



Sensor Fusion Architectures for Ballistic Missile Defense

Donald E. Maurer, Robert W. Schirmer, Michael K. Kalandros, and Joseph S. J. Peri

The project described in this article is developing sensor fusion techniques, within the framework of Bayesian belief networks, that will be feasible for operational missile defense systems. As full Bayesian nets may be too computationally intensive for the engagement timeline, we investigated techniques that combine outputs of several small nets operating in parallel. An important application is the handover problem wherein a ground-based radar transmits its track picture to the intercepting missile to be combined with the interceptor's track picture to maximize the likelihood of selecting the true target from among other objects. We discuss net architectures addressing this problem. Since Dempster-Shafer algorithms have also been proposed as a solution, we compare our results with a Dempster-Shafer approach to understand the advantages and disadvantages of each.

INTRODUCTION

The Missile Defense Agency has been tasked to develop missile defense systems to protect both deployed forces from tactical missiles and the United States from strategic threats. These systems rely on information from a variety of sources, including *a priori* knowledge of the threat and ship-, space-, and interceptor-based sensors, to enhance their ability to select and intercept the lethal object in an inbound threat complex (usually the reentry vehicle or RV). Figure 1 is a schematic showing various functions that must be performed in order to select and guide the intercepting missile to the RV. The scenario depicted on the right is an overview of a ballistic missile engagement illustrating how data gathered from various sensors—including satellites, ground- and ship-based radars, and the intercepting missile's IR seeker—are combined to support target selection. Target selection is difficult because, in all but the simplest cases, the threat

complex will consist of multiple objects, for example, a booster, an attitude control module (ACM), and an RV. In addition, there will likely be spent fuel debris and various countermeasures that further obscure the RV. Identifying the true target among all of the objects tracked by the various sensors is therefore a significant challenge and provides the rationale for combining complementary information from different sensors through a fusion process.

Although all relevant sensor information should be brought to bear against advanced threats, only IR/radar fusion is envisioned for near-term systems. In future work, we anticipate extending our fusion concepts to include more general sensor suites. Even with only two sensors, however, fusion is a complex process, as outlined in Fig. 1: the radar detects, tracks, and classifies objects in the threat complex; the information is transmitted to

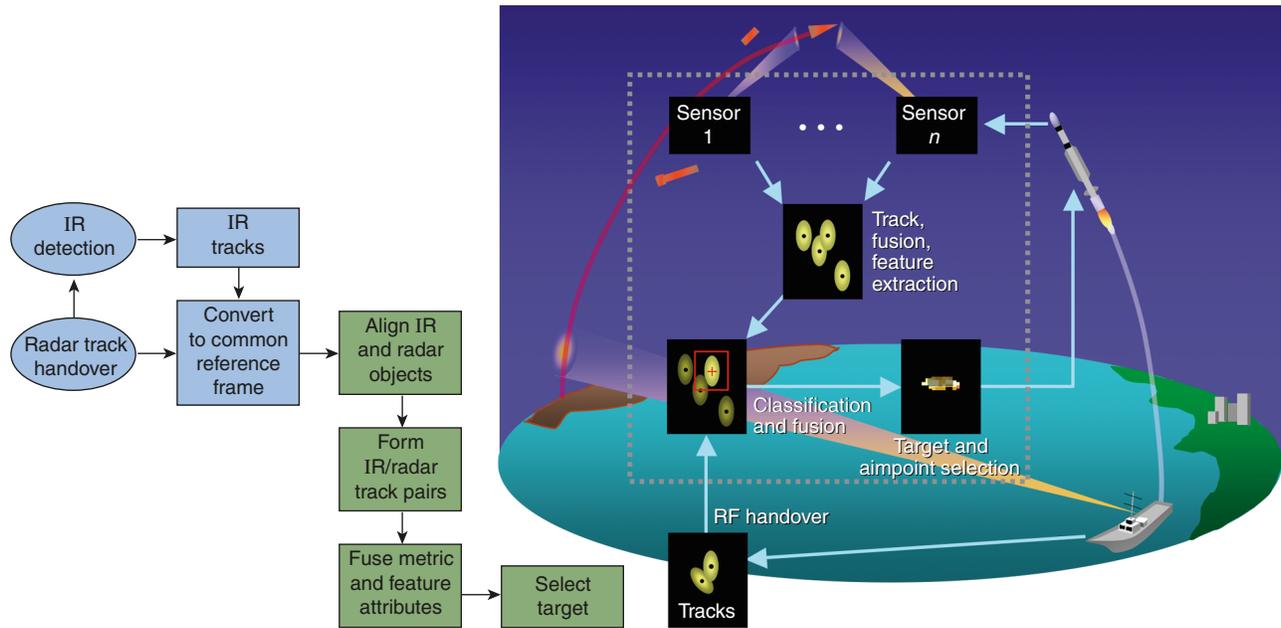


Figure 1. Flow diagram summarizing the major target selection processes in a ballistic missile defense engagement. The green blocks highlight the specific functions that are addressed in this article. Also shown is an overview of an engagement involving data from a number of sensors.

the missile to cue the IR seeker; and, once the seeker has developed its tracks, the information is combined with the track classification information developed from the IR data. A critical step is achieving gridlock between the IR and radar coordinate frames so that tracks corresponding to the same object can be paired. As an example of the problems encountered in this process, missile position and pointing error can manifest as a common offset bias between the two track sets. The green blocks in Fig. 1 indicate the specific target selection functions addressed by our project.

A decision methodology for target selection will be developed later in the project, so we focus here on effective information fusion techniques to provide the optimal probabilistic data on which to base target selection. Ideally, the information fusion process will make maximal use of all available sensor evidence while properly accounting for any interrelationships among the evidence. For example, track state estimates and covariance (i.e., *metric* data) are currently used to correlate tracks in the radar handover with tracks from the IR seeker. The inclusion of feature information should improve correlation accuracy; however, if feature measurements that are strongly correlated with one another are treated as being independent, performance may degrade.

We consider target selection in terms of Bayesian belief networks (BBNs). These networks provide a framework for fusing metric and feature data with prior information that can, theoretically, properly account for interdependencies among data sources. Moreover, there is interest in the Bayesian approach within the ballistic missile defense community because of recent progress in

developing computationally efficient methods for carrying out probability calculations with BBNs.

In view of the short timeline allowed for target discrimination, computational complexity is a major concern. This motivates our approach of combining the outputs of several small nets operating in parallel. To this end, we first investigate networks for single-sensor track classification and then combine them into a system to fuse information provided by two sensors. A well-known alternative to information fusion is the Dempster-Shafer theory.¹ Since this is a competing technique in the missile defense community, we also present a brief comparison of the Bayesian and Dempster-Shafer approaches.

To avoid classification issues related to realistic threat characterizations, engagement scenarios, and high-fidelity performance data, we do not discuss our architectures in terms of the operational performance of real systems; rather, our intent is to show how target selection functions can be synthesized in the case of generic sensors and scenarios. Our results, therefore, are summarized as functions of parameterized sensor performance levels that bracket operational system performance.

The modifications required to apply these algorithms to real systems are nominal. The near-term threat consists of relatively simple two- and three-object ballistic missile threat complexes without debris or countermeasures. Thus, the major modification is to implement the current suite of features for both the radar and IR. (In our notional scenario, only two generic features are defined for each sensor.) The value distributions for these features, which depend on the characteristics of the specific sensors, threats, and interceptors, have been

developed from realistic, but simulated, engagement analyses.

