y_{St+1}\right) = \sum_{j=1}^{n} w_j^{t+1} H\left(N\left(\mu_j^{t+1:t}, \Gamma_j^{t+1:t}\right)\right) - \sum_{j=1}^{n} w_j^{t+1} H\left(N\left(\mu_j^{t+1:t}, \overline{\Gamma}_j^{t+1:t}\right)\right)$$

$$+ \sum_{j=1}^{n_t} w_j^{t+1} \int p(y_{St} \mid j) \ln\left(\frac{p(y_{St} \mid j)}{p(y_{St})}\right) dy_S. \qquad (26)$$

Note that the difference of the sum of weighted normal entropies in (26) captures the reduction in the class covariance due to the measurements, and the remaining term captures the reduction in the class ambiguities.

A key step in the application of entropy based resource management is the approximation of this integral. Results obtained using Monte Carlo integration may be compared with results obtained using another method in order to judge

the efficacy of the proposed technique. Alternatively, quadratic mutual information, which for Gaussian mixture distributions can be computed in closed form may be used as a measure of sensor efficacy[16]. The present paper proposes another approximation in terms of classification error bounds.

$$\sum_{j=1}^{n_t} w_j^{t+1} \int p(y_{St} \mid j) \ln\left(\frac{p(Y_{St} \mid j)}{p(Y_{St})} \right) dY_S \approx 2 \sum_j \sum_{k>j} \left(w_j^{t+1} w_k^{t+1} \right)^{0.5} \sum_{r,s} \left(\alpha_{jr} \alpha_{ks} \right)^{0.5} B\left((\mu_{jr}, \Gamma_{jr}), (\mu_{ks}, \Gamma_{ks}) \right), \quad (27)$$

where $p(y \mid j) = \sum_r \alpha_{jr} N(\mu_{jr}, \Gamma_{jr}; Y)$ and $B(N(\mu, \Gamma), (\nu, \Sigma))$ is the Bhattacharyya bound on the maximum likelihood classifier assuming normally distributed classes. This result is derived and results obtained using various approximation techniques are compared[16].

The optimization criteria (15) is implemented by determining the set of sensors that maximizes the mutual information for a given level of resource utilization and approximating the derivative with the corresponding difference equation. The resource level and hence the corresponding set of sensors is determined by the threshold criterion (15). This approach requires a search over subsets of sensors. The Mobius transformation[18] is employed to reduce the computational complexity of the search. Following[18], suppose that Θ is a finite set and that f and g are functions on 2^Θ the set of subsets of Θ. Then if

$$f(A) = \sum_{B \subset A} g(B) \text{ for all } A \subset \Theta, \text{ and } g(A) = \sum_{B \subset A} (-1)^{|A-B|} f(B) \text{ for all } A \subset \Theta, \quad (28)$$

g is the Mobius transform of f. To reduce the computational complexity of optimizing a function over a discreet set, select $k > 0$ and define

$$\tilde{g}(A) = \begin{cases} g(A) \; if \; |A| \le k \\ -\sum_{|A|-|B|=1} \tilde{g}(B), \; otherwise \end{cases} \quad (29)$$

Define $$\tilde{f}(A) = \sum_{B \subset A} \tilde{g}(B). \quad (30)$$

The extrapolations of the Mobius transform

$$\tilde{g}(A) = \begin{cases} g(A) \; if \; |A| \le k \\ 0, \; otherwise \end{cases} \quad (31)$$

has been applied to feature selection[19]. These methods of extrapolating the Mobius function for the present application are compared[16].

4. APPLICATION TO ROAD CONSTRAINED TRACKING

MITRE's Netted Sensor Program has deployed networks of radar, acoustic and imaging sensors and developed a simulation facility that is used to demonstrate the capabilities of sensors and algorithms. The range radars were built by Multispectral Solutions, Inc. (Germantown, MD) and have a detection range of 512 ft (156 m) and a 24 degree beam pattern. Detections are recorded by determining background clutter statistics and detecting statistically significant blocks of range cells that exceed the background clutter. The detection probabilities are calculated from the signal, S, using the Swerling I model. Each acoustic array consist of four microphones spaced 8 inches apart in a square planar configuration. Angles of arrival are estimated in the time domain from the microphone cross-correlation functions. Angles are reported when the sound level exceeds a threshold, T, defined relative to the background. Experiments show

