Survey of approaches and experiments in decision-level fusion of Automatic Target Recognition (ATR) products

Timothy D. Ross[a], Doug R. Morgan[b], Erik P. Blasch[a*], Kyle J. Erickson[c†], and Bart D. Kahler[d]

[a]AFRL/SNAR, WPAFB, OH 45433, [b]BAE Consultant, [c]Jacobs Engineering, [d]General Dynamics

ABSTRACT

The US Air Force Research Laboratory (AFRL) is exploring the decision-level fusion (DLF) trade space in the Fusion for Identifying Targets Experiment (FITE) program. FITE is surveying past DLF approaches and experiments. This paper reports preliminary findings from that survey, which ultimately plans to place the various studies in a common framework, identify trends, and make recommendations on the additional studies that would best inform the trade space of how to fuse ATR products and how ATR products should be improved to support fusion. We tentatively conclude that DLF is better at rejecting incorrect decisions than in adding correct decisions, a larger ATR library is better (for a constant Pid), a better source ATR has many mild attractors rather than a few large attractors, and fusion will be more beneficial when there are no dominant sources. Dependencies between the sources diminish performance, even when that dependency is well modeled. However, poor models of dependencies do not significantly further diminish performance. Distributed fusion is not driven by performance, so centralized fusion is an appropriate focus for FITE. For multi-ATR fusion, the degree of improvement may depend on the participating ATRs having different OC sensitivities. The machine learning literature is an especially rich source for the impact of imperfect (learned in their case) models. Finally and perhaps most significantly, even with perfect models and independence, the DLF gain may be quite modest and it may be fairly easy to check whether the best possible performance is good enough for a given application.

1. INTRODUCTION

Automatic Target Recognizers (ATRs) are essential for many modern weapon systems. Advanced sensors may produce a large quantity of possibly non-literal data, often making human exploitation too slow or costly. ATRs assist exploitation by computing labels for objects in sensed data. Although ATRs may operate at various levels of abstraction, we are concerned here with ATRs that report a target's presence (detection) and its identity (e.g., T72, M1, BTR70, other). We will use the term Operating Conditions (OCs) for all factors that influence ATR performance, which are often separated into sensor OCs, target OCs, and environment OCs.[1] ATRs occasionally make errors (e.g., labeling a tree as a target or labeling a T72 as an M1), especially in difficult OCs.

The limitations in the reliability of single-source ATR products and the growing sensor data collection capabilities of modern armed forces have made the fusion of ATR products a priority. This data combination could improve the accuracy and detail of information while simplifying its presentation. Even when restricting consideration to decision-level combination of identity data (a subset of level 1 fusion), as we do in this survey, there is a large trade space of fusion approaches. The proper approach depends on many factors, including the type of information available, the accuracy of that information, and how well the uncertainty of that information is known. Similarly, the technology requirements for fusion sources should be driven, at least in part, by what best enables contributions to a fusion system.

Data fusion combines data from multiple sensors (or sensor collections) to (1) improve the reliability and conciseness of the information provided to end-users, (2) increase spatial coverage, and (3) reduce the time to make decisions. There are a variety of approaches to data fusion[2], but we are concerned here only with decision-level identity fusion (DLF). DLF is also known as object-level fusion, identity fusion, or downstream fusion. In fusing ATR products, the ATRs run their individual processes separately and then provide only their final decision (or possibly a list of candidate decisions with confidences, e.g., a likelihood vector) to the fuser. The fuser combines these ATR decisions to produce a single

* erik.blasch@wpafb.af.mil
† kyle.erickson@wpafb.af.mil

Signal Processing, Sensor Fusion, and Target Recognition XVI, edited by Ivan Kadar,
Proc. of SPIE Vol. 6567, 65670V, (2007) · 0277-786X/07/$18 · doi: 10.1117/12.719803

fused decision for the end-user. Although DLF has many limitations compared to fusing intermediate products (e.g., features), there are four key reasons to continue to consider DLF. First, the communication requirements for DLF are small, since only decisions need to be passed to the fuser. Second, DLF algorithms have relatively simple processing requirements. Third, ATRs feeding a DLF system may be designed and operated without having to be adapted to the fusion context. Finally, DLF has a strong theoretical foundation from decision theory that does not yet exist for more advanced fusion approaches.