Realistic test data are difficult to obtain since measurements from the various sensors must be correlated with a common engagement scenario. Given the wide range of models being used to generate such data, this is a significant logistics problem. We have addressed this issue with respect to radar and IR sensors by implementing a process for generating IR/RF-correlated test data for the near-term threat. Current data generation models, however, will not support this process for the advanced threats we intend to address in the future. Therefore, lower-fidelity sensor models and scenario generators, along with parametric variation studies that span the expected battle space, will continue to be important validation techniques. These tools also allow greater flexibility in determining critical algorithm stressors and breakpoints. Finally, because the studies presented here are parametric, our algorithms have applicability beyond the missile defense realm.

BASIC CONCEPTS

Consider a simplified example in which an IR sensor tracks a ballistic missile complex consisting of an RV, an ACM, a booster, and debris. Assume that the IR intensity of each type of object differs sufficiently from the intensities of the other objects. Then the sensor can measure the intensity of each tracked object to determine which track corresponds to the RV (in reality, of course, the situation may not be this clear-cut). However, intensity measurements from the RV will not be identical to measurements obtained from another RV in a different engagement, or even from the same RV at different times, because of multiple unknown factors such as environmental and flight variations. Therefore, the intensity I of a tracked object is treated as a random variable. The ability to identify the RV is based on the degree of separation between the probability distribution function $P_{RV}(I)$ and $P_X(I)$ for the other types, where X is an ACM, a booster, or debris.

The $P_X(I)$ can be combined in a single binary, or joint, distribution $P(I, X)$, where the random variables X and I are object type and intensity, respectively. Then $P(I, X) = \mu$ means that μ is the probability of an intensity measurement I and object-type X . Moreover,

$$P(I) = \sum_X P(I, X)$$

is simply the probability distribution for I alone. $P(X)$ is similarly defined; however, since I is continuous, the summation in the definition of $P(X)$ should be replaced by an integration over the range of I .

The fundamental concept in the Bayesian treatment of uncertainty is *conditional probability*. In our

example, the conditional probabilities are $P(I|X)$ and $P(X|I)$. $P(I|X) = \mu$ means that μ is the probability that the intensity is I given X ; that is, $P(I|X) = P_{X|I}(I)$. Similarly, $P(X|I) = \mu$ means that μ is the probability that the object is X given I . The joint and conditional probabilities are related by the factorization

$$P(I, X) = P(X|I)P(I).$$

By symmetry, we also have $P(I, X) = P(I|X)P(X)$, and comparing these two factorizations for $P(I, X)$ yields Bayes' theorem:

$$P(X|I) = \frac{P(I|X)P(X)}{P(I)}.$$

Bayes' theorem provides the mechanism for using measurements of one variable to make inferences about another variable. For example, the conditional probabilities $P(I|X)$ may be determined beforehand, based on the physical characteristics of each object type. Before any intensity measurements have been made, $P(X)$ is the *a priori* object-type distribution. Unless there is prior information favoring one object type over another, this should be uniform; e.g., in our example, there are four object-type classes, and so $P(I, X) = 0.25$. When an intensity measurement I is obtained, the updated class distribution $P(X|I)$, which includes the information provided by the measurement, is calculated using Bayes' theorem. This procedure can be repeated to incorporate multiple measurement updates. A simple target selection technique is to select as the RV the track for which $P(X|I)$ is maximal if $X = RV$.

Now, intensity is influenced by factors other than object type, such as whether the sensor is looking at an object broadside or head on (i.e., aspect). If aspect is modeled as a random variable A , we then must deal with the joint distribution $P(I, X, A)$ and associated conditional probabilities. In particular, the factorization for a binary distribution extends to three variables as follows:

$$\begin{aligned} P(I, X, A) &= P(I|X, A)P(X, A) \\ &= P(I|X, A)P(X|A)P(A). \end{aligned}$$

Both factorizations are special cases of a general product formula

$$P(x_1, \dots, x_n) = P(x_n) \prod_{i=1}^n P(x_i | x_{i+1}, \dots, x_n),$$

showing that any joint probability distribution, $P(x_1, \dots, x_n)$, is uniquely determined by its associated conditional probabilities. However, x_i may not be dependent on *all* the x_{i+1}, \dots, x_n , so a more concise expression is

$$P(x_1, \dots, x_n) = P(x_n) \prod_{i=1}^{n-1} P(x_i | \text{parents}(x_i)),$$

where $\text{parents}(x_i)$ is the subset of the x_{i+1}, \dots, x_n on which x_i does depend. For instance, object type is independent of aspect, so the above factorization for $P(I, X, A)$ reduces to $P(I, X, A) = P(I|X, A)P(X)P(A)$.

A *Bayesian net* is a compact, graphical representation of $P(x_1, \dots, x_n)$ reflecting this factorization. It consists of a set of vertices, or nodes, corresponding to the random variables and a set of directed edges between pairs of vertices (the initial and terminal nodes that determine a directed edge are called, respectively, *parent* and *child* nodes). Together they form a directed acyclic graph (i.e., there are no directed paths $x_{i_1} \rightarrow \dots \rightarrow x_{i_k}$ such that $x_{i_1} = x_{i_k}$) in which the parents of the vertex corresponding to x_i are precisely the set $\text{parents}(x_i)$. Nodes without parents are called *root nodes*. In addition to node connections, the network specification includes the conditional probabilities $P[x_i | \text{parents}(x_i)]$ and the *a priori* probability densities for the root nodes. Thus, the conditional independence structure of the random variables is represented in graphical form. In this article, we assume that all variables are discrete, although this constraint can be relaxed. Therefore, the net probabilities can be specified in tables. Finally, it is worth noting that different factorizations are possible, and each one determines a different Bayes net representation of $P(x_1, \dots, x_n)$, while each Bayes net, on the other hand, corresponds to a unique joint distribution. Standard references²⁻⁴ are available that provide a more detailed exposition of the theory.

As a first example, consider the simple network shown in Fig. 2a for gauging the power of a single feature to discriminate objects, where the root node a represents the object class variable. The root node is discrete and takes on as many values as there are object types. We

assume that the *a priori* probability distribution $p(a)$ is uniform unless more specific information is available. The child node d represents the feature-value random variable. Observed values (i.e., sensor measurements) are denoted by e_1 in Fig. 2a and are also indicated graphically by a nondirected line segment connected to the proper node.

On the other hand, if we want to determine the performance improvement gained by fusing multiple features, the simplest architecture for doing so is shown in Fig. 2b. Since the feature nodes are independent, this net may ignore some dependencies. In many cases, however, networks with such implied independence assumptions actually outperform more complex models.⁵ Figure 2c is a more advanced network that includes an additional root node b , which can be used to model a parameter that may influence the feature nodes. For instance, some feature measurements may depend on the angle at which the sensor views an object, and this dependency on the aspect angle can be represented by the directed edges from node b , where b denotes aspect angle. The increase in computational demand over the simple networks is relatively small. Although the nets in Fig. 2 show at most two features, denoted by the variables c and d , they can be extended in an obvious way to include any number. The primary output is the *a posteriori* probability distribution of the class variable a after the net has been updated based on evidence in the form of measured values for a subset of the feature set. Node b is not measured directly, but is estimated along with the object class and can be considered as another output.