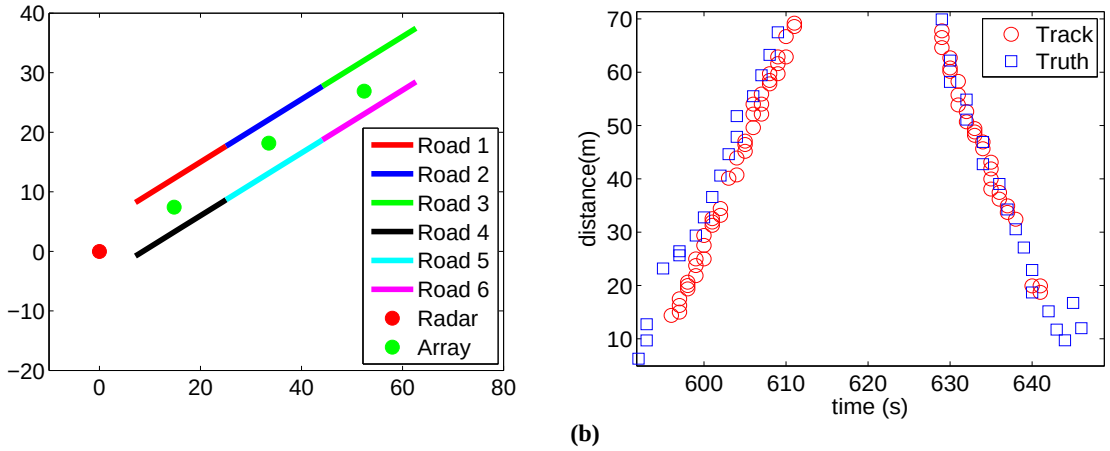
a normally distributed error with 5^o accuracy and a detection range of 10-15 meters. The estimated bearing angle generally points to the loudest (and often closest) vehicle, although more complicated behaviors occasionally occur.

4.1 Tracking

The methods described above were applied to the problem of tracking vehicles traveling on a road without intersections as depicted in Figure 2(a). One range radar and three acoustic arrays were utilized, and they were positioned as shown. Ground truth was obtained by instrumenting the target vehicle with a GPS. This vehicle traversed the road while targets of opportunity enter the road network and act as decoys. The sensors and the FISST tracker were run at 3 Hz.

The FISST global density reported tracks determined by: 1) determining the most likely hypothesis through marginalizing over the geokinetic variables, $P_{r_1 \ldots r_n} = \frac{1}{n!} \sum \int d\{x^{(n)}\} f_{r_1 \ldots r_n}(\{x^{(n)}\})$, where the summation is performed over the permutations of $r_1 \ldots r_n$. 2) The most likely positions of the most likely hypothesis (the MAP estimate) are reported, $\arg\max_{\{x^{(n)}\}} f_{r_1 \ldots r_n}(\{x^{(n)}\})$.

Figure 2(b) compares a typical track on the linear road network for the fusion algorithm with all sensors collecting data at every time step against a typical track for the fusion algorithm. The measured distance corresponds to the distance from the radar to the target or track.



(a) **(b)**

Figure 2: (a) The straight line road network geometry. The axes are topocentric with distances in meters. (b) Tracking on the linear road network in (a) with all sensors running at 3 Hertz. Distances are measured in meters from the radar. Errors in estimated positions vary with time and the measurements but have an average value of approximately 4m. Truth error estimates average approximately 3m and may be systematic, due to GPS.

4.2 Resource Management

Simulation studies were performed to evaluate the resource manager. The simulator allows for an arbitrary road network instrumented with an arbitrary number and placement of range radars and acoustic arrays. Whereas the radar can simultaneously detect multiple targets, each in different range bins, the acoustic sensor, as described above, can only detect one target in any given measurement. For each sensor type, the simulator incorporates sensor-target geometry dependent models of the accuracy and the probabilities of detection and models of the probabilities of false alarms. As the acoustic sensor reports only one, generally the loudest, detection, its probability of detecting a target depends upon the totality of targets in the scene, their positions, and their amplitudes. At present the simulator assumes

that the acoustic source levels and radar cross sections of all targets are identical. Targets enter the scene according to the target birth model, travel through the scene according to the motion model and exit according to the target death model. At each time step, given the configuration of targets, the radar and acoustic models are used to generate synthetic radar and acoustic detections. These data are used by the tracker, as real data are used, to update the global density. Given the current global density, the resource manager determines what sensor data, or simulated sensor data, to collect at the next time instant. The average number of sensors deployed is determined by the value of the threshold in (15). The resource manager is evaluated by computing average entropy of the global density as a function of the average number of sensors deployed for various target-sensor scenarios.