There are a variety of formalisms available for designing DLF algorithms. We are mostly concerned here with the Bayesian formalism.[3] In a Bayesian approach, we compute the *a posterior* probability of the true target types (or, alternatively, the likelihood of the observations) based on the available evidence and then select the most probable decision to report (i.e. Maximum *a posteriori* (MAP) decision). The computation of the posterior involves prior distributions (knowledge about what to expect) and conditional distributions (knowledge about report reliability). Limitations in our knowledge (or in complexity of representation) of these distributions will be a key aspect of the FITE survey. We view the problem from the Bayesian perspective, not because we believe it is always the best approach, but because it is a valuable reference point. It would be of interest to know the kinds of knowledge and complexity required for an effective Bayesian fuser, what the payoff of that would be, how the payoff diminishes as the knowledge and complexity are compromised, etc.; even when we ultimately decide to take a non-Bayesian approach. The Bayesian formalism provides a framework for problem understanding and a baseline for performance.

This paper will begin by summarizing the FITE framework in Section 2. Section 3 describes a few additional considerations in the survey. Section 4 presents the results of several preliminary surveys. Section 5 summarizes the paper. The survey does not yet have the breadth or depth needed; however, perhaps this paper is a sufficient start for others to provide feedback on the approach and suggest particular pieces of work to consider as the survey progresses.

2. BACKGROUND

The US Air Force Research Laboratory (AFRL) Fusion for Identifying Targets Experiment (FITE) program aims to assess the benefits and costs associated with fusing ATR decisions. The assessment results will help guide the direction of research within the ATR Thrust of AFRL's Sensor Directorate to improve the development and transition of fusion technology and the ATR technologies contributing to fusion systems. The trade-space developed in this paper is a product of the FITE program and will serve as the basis for FITE's experiments. Note that this paper only reports on the FITE trade-space. Although FITE plans to conduct trade studies within this space, such results are not reported here.

The FITE framework[4] provides background for this survey and is summarized here. The trade-space for decision-level identity fusion of ATR outputs is developed from the elements of the FITE equation. The FITE equation relates fuser performance to four basic conditional probability distributions (CPDs). The four CPDs model the <u>R</u>eal and <u>V</u>isible worlds, <u>A</u>TR outputs, and <u>B</u>ayes Fusion outputs. Key dimensions of the main variables ("subscripts") are <u>T</u>ype, operating <u>C</u>onditions (OCs), and <u>U</u>ncertainty. Two sets of models were developed, one for the real world (unprimed) and a second set of approximate models for use by the Bayesian fusion algorithm (primed). The FITE equation does not lend itself to analytical solution because of the complexity of the CPDs, especially the ATR model. A Monte Carlo solution though does appear to be tractable, but may depend on avoiding hyperdistribution representations in the fuser and thoughtful source-to-source dependency modeling (link models). Even when we can solve the FITE equation for a point in the trade-space, we cannot completely explore the trade-space because of the large number of models that must be considered. However, FITE is especially interested in exploring the effects of CPDs with OC conditioning, since there are reasons to believe conditioning will affect both the fusion performance and the requirements on ATRs as fusion sources. DLF trade studies may be thought of as setting the elements of the FITE equation and then solving for performance. Some of the settings will be based on the scenario and application context for the fuser; while others are essentially specifying the fusion technology. By solving for performance, we can better understand what applications are appropriate for decision-level fusion, how much benefit to expect, and which fusion technology (i.e., CPD models) to consider.