The product formula corresponding to the multiple-feature classifier with scenario variable is $p(a, b, c, d) = p(c|a, b)p(d|a, b)p(a)p(b)$; thus, the *a posteriori* joint distribution is

$$p(a, b, e_c, e_d) = p(a)p(b) \times p(c = e_c | a, b)p(d = e_d | a, b),$$

where $\mathbf{e} = (e_c, e_d)$ denotes a vector of feature measurements. Hence, the *a posteriori* probability distribution for a is

$$p(a | \mathbf{e}) = \frac{p(a, \mathbf{e})}{\sum_{a_i} p(a_i, \mathbf{e})} = p(a) \frac{\sum_{b_j} p(c = e_c | a, b_j)p(d = e_d | a, b_j)p(b_j)}{\sum_{a_i, b_j} p(c = e_c | a_i, b_j)p(d = e_d | a_i, b_j)p(a_i)p(b_j)},$$

where the summations range over the domains of the indicated random variables. Similarly, $p(b | \mathbf{e})$ can be obtained by interchanging a and b in the above

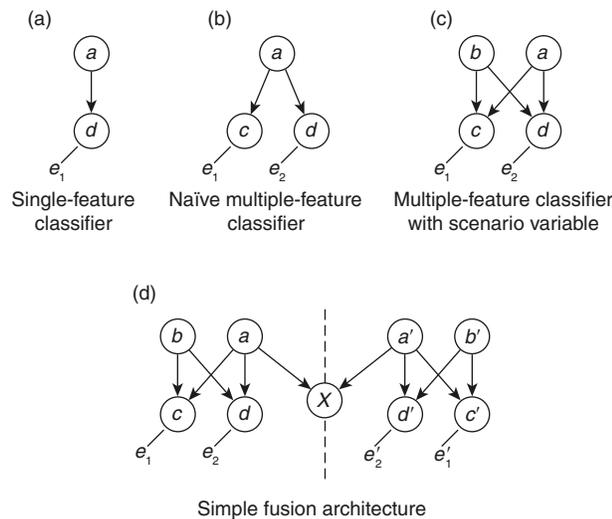


Figure 2. Simple Bayesian classification nets.

equation. If, for instance, there is no measurement for feature c (e.g., $e = e_d$), then the term $p(c = e_c | a, b)$ must be replaced by 1 in the update equation for $p(a|e)$. This result generalizes to any number of features. Although other net structures are possible, our objective here is to illustrate ways of combining multiple Bayes nets into fusion architectures.

APPLICATION OF A SINGLE-SENSOR BAYESIAN NET

Our first step was to investigate Bayesian nets for single-sensor discrimination in order to develop a basic building block for multisensor architectures. In this section we discuss the application of single-sensor nets to analyze the discriminatory power of a set of features and select an optimal subset for classification. Smith⁶ proposed a set of 13 radar features for discriminating a cone from a cylinder. (The RV is roughly conical and the ACM is roughly cylindrical.) These included, among others, several length-related features, the power returned by various types of scatterers, and the maximum of the absolute value of the return.

A database of feature measurements was generated by an X-band radar simulation and included, for each object, values of each feature indexed by the aspect angle between the object and the radar. (The aspect angle is zero when the front of the object—the point of the cone, for example—is facing the radar and increases to 180° when the back of the object is facing the radar.) The aspect angle varied in steps of either 0.10° for the cone or 0.05° for the cylinder. (No roll angle is assumed since the radar properties of these objects are roll-angle invariant.) Figure 3a is a plot of the first length feature as a function of the aspect angle of the target. The data are without noise, so the measured features are deterministic functions of the object class and its aspect angle. However, we assume the aspect angle to be stochastic with a uniform distribution, and thus the radar features become stochastic as well. But without noise, the training data used to create the conditional probability distributions are the same as the test data, which could result in an overly optimistic performance expectation.

To gauge the discriminatory power of each feature individually, we use the single-feature classifier (Fig. 2a) consisting of a class node with possible states of cone and cylinder and a measurement node consisting of possible measurement states (although measurements are continuous, they can be divided into separate bins to simplify processing within a BBN). The *a priori*

probability density for the class node is uniform, i.e., $P(\text{cone}) = P(\text{cylinder}) = 0.50$, and the conditional probability table for $P(d|a)$, shown in Fig. 3b for the length feature, was derived by binning the data in Fig. 3a. A separate BBN is assigned to each object, but the conditional probability tables are the same. Since this net does not contain a node for aspect angle, body orientation must be incorporated in the conditional distributions. The time since the two objects separated is assumed to be long enough for their aspect angles to be independent of one another. Thus, we may assume that the aspect angles for each object are independent and uniformly distributed.

Feature measurements are simulated by first randomly drawing an aspect angle and then generating the measurement from a lookup table. True aspect angles are only used to generate the data and are never passed to the BBN, since they would not be available in an actual engagement. The BBN processes the measurement data to arrive at the probabilities $P(\text{cone}|feature)$ and $P(\text{cylinder}|feature)$ for each object. Target selection can be performed as a simple hypothesis test by selecting the object with the highest $P(\text{cone}|feature)$ or, if the number of targets is unknown (e.g., engagements with threats consisting of multiple RVs), all objects for which $P(\text{cone}|feature)$ exceeds some threshold can be selected.

Although the individual performance of several of the features was very good, we also wanted to determine the improvement gained by fusing multiple features. One fusion approach is the naïve multiple-feature classifier (Fig. 2b), so named because it ignores dependencies among the features and does not fully exploit the aspect information contained in the measurement data. On the other hand, the features as modeled all have a correlated noise input: the aspect angle. Thus, an alternative is the multiple-feature classifier with an additional node to model aspect angle (Fig. 2c).

To construct the conditional probability tables for this new node, the full range of aspect angles is divided into bins and a separate feature histogram is created for

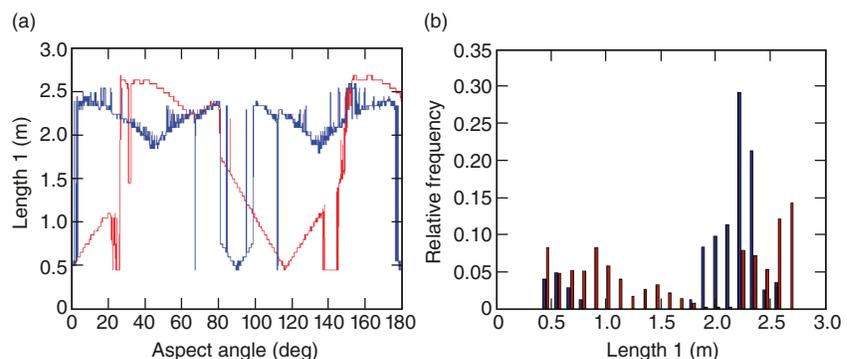


Figure 3. (a) Raw data (simulated) characterizing radar length measurements of a threat object as a function of aspect angle. (b) A discrete conditional probability distribution derived from the data. (Red denotes cone; blue denotes cylinder.)

each bin. The *a priori* probability density for aspect angle is assumed to be uniform across the bins. The performance of these two network architectures was examined for several subsets of features to try to determine an optimal choice. Both architectures were able to accurately discriminate the cone from the cylinder; however, the design that estimates aspect angle performed better, even when the number of features was reduced. The strongest overall performer used only four features but was 99.9% accurate when configured as the simple hypothesis test described in the previous paragraph (selecting the most “cone-like” object) and had one of the lowest computational demands.