The resource manager was utilized as follows. The mutual information of a given set of sensors was approximated using (26) and (27). The Mobius transform of the approximate mutual information on subsets up to size 2 was computed and extrapolates to larger sets using (29). Approximate mutual information was then computed using (30). In this work, a sensor was either on or off and all sensors were assumed to consume the same quantity of resources when on, so that the resources devoted to a sensor were either 0—it is off—or 1—it is on. For each level of resource utilization, the subset of sensors having the largest approximate mutual information was determined, and the set of such sensors such that the increment in mutual information from adding the next sensor is below a threshold was selected for the next set of measurements. The threshold was fixed for the duration and not time dependent as in (15).

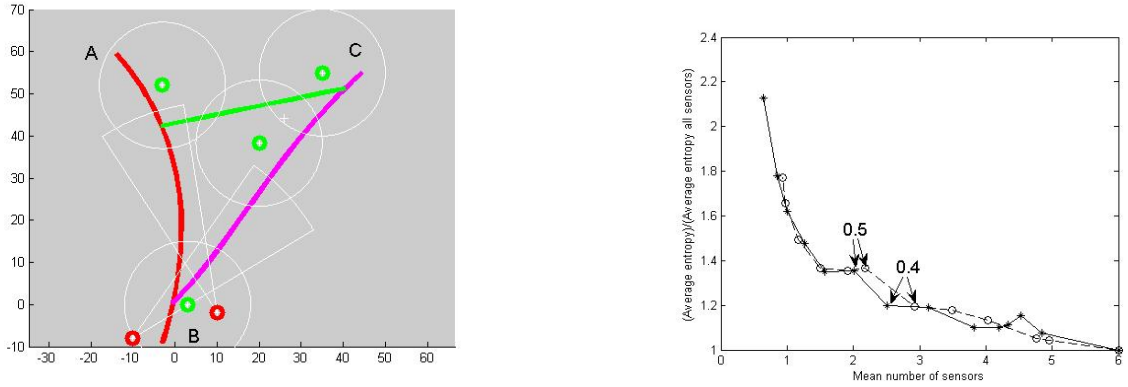


Figure 3. (a) The road network and sensor configuration for the resource management experiments in topocentric coordinates with distance measured in meters. Position of the acoustic arrays and radars indicated in green and red, respectively. Note that the detection range of the radars is 156 m, and the detection range of the acoustic arrays is 10-15 m. (b) Curves showing the average number of sensors deployed and the ratio of the average entropy of the global density using resource management to the average entropy of the global density without resource management for a range of threshold values in equation 15: scenario 1-dashed, and scenario 2-solid. Points obtained at threshold values of 0.4 and 0.5 are indicated.

Simulation results are provided in Figure 3. The sensors were deployed as illustrated in Figure 3a. Two scenarios—the first does not and the second does include a vehicle path-crossing—are considered. In the first scenario, five seconds into the simulation, a vehicle enters the road network at A, takes the direct route to C, and leaves the network; eight seconds into the simulation a second vehicle enters the network at C, takes the direct route to B and leaves the network. In the second case, the first vehicle enters the network at B one second into the simulation, travels to C, travels to A and leaves the network, while the second vehicle enters the network at A at time 5 seconds, travels to C, and leaves the network. The vehicles arrive in the vicinity of intersection C at approximately the same time. Scenario two, is thus the more difficult tracking problem. Note that the vehicles, when in the vicinity of C, may be difficult to distinguish as each report from an acoustic array, as described above, includes at most one detection. Subset selection threshold values ranged from 0.1 to 1.0 for case 1 and 0.0125 to 1.0 for case 2. For each threshold value 10 runs through the network were computed. For each threshold value the average number of sensors selected and the average entropy of the global density were computed with the averages taken over time, as the simulation proceeds and over repeated trials. For each scenario, the average entropy of the global density using all sensors was also calculated. The curves in Figure 3b plot, for the two scenarios and a range of threshold values, the average number of sensors deployed and the ratio of the

average entropy of the global density using resource management to the entropy of the global density obtained using all sensors.

5. SUMMARY AND CONCLUSIONS

An integrated approach to tracking and dynamic resource management that is applicable to complex surveillance applications has been developed. The FISST allows for target birth and death, clutter and missed detections, and multiple possible trajectories that, for example, occur at intersections of roads. A MHT can successfully track branching processes by intelligent pruning of hypotheses[15], but the FISST examined in this paper introduces hypothesis branching reduction through merging similar hypotheses, which avoids information loss from pruning processes. In application to tracking a single object on a road with no intersections, the FISST achieved an average error of 4m relative to GPS locations which had an average error of 3m. Initial applications to the triangle network and other road networks show successful tracking capabilities, which will be quantified in future publications[12]. The application to several experimental scenarios demonstrates that the algorithm does not depend on the specifics of the road network and is extensible.