The four general classes of random variables used in FITE are:

- Simulation Real-world OCs (R_C) and target type (R_T): the state of the sensed world and the sensor systems. For notational simplicity we separate target type from other OCs.

- Contextual information (V_C, V_U): estimates of quantities related to OCs used for tuning ATR and Fusion algorithms.
- ATR Products (A): ATR algorithm target calls (A_T) and associated score values or any other ATR outputs used in fusion (A_C, A_U). These cover the set of ATR algorithm outputs to be fused.
- Bayesian Fusion Decisions (B): Final fusion call of target existence, type, state and uncertainty.

The joint distribution $p(R_T, R_C, V_C, V_U, A_T, B_T)$, (which we may refer to as $p(R, V, A, B)$) is decomposed using assumptions of conditional independence into the Bayes Net of Figure 1.

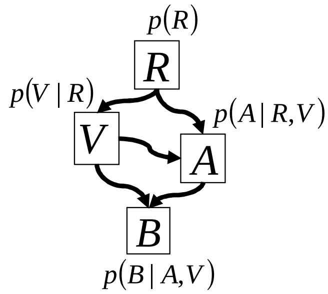

Figure 1 Bayesian network for a FITE system model representing $p(R, V, A, B)$

There is then a parallel set of approximate CPDs (the primed models) used in Bayesian fusion. The expected differences, and the impact of those differences, between the unprimed and primed CPD is a key factor of the FITE trade-space.

3. SURVEY POINTS OF INTEREST

3.1 The Questions of Interest

FITE is intended to provide guidelines for future research and development investments in DLF and the contributors to DLF, especially the ATR related technologies. This section outlines the questions being considered in the development of those guidelines. The remaining sections of the paper begin to gather answers to those questions as they appear in the open literature. A few sources have been considered and their contributions summarized and included in the conclusions. As will always be the case, there are many further sources that should be considered; a few of those are also identified and included in the list of promising papers for further consideration.

At the highest level, the question of interest in FITE is, "What is the expected system performance as a function of the trade-space?" We think of these as independent variables (the trade-space) and dependent variables (performance) and the FITE guidelines are built around this relationship. The guidelines should say where the better places are to be in the trade-space and what performance you might expect there.

Performance considerations include the accuracy of the resulting fused decisions and estimates (e.g., Pid and Pd/Pfa) and the costs associated with the system. The costs that might be considered include development, deployment and maintenance costs (e.g., cost of data for modeling training or tuning) and operation costs (e.g., cost of communications, processing, operator labor). Some costs are actually monetary; others are not, for example the bandwidth requirements of communications). It may be preferable to consider latency as a performance issue rather than (or in addition to) processing as a cost issue.

The trade-space includes the application / problem domain for the system (R model) and the technical approach of the system (V and A models and the fuser B, with B possibly including the prime models R', V', and A').

Within the broad problem of how R, V, A, and B map to performance there are several specific issues of special interest to FITE. The following sections discuss these by going through the FITE framework's major elements.

3.1.1 R-related Questions

- What are the scenarios and conops that the system may encounter (i.e., what are the OC distributions – types of targets, OCs of the targets, environment, and sensors, number of sources)?
- Where should the controllable OCs be set to maximize fused performance?

3.1.2 V-related Questions

- What is the context for the ATRs and the fusers (i.e., where those computations (A and B) are performed, when they are performed, and what is available (machine-readable; V_C and/or V_U) at that place and time)?
- What is the accuracy of the decisions and estimates made available by V (V_C and/or V_U)?

3.1.3 A-related Questions

- How well do the sources perform?
- What are their sensitivities to OCs?
- What is the impact of the parameter settings for A (e.g., detection threshold)?
- What functionality will A have (A_C, A_U, full hypothesis list or only best guess, …)?
- What if sources have different target and/or library types?
- How do sources use V information (e.g., whether V information is generic or specific to source and time of collection)?
- What if the sources' functionalities include no-decision options (for detection and/or ID)?
- What if the sources have different nominal or training OCs (i.e., the prime models may need to be concerned with absolute OC values rather than simply whether they are nominal or not)?