When comparing the *a posteriori* probabilities to a set threshold rather than across objects (again, as described in the previous paragraph), the results can be represented as a “receiver operating characteristic” (ROC) curve that plots the probability of detection (the object is declared a cone when it is actually a cone) as a function of the probability of false alarm (the object is declared a cone when it is not a cone). The detection threshold governs the relationship between the two values. Varying that threshold will produce multiple pairs of detection and false alarm probabilities that can be plotted as the ROC curve. An ideal ROC curve will have very high probabilities of detection for very low probabilities of false alarm, resulting in a sharp “elbow” on the graph. Although these curves are generally used to detect a signal in noisy data, the representation of the aspect angle as a random variable replaces the noise of our system and allows us to use a similar representation to empirically evaluate the performance of our classifier. The ROC curves for the two nets are plotted Fig. 4. We selected the multiple-feature classifier with the additional aspect angle variable as our *fundamental net* for single-sensor discrimination based on this analysis.

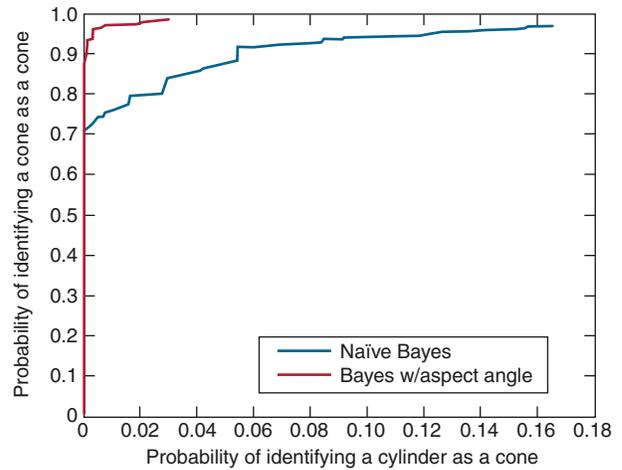


Figure 4. ROC curves, obtained by incorporating information on the aspect angle between radar and target object, showing the improvement in the probability that a cone is correctly classified as a cone.

BAYESIAN NET-BASED SENSOR FUSION ARCHITECTURES

Recall that the success of multiple-sensor fusion depends on several factors, including the ability of individual sensors to determine what objects their tracks correspond to, how well the transmitted information can be correlated with the appropriate seeker-developed tracks, and how the object-type probabilities can be combined to extract the optimal targeting information. We developed several fusion architectures incorporating these factors by combining the fundamental sensor nets in various ways to attain a unified target selection system. Two examples are discussed in this section.

The effectiveness of these architectures was analyzed using the notional test scenario shown in Fig. 5, which

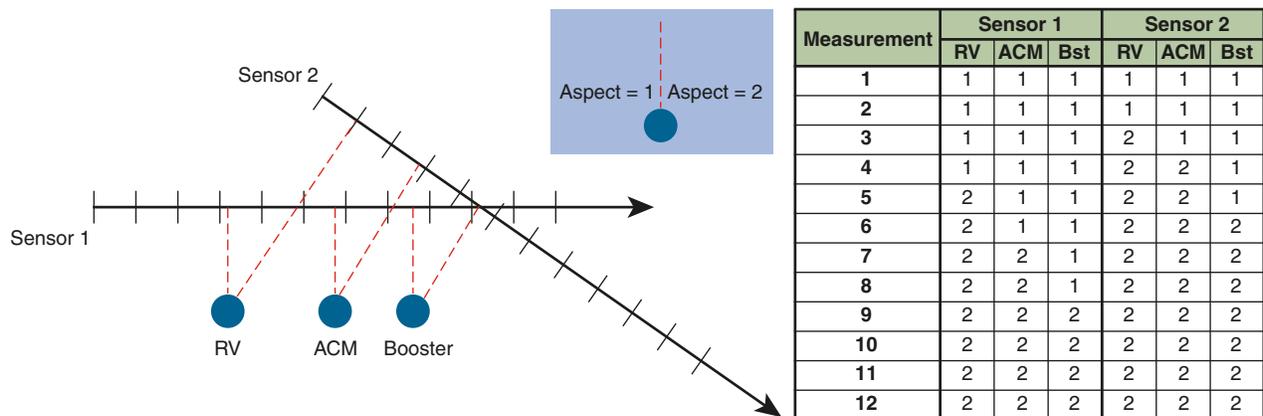


Figure 5. A diagram illustrating a test scenario involving two generic sensors, each tracking three objects. During the engagement, the sensors move along their respective linear trajectories, making measurements at regular time increments as indicated by the tick marks. Aspect is considered to be a binary variable. The table summarizes the variation in aspect with respect to each sensor and object throughout the engagement.

models two sensors collecting information on a threat complex consisting of an RV, a booster, and an ACM. The information collected by each sensor consists of measurements of two generic features from each object. In this scenario the objects are assumed to be stationary, and only the sensors move. As the sensors move along their linear trajectories, they measure features on each object at equally spaced intervals as indicated in the figure. Generally, features depend on a number of variables defining the geometry of the encounter and the environmental conditions. As noted earlier, the aspect angle is important because it determines the target cross section presented to the sensor (e.g., head-on vs. broadside). Our simplified scenario, therefore, includes aspect angle as a representative scenario-dependent variable. The table in Fig. 5 summarizes the aspect variation for each sensor and object.

We can fairly accurately model a real sensor's time series of measurements for a variety of features derived from IR intensity, radar reflectivity, and rotational properties of the target objects by appropriate adjustments of the feature node conditional probabilities. In particular, we consider two capability levels that bracket the expected performance of these sensors. In the first case, the sensors provide relatively strong class discrimination through their feature measurements; in the second case, they provide relatively weak class discrimination. The conditional probabilities defining the first case are shown graphically in Fig. 6a and those defining the second case are shown in Fig. 6b. The scenario proceeds through 12 time steps, and independent measurements of the features for each track are made at each time. Drawing random values from these distributions simulates this sequence of measurements. The aspect angles to the target objects are unknown to the sensors.

The goal of this scenario is to determine which object is the RV based on information provided by the measurement data. Each sensor can independently process its measurements and make a decision, or they can combine their information to make a more accurate decision. Before fusion can occur, it is necessary to attempt to correlate, or pair, the tracks of one sensor with those of the other, such that the paired tracks correspond to the same ground truth object. Specifically, suppose the

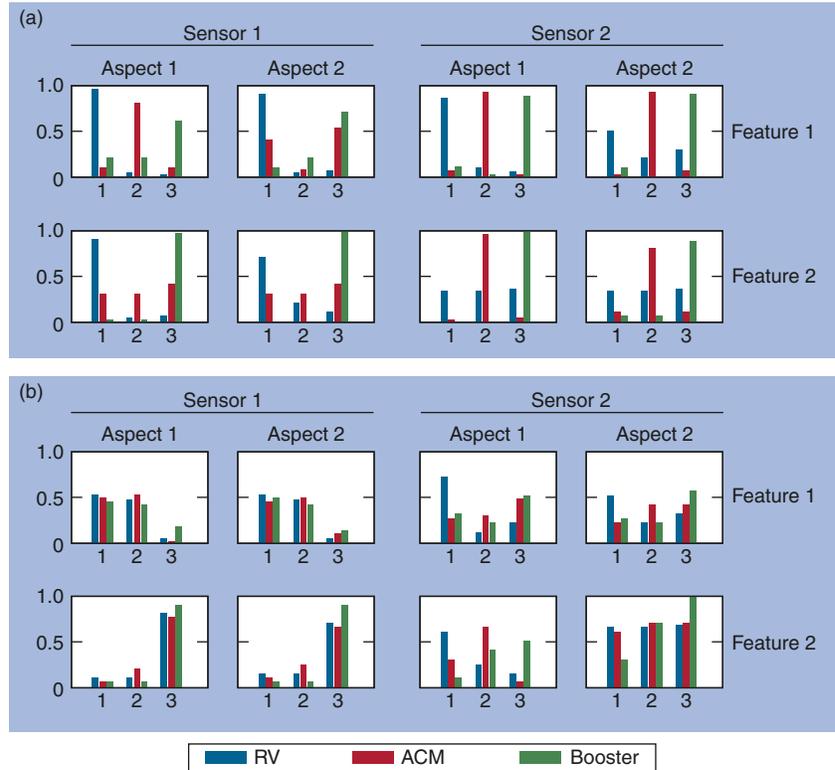


Figure 6. Conditional probabilities: (a) the strong discrimination case and (b) the weak discrimination case. In these histograms the vertical axes represent the probability of each of the three possible feature values (1, 2, or 3) being measured by each sensor. The probabilities are functions of target type, aspect to the target, and the feature being measured (1 or 2).