A resource manager based on minimizing average entropy of the state space PDF was developed and applied to the FISST implemented with Gaussian mixture distributions. Key to implementing the resource manager are the approximations (27) using the Bhattacharyya bounds and the extrapolation of the Mobius transform (29,30). Simulations of two vehicles moving through a network of intersecting roads show that dynamically choosing the sensors with this technique results in approximately a 20% increase in global density entropy while using on average one-third to one half of the sensors at any given time. Consistent results in terms of entropy increase and resource utilization were achieved with a similar threshold (0.4—0.5) applied in equation (15) for two very different scenarios.

REFERENCES

1. Deborah Estrin, Ramesh Govindan, John Heidemann, and Satish Kumar, "Next Century Challenges: Scalable Coordination in Sensor Networks," *Mobicom '99*, Seattle, Washington USA, pp. 263—270.
2. James Witkoskie, Walter Kuklinski, Stephen Theophanis and Michael Otero, "Random Set Tracker Experiment of a Road Constrained Network with Resource Management," Fusion 2006, *Proceedings of the 9th International Conference on Information Fusion*, Florence Italy, 10—13 July 2006.
3. I.R. Goodman, R.P.S. Mahler, and H.T. Nguyen. *Mathematics of Data Fusion*. Kluwer Academic Publishers. Dordrecht, Netherlands. 1997.
4. M.R Morelande and S. Challa. "A Multitarget Tracking Algorithms Based on Random Sets." IEEE Proceedings of the 6th International Conference of Information Fusion. 2:807—814, 2003.
5. Juan Liu, Feng Zhao, and Dragan Petrovic, "Information-Directed Routing in Ad Hoc Sensor Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 23 Vol. 4, 2005, pp. 851—861.
6. Feng Zhao, Juan Liu, Juan Liu, Leonidas Guibas, and James Reich, "Collaborative Signal and Information Processing: An Information-Directed Approach," *Proceedings of the IEEE*, Vol. 91. No. 8, August 2003, pp. 1199—1209.
7. Juan Liu, James Reich, and Feng Zhao, "Collaborative In-Network Processing for Target Tracking," EURASIP Journal on Applied Signal Processing 2003: 4, 378—391.
8. Christophe Andrieu, Arnaud Doucet**,** Sumeetpal S. Singh, and Vladislav B. Tadic. "Particle Methods for Change Detection, System Identification and Control," *Proceedings of the IEEE*, vol. 92, No 3, March 2004, pp. 423—438.
9. C.A. Scott and C.R. Drane. "An Optimal Map-Aided Position Estimator For Tracking Motor Vehicles." *IEEE Proceedings of the 6th International Conference on Vehicle Navigation and Information Systems.* 360--367, Jul. 1995.
10. D. Musicki and R. Evans. "Joint Integrated Probabilistic Data Association." *IEEE transactions on Aerospace and Electronic Systems*. 40(3): 1093—1099, 2004.
11. A. Doucet, B.N. Vo, C. Andrieu, and M. Davy. "Particle Filtering for Multi-Target Tracking and Sensor Management." *Proceedings of the Fifth International Conference on Information Fusion*. 1:474—481, Jul. 2002.
12. James Witkoskie, Walter Kuklinski, David Stein, and Stephen Theophanis, "An Information Theory-Multiple Hypotheses Tracking algorithm based on random set theory," in preparation.
13. J.H. Kotecha and P.M Djuric. "Gaussian Sum Particle Filtering." IEEE Transactions on Signal Processing. 51(10):2602—2612, Oct. 2003

14. W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov. "Reconstruction of elections with the Gaussian-sum filter in CMS tracker at LHC." *Journal of Physics G-Nuclear and Particle Physics*. 31:N9-N20, Mar. 2005.
15. S.S. Blackman. "Multiple Hypothesis Tracking For Multiple Target Tracking." IEEE Aerospace and Electronic Systems Magazine. 19:5—18, 2004.
16. David Stein, James Witkoskie, Walter Kuklinski, and Stephen Theophanis, "Mutual Information based resource management applied to state space estimation," in preparation.
17. Yaakov Bar-Shalom, X. Rong Li, and Thiagalingam Kirubarajan, *Estimation with Applications to Tracking and Navigation,* John Wiley and Sons, New York, 2001, Chapter 5.
18. Alain Chateauneuf and Jean-Yves Jaffray, "Some Characterizations of Lower Probabilities and other Monotone Capacities Through the Use of Mobius Inversion," *Mathematical Social Sciences* Vol. 17 (1989) pp. 263—283.
19. Ivan Kojadinovic, "Relevance Measures for Subset Variable Selection in Regression Problems Based On K-Additive Mutual Information," *Computational Statistics and Data Analysis*," Vol. 49, 2005, pp. 1205—1227.