3.1.4 B-related Questions

- Type of fusion to use (decision, feature, or signal level)? We are concerned here only with DLF in the decisions-in and decisions-out sense, except allowing for scores with the inputs, but this important question is acknowledged.
- What fuser architecture (multi-sensor versus single-sensor with multiple ATRs; serial versus parallel, centralized or distributed fusion)?
- What fuser approach (e.g., AND, Majority, M of N, Bayes MAP/ML, Naïve Bayes, most-reliable-source, Dempster-Schafer, Fuzzy, …)?
- For applicable approaches, which OCs are conditioning variables in the prime models?
- Does the fuser use all of the products from V and A (including possibly a full hypothesis list, A_C or A_U)?
- What is the effect of the fidelity of the fuser's models (i.e., similarity between the prime models and their respective unprimed versions)? How does this fidelity evolve with time? How do the fidelity requirements vary with the sophistication of the fusion approach?
- Does the fuser produce OC (B_C) or uncertainty estimates (B_U)?
- What is the association performance and how is it measured?
- What does the fuser need to know about the sources' target / library types?
- What is the computational complexity and how can it be managed?
- How can DLF be integrated into systems so that it can be used when helpful and not interfere otherwise?

- What is the system's performance with DLF? To what extent are dependencies between the sources a factor in performance for a given approach?

3.1.5 Summary

The questions are mostly posed as "what is", but it is of equal interest to know "what should" these various elements be for an improved system. The answers to the latter form of the question will help set technology requirements for those sub-systems.

Many of the questions could be broken into many more specific questions, FITE is doing this in some cases. In other cases, FITE is fixing a baseline setting and not further considering the issue. FITE is also not considering the broader implications of fusion, such a registration, sensor management, or on-line adaptation of the source parameters for fusion.

In answering any of the questions above, there is the issue of confidence in the answer. This might also be thought of in terms of the breadth of applicability of the answer. FITE's approaches to characterizing confidence in any results will include

- Component Testing (we will get some idea of how close our components are to the real world, esp. A and B)
- Sensitivity Experiments (we will try to learn what aspects of the components' fidelity have the most impact and contrast that with the available degree of fidelity)
- Bounding Experiments (esp. R and V) - to the extent we can characterize the uncertainty in model parameters, we may try to do a test at worst case settings within that uncertainty and then again at best case settings
- Reproducing results from the literature.

Answers to the above questions must eventually support a process that also considers the available technologies and the potential for improving those technologies as investment decisions are made. Examples of such technologies that might be considered include prime models (R', V' and A' modeling), adaptive acquisition of prime models, use of consistency / inconsistency in the sources' OC estimates to better model uncertainty, extraction of OCs from the sensor data for use as conditioning variables (A_C production), making machine-readable other information available at the point-of-fusion (i.e., making V available), ATRs that produce accurate likelihoods, posteriors, ..., and ATRs and fusers that reason at the class level (e.g., to get more accurate performance at cruder levels of discrimination).

There are other indirect investment decisions that FITE should help inform, including the need for data collections, ATR tests, potential for transition of a given fusion technology, and recommendations on how to employ the technology (such as the diversity of phenomenology, perspectives, and times of the sources).

3.2 Survey Considerations

Papers were selected for review based on their relevance to the FITE program. Consideration was limited to works appearing in the open literature. The paper selection is not completed, but enough papers were found to illustrate the process. The FITE review process attempts to place a given work within the FITE framework. To help with this, we maintain a list of the various dimensions of the DLF problem that would help place each piece of work. Some papers cover multiple points in this space. Occasionally a paper will not completely define its place in this space, perhaps leaving out elements that are commonly set a particular way in the community of the publication. We do not report here where the reviewed papers fit in the space, but do provide a description of the space, whose key dimensions are the fuser itself and the fuser's performance. The considerations in the description of the fuser itself include:

- Functionality (detection, ID, OC estimation, uncertainty characterization, …)
- Approach
- Level of detail in internal (primed) models or, perhaps equivalently, the simplifying assumptions
- Accuracy of the assumptions (i.e., accuracy of the prime models (R', V', and A') of the FITE framework)
- Association performance (note that FITE is not attempting to work the association problem, therefore we treat this as a property of the environment within which the fuser performs)

The considerations in the description of the performance assessment's independent variables include:

- Sources (A and V models)
  - Number and type of sources
  - Functionality of each source (detection, ID, OC estimation, uncertainty characterization, …)
  - Accuracy of each functionality and each source
  - Number and types of objects that each source is trained to recognize
  - Parameters of each source (e.g., ROC operating point)
- Scenarios
  - The OCs and their distributions (R model)

The considerations in the definition of the performance assessment's dependent variables include:

- Accuracy
  - There might be one or more metrics per functionality
- Cost
  - Considering fuser algorithm runtime, cost of satisfying prior knowledge and modeling requirements, cost of communications from sources to fuser, etc.
- Uncertainty characterization
  - For empirical results, typically indicated by reporting the number of samples involved or confidence intervals

4. SURVEY RESULTS

4.1 Statistical Decision Theory

Statistical decision theory[5] provides a theoretical framework for DLF. If the models (typically as probability distributions) are known and tractable then fusion is the straight forward multiple-observation generalization of the single observation decision theoretic problem (i.e., some likelihood ratio test, where the threshold depends on what is known (priors and/or costs) and desired optimality criterion). This applies whether the multiple observations are two pixels within a given image or features from two different images that are being fused. Some problems do have known models and are tractable, e.g., with multivariate Gaussian distributions. The disciplines of pattern recognition and machine learning concern themselves with the situation where the models are not known, but data samples and perhaps a few other constraints are known. The fusion community also considers the problem of various cases of known models, where tractability or cost constraints are the driving issue.

4.2 DLF Baseline Analysis

4.2.1 Dasarathy 1994 Chapter 2

There is a collection of assumptions that make performance analysis relatively straightforward. We take Chapter 2 of Dasarathy'94[6] for a baseline case and consider several excursions from that baseline. Several of the excursions are from Dasarathy'94 while others are from personal communications from D. Morgan in 2007.

The baseline case assumes:

1. independent sources
2. all sources have the same performance
3. all test classes are equally well identified (the confusion matrix diagonal terms are all the same)
4. all library classes are equally confused (probability mass equally distributed across off-diagonal confusion matrix elements)

5. sources are forced-decision (i.e., always provide an identification decision)
6. AND fuser (aka consensus)
7. fuser is not forced-decision (i.e., may not provide an identification decision)

The performance analysis is then with respect to independent variables: number of sources (N), number of library classes (L), performance of the sources (Pid or x) and involves analytic expressions for fuser performance metrics: rate of correct decisions, P(Correct), rate of decisions, P(Decision), and conditional rate of correct decisions, P(Correct | Decision)[‡].

Excursions are then made on that baseline, i.e.,

- Assumption 3 relaxed, not requiring that all classes be equally confused (for L = 2 case only, requires introduction of priors)
- Assumption 5 relaxed, not requiring the sources to be forced decision (for N = 2 case only).
- Assumption 6 replaced with majority agreement fuser (for N = 2 or 3 only)
- Assumption 6, M-of-N fuser is discussed but not developed, noting that the performance equations "... can be worked out ..." but "... are rather complex ...".

Many special cases of the above are developed in Dasarathy'1994, but the results of most interest here are Figures 2-4[§] (plot of fuser metrics versus source Pid for Baseline case, except N=2, L=2), Figure 2-11 (plot of fuser metrics versus source Pid for Baseline case, except N=3, L=2), and Figure 2-14 (plot of fuser metrics versus source Pid for Baseline case, except N=3, L=2, Majority fuser[**]).