first sensor is tracking m objects and the second sensor is tracking n objects (without loss of generality, we shall assume $m \leq n$). A *correlation hypothesis* is a mapping $ch: \{1, 2, \dots, m\} \rightarrow \{0, 1, 2, \dots, n\}$, such that $ch(i) = 0$ if the first sensor's i th track does not correspond to any track from the second sensor, and $ch(i) = j$ if it corresponds to the j th track. We shall also assume that the following holds: if $ch(i) \neq 0$, then $ch(j) = ch(i)$ if and only if $j = i$.

Our first architecture is based on a simple system, shown in Fig. 2d, for determining whether a track pair from two sensors should be correlated. It consists of a fundamental net attached to each track for each sensor. All pairs of nets, one from each sensor, are connected via nodes corresponding to a binary random variable X , where $X = 1$ if the corresponding tracks are from the same object and $X = 0$ otherwise. The connection is such that the two class nodes are the parents of node X . We denote probabilities related to this net by P and use primes to reference the second sensor. The *a priori* probabilities for the root nodes are uniform. Defining the conditional probability

$$P(X | a, a') \approx \begin{cases} 1 & \text{if } a = a' \text{ and } X = 1, \text{ or if } a \neq a' \text{ and } X = 0 \\ 0 & \text{otherwise} \end{cases}$$

completes this system definition, where “ \approx ” denotes equality up to a multiplicative normalization factor. It can be shown that

$$\begin{aligned} P(a|e,e') &= p(a|e), \\ P(b|e,e') &= p(b|e), \\ P(a'|e,e') &= p'(a'|e'), \\ P(b'|e,e') &= p'(b'|e'). \end{aligned}$$

Hence, each net can be updated independently of X . To determine the most likely correlation hypothesis, we calculate $P(X|e,e')$ for all possible track pairs and use an assignment algorithm, such as Munkres⁷ or its generalization,⁸ to find a set of pairings that maximizes the sum of these probabilities.

In the fusion architecture, each track net operates independently. Probabilities are updated each time the sensor makes a measurement: the updated root node probabilities become the *a priori* probabilities for the next update, whereas the conditional probabilities remain unchanged. Also, the distributions of the random variables X are updated and the assignment algorithm is applied to determine a correlation hypothesis at each time step. The class node distributions for the correlated net pairs are then fused using a net of the type shown in Fig. 2b.

Figure 7 compares the statistical performance of this architecture to a single-sensor net. The data are the averaged results for 1000 independent realizations of the notional scenario. The strong and weak discrimination cases are shown in Figs. 7a and 7b, respectively. The curves show the time evolution of the average probability that each object is classified as an RV. In both strong and weak discrimination cases, the fused result is only marginally better than it is for the simple track net. However, the separation between the curve for the

true RV and the curves for the non-RV objects is greater with fusion, especially in the weak discrimination case. Evidently, the performance advantage of this simple architecture is limited. Moreover, it requires a link between all pairs of nets from the two sensors (although the fusion nets can be operated in parallel) as well as an external assignment algorithm.

A more integrated concept is shown in Fig. 8. This architecture processes feature and metric data. It fuses a single track observed by sensor 1 with one of two tracks observed by sensor 2. The network structure is easily generalized to M -to- N track mappings for an arbitrary number of sensors. Each individual track is processed by a fundamental sensor net, where for the i th track ($i = 1, \alpha, \beta$), a_i is the classification, b_i is the aspect angle, c_i and d_i represent two features, and $e_i^{(x)}$ represents evidence for feature (or metric) x . Primes denote variables associated with the second sensor. Node \tilde{a}_i is the fused classification for track 1 as determined by the combined evidence of both sensors. Node ch enumerates track correlation hypotheses that map either track α or track β to track 1. For an M -to- N track mapping problem, the number of possible correlation hypotheses, and hence the computational load, increases rapidly with M and N . Therefore, an external algorithm may still be required to limit the number of possible track mappings considered by the BBN. Here, node m represents metric evidence that supports selection of an association hypothesis (e.g., estimated track position).

A given track correlation hypothesis establishes the conditional dependence relationship between \tilde{a}_i and a'_i . Specifically, the conditional distributions defined by

$$p(a'_i | \tilde{a}_1, ch) = \begin{cases} p(a'_i | \tilde{a}_1) & \text{for values of } ch \text{ that} \\ & \text{map track } i \text{ to track 1} \\ p(a'_i) & \text{otherwise} \end{cases}$$

are used, where $i = \alpha, \beta$. Note that in the second part of this definition a'_i is independent of \tilde{a}_1 . Furthermore, the relationship

$$p(a | \tilde{a}_1) = \begin{cases} 1 & a = \tilde{a}_1 \\ 0 & a \neq \tilde{a}_1 \end{cases}$$

(where $a = a_1, a'_\alpha,$ or a'_β) enforces the requirement that a track have one true classification. In effect, if the state of ch is known, all but one of the links between \tilde{a}_i and the various a'_i nodes is removed. The remaining link connects track 1 to the sensor 2 track with which it is correlated.

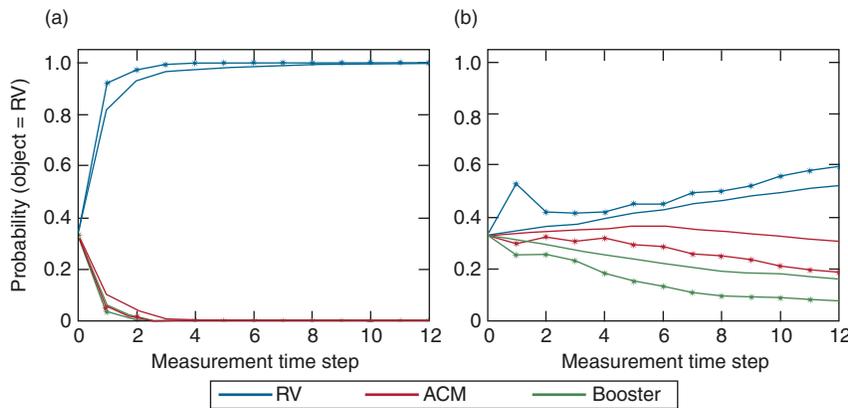


Figure 7. These subplots show the temporal evolution of the probability that each of the three objects is an RV as generated by the first simple fusion concept for both the strong (a) and weak (b) discrimination cases. Solid curves denote results generated by single-sensor track nets; curves with asterisks denote fused results.

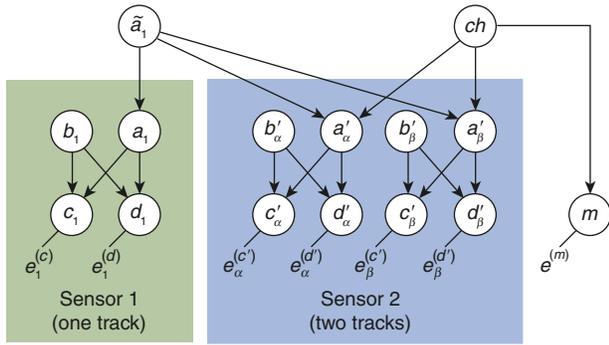


Figure 8. A Bayesian network for multisensor track handover and fusion. The first sensor tracks one object and the second tracks two objects.

Both feature and metric evidence influence the probabilities assigned to the various track correlation hypotheses. A pair of tracks from different sensors that are correlated with each other should share both the same classification (e.g., RV tracks map to RV tracks) and the same metric state (e.g., the same position and velocity). If an external algorithm is used to select a correlation hypothesis, then the selection can be instantiated as evidence for node ch .

This architecture was extended to include three tracks from each sensor and multiple measurement times, and then exercised using the same test scenario as in Fig. 5. To incorporate multiple measurement times, each track is given an aspect node and two feature nodes for each time t , which represent the values of these variables at t . The aspect nodes are all root nodes. The feature nodes for time t are children of their tracks' aspect node for time t and also of their tracks' classification node. In addition to the strong and weak discrimination cases, two track correlation cases are also considered. In the first case, evidence provided for node ch selects the true track association with unity probability. In the second case, no evidence is provided for node ch , and no metric measurement is available, so the correct correlation hypothesis must be inferred from feature measurements.

The prior probability for the fused classification of each track is taken to be uniform so that $p(\tilde{a}_i) = 1/3$ for each object track and class ($i = 1, 2, 3$). Six possible correlation hypotheses map the three tracks of sensor 2 to the three tracks of sensor 1. Each of the six correlation hypotheses is given equal prior probability so that $p(ch) = 1/6$ for every state of ch . (In reality, additional correlation hypotheses in which sensor 2 tracks do not correlate with any sensor 1 tracks would

need to be considered.) The two possible aspect angles are taken to have equal prior probability at every time for every track, so $p(b) = 1/2$ for both angles and any aspect node b .

Figure 9a shows the average probability assigned to the classification hypothesis $\tilde{a}_i = RV$ for each object as a function of time as calculated by this architecture for the strong discrimination case. As noted previously, the average is over 1000 independent realizations of the feature measurement process for both sensors. The curves without asterisks show the results for the case where the correct track correlation is known with unity probability via external evidence. The curves with asterisks show the results for the case where the correct correlation is unknown. In both cases, the RV is assigned probability ≈ 1 of being an RV by the second measurement, while the booster and ACM are given probability ≈ 0 of being RVs. Certain knowledge of the track correlation slightly increases the probability assigned to the RV hypothesis for the actual RV and slightly reduces the probability assigned to the RV hypothesis for the booster and ACM, relative to the case where the correlation is uncertain.

Figure 9b shows similar curves for the weak discrimination case. The average trend with time is toward assigning probability ≈ 1 to the RV classification hypothesis for the RV and ≈ 0 to the RV hypothesis for the booster and ACM. The primary result of the weaker discrimination capability is that more time (i.e., more feature evidence) is required to classify the objects with a given level of confidence, relative to the strong discrimination capability case of Fig. 9a. Again, if the correct correlation is uncertain, and therefore must be inferred from feature evidence, the average trend is toward the correct classification, but the probability assigned to the true classification hypothesis for each track grows more slowly with time than when the correct association is known. Conversely, the weak class discrimination

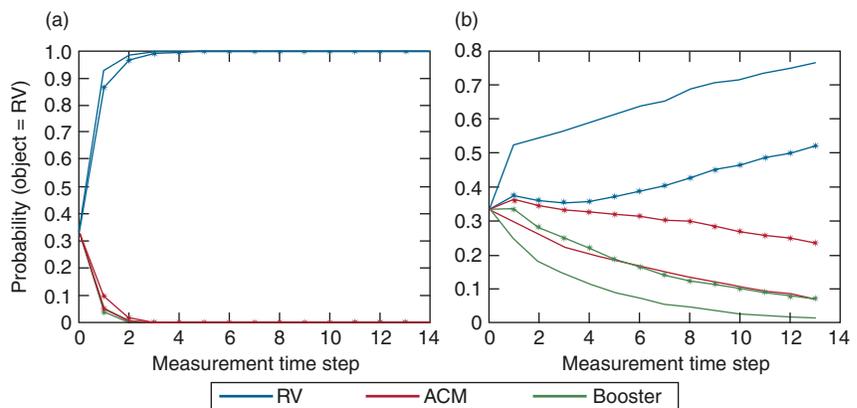


Figure 9. Subplots for the strong (a) and weak (b) discrimination cases for the second fusion architecture. Solid curves show the results when the correct track association is known with unity probability; asterisks show the results when the correct correlation is unknown and must be inferred from feature evidence. For the strong discrimination case (a), the results are nearly identical whether or not correct correlation is known.

provided by the feature evidence impairs the architecture's ability to infer the correct track correlation hypothesis. This, in turn, interferes with the ability of the two sensors to work together. As a result, performance is worse than when the track correlation is known and the sensors can readily combine their evidence.

Finally, consider the probability assigned to the correct track correlation hypothesis. As noted earlier, performance at small values of t is important. In the strong discrimination case, as expected, the feature evidence rapidly generates support for the true track correlation hypothesis, which attains probability 0.9 by $t = 2$. In the weak discrimination case, however, the true correlation hypothesis probability increases gradually (and almost linearly) from $1/6$ at $t = 0$ to $1/2$ at $t = 12$ after all evidence is collected.

COMPARISON OF BAYESIAN AND DEMPSTER-SHAFFER ARCHITECTURES

Recall that there are alternative approaches to information fusion, notably the Dempster-Shafer theory. Since the advocates of Bayesian and Dempster-Shafer methodologies are generally highly polarized, it is interesting—at least briefly—to compare the performance of these techniques on a common scenario.

Although the Dempster-Shafer theory of evidential reasoning^{1,9} deals with concepts of the *belief* and *plausibility* of hypotheses, the theory is a generalization of Bayesian probability. For instance, the fundamental object is a set Θ called the *frame of discernment*, the elements of which we may think of as the possible target types (more generally, they could also be hypotheses or propositions). Bayesian theory would refer to these elements as elementary events and Θ as the set of possible outcomes. A mapping, $m:2^\Theta \rightarrow [0, 1]$, of the power set 2^Θ (i.e., the set of all subsets of Θ) into the unit interval is called a *basic probability assignment* (BPA) if it satisfies the following two conditions:

$$m(\emptyset) = 0, \quad \sum_{A \subseteq \Theta} m(A) = 1,$$

where \emptyset denotes the empty set, and the sum is taken over all subsets A of Θ . Note also that $m(A) = 0$ for $A \neq \emptyset$ is possible. Now, if m were defined on all the singleton subsets of Θ , it would be an ordinary probability distribution on the elementary events. The generality of the Dempster-Shafer theory lies in its ability to operate with *less* information than a probability distribution requires. Thus, we may interpret $m(A) = 0.2$ as our having a *confidence* of 0.2 that the object under consideration is one of the types in A , but having no additional information indicating which one.