The most striking aspect of the first two figures is that the fused correct decision rate is well below that of the individual sources (which would be the lower-left to upper-right diagonal on those plots). The desirable effect of the fuser in these cases is to reduce the incorrect decisions. It also eliminates some correct decisions, but the conditional rate of correct decision does improve.

4.2.2 Baseline Excursions by Morgan

The main difference between this analysis and the baseline is that the fuser used here is M-of-N voting (Assumption 6). This is of special interest in FITE because M-of-N voting and thresholded-MAP give the same fuser results. For every (M,N) there is an equivalent posterior probability threshold and for every threshold there is an equivalent (M,N). This is a "hard/best case" for fusion: maximum entropy given L and x.

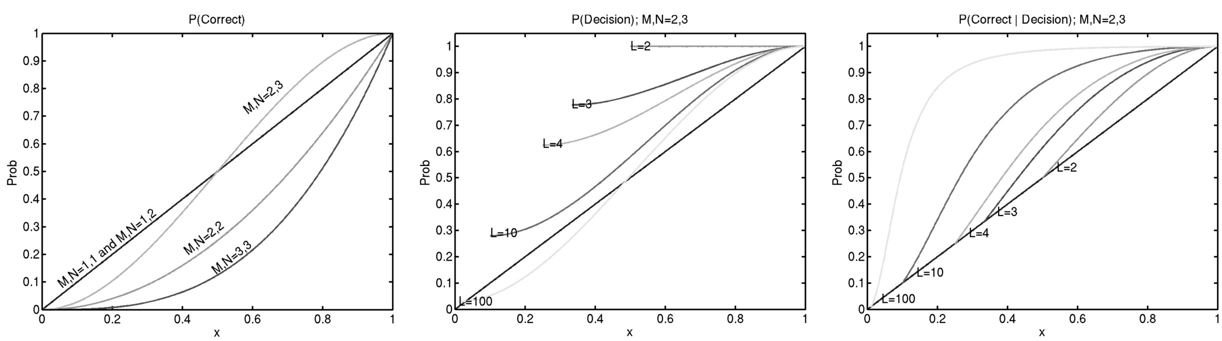

Figure 2 Decision and correct decision rates as a function of source performance, no. of sources, and fusion rule

The direct observations included with these results were that DLF is better at rejecting incorrect decisions than at adding correct decisions, DLF gain varies with operating point (in this case, M, N, L, and x), more sensors is better, higher x is

[‡] This metric was provided by Morgan only.
[§] Figure numbers 2-4, 2-11, and 2-14 refer to figures in Reference 6.
[**] This result is among the lines in the left-hand plot of Figure 2.

better, and larger L is better (for a constant x). Noted from this is that uncertainty in operating point corresponds to uncertainty in performance.

Since the performance curves give performance versus system parameters (M, N, L, and x), they also describe the sensitivity of performance to changes in system parameters. For instance for a given M, N, and L, a user might assume that x is 0.8 when it is really 0.7. Two looks at the curves then give what the user will think should be the fusion performance versus what the fusion performance really is. Therefore, several further inferences were made by noting how some excursions from the baseline conditions might have a corresponding baseline operating point. That correspondence may then allow conclusions about the otherwise intractable excursions from the baseline analysis. First, it was noted that if the off-diagonal elements of the source confusion matrix were not all the same (contrary to Assumption 4), the overall behavior might be approximated by a baseline setup with L reset to only count the main attractors[††] and adjusting x to maintain rows that sum to one. From this, one can conclude that a better source ATR has many mild attractors rather than a few large attractors. A second deviation from the baseline conditions is when there are dominant sources (perhaps OC dependent), contrary to Assumption 3. This case may be approximated with a baseline set up with N reset to only count the better performing sources. Clearly fusion will be more beneficial when there are no dominant sources.