The *belief* function, defined as

$$Bel(A) = \sum_{B \subseteq A} m(B),$$

is interpreted as the degree of belief in A , while $Bel(\bar{A})$ expresses one's degree of doubt in A , where \bar{A} is the complement of A in Θ . The plausibility of A ,

$$P^*(A) = 1 - Bel(\bar{A}),$$

expresses one's belief in all those subsets of Θ that "have anything to do with A ." We see immediately that $Bel(A) \leq P^*(A)$. Dempster called P^* and Bel upper and lower probabilities, respectively.¹⁰

The fundamental operation in the theory is Dempster's rule of combination, a method for combining two BPAs. It is defined as follows. If $m(A) > 0$, then A is called a *focal element* of the belief function. The collection of all focal elements of a belief function is called its *core*. Now, suppose we define two belief functions over the same frame. Let Bel_1 have BPA m_1 with a core consisting of focal elements A_1, \dots, A_m . Let m_2 be the BPA for Bel_2 and denote the focal elements of its core by B_1, \dots, B_n , where n need not equal m . Dempster's rule defines the new BPA as follows:

$$m(C) = \frac{\sum_{\substack{i,j \\ A_i \cap B_j = C}} m_1(A_i) m_2(B_j)}{1 - \kappa},$$

where κ is defined by

$$\kappa = \sum_{\substack{i,j \\ A_i \cap B_j = \emptyset}} m_1(A_i) m_2(B_j).$$

The combined belief function, associated with m , is called the *orthogonal sum* of Bel_1 , and Bel_2 and is denoted by $Bel_1 \approx Bel_2$. The denominator is a measure of the extent of conflict between the two belief functions. The greater the number of table entries committed to \emptyset , the greater the amount of conflict. A useful measure of the conflict,

$$Con(Bel_1, Bel_2) = -\ln(1 - \kappa),$$

is called the *weight of conflict* between Bel_1 and Bel_2 . If there is no conflict, $\kappa = 0$ and $Con(Bel_1, Bel_2) = 0$. If, on the other hand, Bel_1 and Bel_2 flatly contradict each other, then $\kappa = 1$ and $Con(Bel_1, Bel_2) = \infty$.

Now, consider a Dempster-Shafer fusion architecture in which each measurement step defines a BPA on $\Theta = \{RV, ACM, booster\}$, which is fused with the BPA from the previous step using Dempster's rule. The theory tells us how to use BPAs once they are defined; however, it does not tell us how to construct them in the first place.

Recall that Figs. 6a and 6b plot the conditional probability distributions of features (c and d) for each of the objects in our scenario. Suppose the measured evidence points only to singleton sets o_j . Then, BPAs can be constructed as follows:

$$m(o_j) = \frac{\sum_{k=1}^2 p(c_i | o_j, a_k) p(o_j) p(a_k)}{\sum_{j'=1}^3 \sum_{k'=1}^2 p(c_i | o_{j'}, a_{k'}) p(o_{j'}) p(a_{k'})}$$

where

- $(o_1, o_2, o_3) = (RV, ACM, booster)$,
- $(a_1, a_2) = (\text{left aspect, right aspect})$ (see Fig. 5),
- $p(o_j)$ = the prior probability that the object observed is o_j ,
- $p(a_k)$ = the prior probability that the aspect of the observed object is a_k , and
- $p(c_i | o_j, a_k)$ = the conditional probability that feature c takes the value c_i , given that the object is o_j (e.g., the conditional probabilities defined in Figs. 6a and 6b).

Under this formulation, we can prove that the Dempster combination rule reduces to Bayes' theorem, and the combined BPA is just the posterior probability.

On the other hand, suppose the measured evidence points to arbitrary sets. We will address two possible BPA constructions. Without loss of generality, we may assume $p(ACM|c_k) < p(booster|c_k) < p(RV|c_k)$ for a feature measurement for which $c = c_k$. Then, define the mapping m using Shafer's *consonance* (i.e., define m on a sequence of nested subsets) as follows:

$$\begin{aligned} m(RV) &= p(RV | c_i) - p(booster | c_i), \\ m(RV, booster) &= p(booster | c_i) - p(ACM | c_i), \\ m(RV, ACM, booster) &= 1 + p(ACM | c_i) - p(RV | c_i). \end{aligned}$$

The Dempster-Shafer approach with this definition of the BPAs was applied to the same scenario and 1000 Monte Carlo cycles used to test the Bayesian net architectures. Figure 10a compares the evolution of the belief and plausibility of the RV for the strong (top) and weak (bottom) correlation cases to the second Bayesian architecture average probability curves in Fig. 9, under the assumption that the sensor track correlation is known.

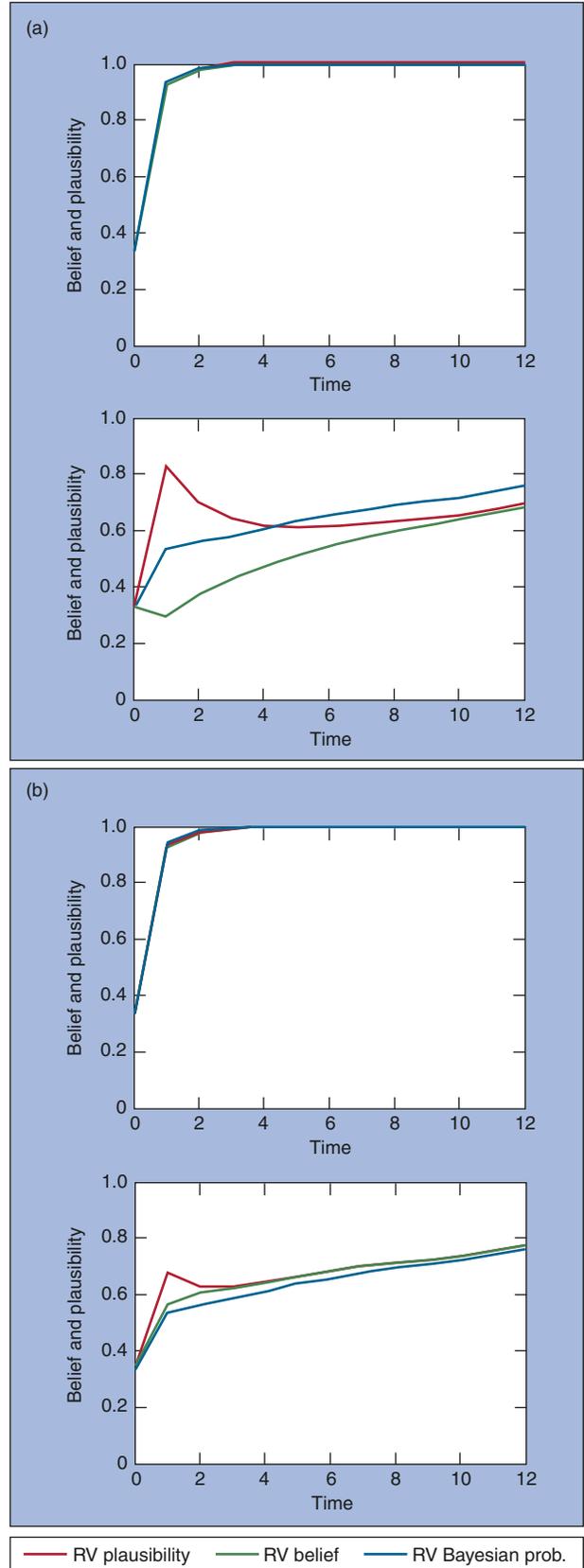


Figure 10. Comparing Bayesian and Dempster-Shafer architectures for the first (a) and second (b) definition of the BPA. Bayesian probability = the average probability assigned to the classification hypothesis that the object = RV when the object is the RV.

Although the plots only show results for the RV, the belief and plausibility for all objects converge to steady-state values, asymptotically approaching the corresponding object probability curves calculated by the Bayesian net. As may be expected, convergence occurs quickly for the strong discrimination case but slowly for the weak case. Moreover, for the weak discrimination case, that convergence lies below the Bayesian probability curve. This behavior is entirely due to the way we have constructed the BPA. (The scenario was designed to be perfectly suited to a Bayesian approach.)

Now, suppose we construct slightly different BPAs as follows:

$$m(\text{RV}) = p(\text{RV} | c_i),$$

$$m(\text{RV, booster}) = \frac{p(\text{RV} | c_i) + p(\text{booster} | c_i)}{N},$$

$$m(\text{RV, ACM, booster}) = \frac{1}{N},$$

where

$$N = \frac{1 + p(\text{RV} | c_i) + p(\text{booster} | c_i)}{1 - p(\text{RV} | c_i)}$$

and

$$p(o_j | c_i) = \frac{\sum_{k=1}^2 p(c_i | o_j, a_k) p(o_j) p(a_k)}{\sum_{j'=1}^3 \sum_{k'=1}^2 p(c_i | o_{j'}, a_{k'}) p(o_{j'}) p(a_{k'})},$$

the remainder of the variables being defined as above.

Figure 10b compares the evolution of the belief and plausibility of the RV with the Bayesian net architecture as before. Again, convergence occurs quickly for the strong discrimination case (top). However, in the weak case (bottom), convergence is attained more rapidly, and the limiting value lies above the Bayesian probability curve. These results show good agreement between the Dempster-Shafer and Bayes techniques. This second BPA construction, therefore, produces more satisfactory results than those of the previous construction.

SUMMARY AND FURTHER RESEARCH

This section summarizes our conclusions based on the parametric analyses presented in this article, as well as the results of higher-fidelity simulated engagements that were not described here. These studies indicate a performance enhancement over single-sensor target selection. We have shown that even for a pair of weakly

discriminating sensors, the fusion architectures provide a performance improvement over either sensor used alone by at least increasing the separation between RV probability curves.

Furthermore, if the track-to-track correlation is known beforehand to a high degree of probability, the improvement relative to the case where the correlation is uncertain can be significant. There is an intrinsic difficulty in the unknown case that is not evident in our test scenario, where the number of possible correlation hypotheses is relatively small. If more objects, such as debris and countermeasures, are present, the total number of hypotheses increases exponentially, making it impossible, in practice, to evaluate the conditional probability table for a correlation node. Therefore, it appears that an external preprocessing step may be necessary to prune the correlation hypotheses to a manageable few.¹¹ However, more work is required to settle this issue.

Even if the individual sensors are strong discriminators, fusion allows faster discrimination. Processing speed is important for time-limited operations such as intercepting an incoming missile. We have shown good performance, with low computational burden, by neglecting some correlations between features. This was shown here, for example, by our single-sensor radar feature results. This conclusion, moreover, is supported by results from a number of independent investigators who have shown that naïve nets often perform nearly as well as complex nets that more carefully model all correlations. However, we plan further work to determine the impact of including evident correlation between features. In addition, based on the comparison results shown here and in our more realistic threat engagement studies, the Bayesian architecture is competitive with Dempster-Shafer approaches in both classification performance and processing speed. Thus, a Bayes net target-selection architecture, along the lines described in this article, appears to be a feasible approach to an operational system.

To develop a capability against future advanced threats, we intend to investigate the utility of combining an extended onboard missile sensor suite, such as multi-color IR and ladar, with Aegis/SPY-1 and another sea-based sensor, where fusion of the sea-based sensors occurs before missile handover. These handover data would be fused again with onboard sensor data. The Mk-99 illuminator is a candidate for the second sea-based sensor. In conjunction with the SPY radar, the Mk-99 can potentially provide a description of the incoming threat that includes whether it is unitary or separating; the number of booster stages; expected countermeasures sophistication; spin, precession rate, and relative mass ratios; and more accurate propagation of track errors. This application will require our net architectures to be expanded to include fusion across multiple sensors and time periods.

ACKNOWLEDGMENT: The authors are grateful for the IR&D support that enabled this research and wish to thank the reviewers for very helpful comments and suggestions.

REFERENCES

- ¹Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ (1976).
- ²Lauritzen, S. L., *Graphical Models*, Clarendon Press, Oxford (1996).
- ³Jensen, F. V., *Bayesian Networks and Decision Graphs*, Springer-Verlag, New York (2001).
- ⁴Koster, J. T. A., "Gibbs and Markov Properties of Graphs," *Ann. Math. Artif. Intel.* **21**, 13–26 (1997).
- ⁵Domingos, P., and Pazzani, M., "Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier," in *Proc. 13th Int. Conf. on Machine Learning (ICML '96)*, pp. 105–112 (1996).
- ⁶Smith, P., *HRRP in X/S: Considerations Relevant to the Use of High Range Resolution Profiles Formed at Different Carrier Frequencies for the Purpose of Automatic Target Recognition*, A2F-01-U-3-007, JHU/APL, Laurel, MD (13 Jul 2001).
- ⁷Munkres, J., "Algorithms for the Assignment and Transportation Problems," *J. Soc. Indust. Appl. Math.* **78**(1), 32–38 (1957).
- ⁸Bourgeois, F., and Lassalle, J-C., "An Extension of the Munkres Algorithm for the Assignment Problem to Rectangular Matrices," *Comm. ACM* **14**(12), 802–806 (1971).
- ⁹Blackman, S. S., *Multiple-Target Tracking with Radar Applications*, Artech House, Inc. (1986).
- ¹⁰Dempster, A. P., "Upper and Lower Probabilities Induced by Multivalued Mapping," *Ann. Math. Statistics* **38**(2), 325–339 (1967).
- ¹¹Maurer, D. E., *Efficient Radar-to-IR Correlation and Bias Estimation*, Report ADB284439, Defense Technical Information Center (2002).

THE AUTHORS

Donald E. Maurer is the Principal Investigator for Sensor Fusion in Ballistic Missile Defense and a member of the APL Principal Professional Staff. Dr. Maurer holds a B.A. from the University of Colorado and a Ph.D. from the California Institute of Technology. Although currently developing and evaluating radar-to-IR handover algorithms, he has also been involved with a number of signal processing algorithm development and simulation tasks. He is a member of the American Mathematical Society, the New York Academy of Sciences, and the Society of the Sigma Xi.



Donald E. Maurer



Robert W. Schirmer



Michael K. Kalandros



Joseph S. J. Peri

Robert W. Schirmer, a member of the APL Senior Professional Staff, received his B.S. in engineering physics from the University of California at Berkeley, and M.S. and Ph.D. degrees in physics from Cornell University. Dr. Schirmer's work includes missile seeker simulations and sensor fusion. He is a member of APS. **Michael K. Kalandros** is a member of the APL Senior Professional Staff. He received a B.S. in engineering science from Trinity University, San Antonio, and an M.S. and Ph.D. in electrical engineering from the University of Colorado. Dr. Kalandros continues to work on target tracking, discrimination, and sensor fusion for ballistic missile defense. **Joseph S. J. Peri**, also a member of the APL Senior Staff, received his Ph.D. in physics from The Catholic University of America. Dr. Peri has been involved in various tasks including submarine communications, IR propagation modeling, and IR systems performance prediction. He also developed an age-dependent track promotion algorithm still used in the Cooperative Engagement Capability. His activities include data fusion, genetic algorithms, and autonomous robot navigation. All team members are in the Air Defense Systems Department and can be contacted through Donald Maurer. His e-mail address is donald.maurer@jhuapl.edu.