Finally, it is noted that, even with perfect models (MAP with perfect models corresponds to some M-of-N voting case), the DLF gain may be quite modest and it may be fairly easy (utilizing the baseline analysis with excursions and inferences) to check whether the best possible performance is good enough for a given application.

4.3 DLF Papers

4.3.1 Drakopoulos'1991[7]

Among the selected studies included in Dasarathy'1994, this work develops an optimal (in the Neyman-Pearson sense) fuser for correlated sources. The degree of correlation is parameterized with a mutual-information based metric. Test results are reported for a 2-class problem. In many of the cases studied, performance degraded immediately with the introduction of correlation among the sources. In all cases, by the point of correlation equal to 0.2 there was noticeable performance loss.

4.3.2 Butler'2003[8]

This US Navy sponsored effort to perform fusion of aircraft identification information has a number of parallels with FITE. Their use of a Bayes Net and standard test data sets is particularly relevant. Other than comparing two not-in-library recognition approaches, the paper does not report on trade studies; however considering the depth of the work and its similarity to FITE, discussions with the authors are probably warranted.

4.3.3 Chong'2005[9]

The main point of this BAE work for MDA and USAF is to assess distributed fusion (i.e., where multiple entities are each requesting and fusing information from the others as needed). Distributed fusion is contrasted with centralized fusion (where all of the information comes to a single entity for fusion). A key issue in distributed fusion arises because a given entity might receive the same measure through different cooperating entities. The study considers that form of dependency and the form of interest in FITE, where it uses a Bayes Net for the primed-models of a similar character to that of FITE's. Centralized fusion is better in identification performance, but distributed fusion may approach the centralized fusion results at much lower communication costs under some circumstances. The consequences of the dependencies were not immediately apparent in the reported results, but a more careful reading or discussions with the authors may yield additional information.

4.4 Multi-ATR Fusion

There is a lot of activity in this area, much of it stemming from ensemble methods[10] in machine learning. There is recent work by the US Army and US Navy reported in Rizvi 2003[11] and Lynch'2003[12] respectively. Rizvi'2003 considers the fusion of four FLIR ATR algorithms, each trained with the same data, using several different fusion approaches. Noted there is that the degree of improvement depends on the fraction of test instances where not all sources are correct (so there is room to improve) but enough are correct to allow a correct fused decision. This will most likely occur when the

[††] An "attractor" is a library type that tends to be used when incorrect type is used.

participating ATRs have different OC sensitivities. Lynch 2003 reports on a particular method of constructing a multi-ATR fuser and demonstrates its strength in dealing appropriately with dependencies on simulated data.

4.5 Bayesian Classifiers

4.5.1 Introduction

Bayesian classifiers are an approach to "feature-level fusion" (FLF). The problem is to make an optimal decision when there are multiple measurements. Although FITE is concerned only with DLF, Bayesian classifiers are closely related to Bayesian DLF. Also, when the input decisions to a DLF are accompanied by confidence estimates (e.g., match scores) the DLF is a form of FLF. There has been considerable work with Bayesian classifiers which may provide insights into DLF, especially with respect to the effects of dependencies in the data.

4.5.2 Shi 2001[13]

Two applications are considered in image segmentation (pixel labeling) where a mutual-information based "generalized correlation" metric[‡‡] shows one problem to have mild dependencies and the other to have much stronger dependencies. For accurate models (50% or more of the data used in training), the benefit of modeling the dependencies (over Naïve Bayes fusion) was greater than the benefit of Naïve Bayes fusion (over the single best source). The examples involve cases where one source is much better than the other source, so the benefit of Naïve Bayes fusion over the single best source is perhaps unusually small. This paper also includes a plot of probability of error (Pe) as a function of correlation and Mahalanobis distance (or classifier performance) for two sources with a known Gaussian joint distribution.

4.5.3 Shi 2001 Figure 7 Additions

Using a similar model to that of Shi'2001 Figure 7, except with N sources and focusing on smaller Mahalanobis distances, the trends are as in Figure 3.

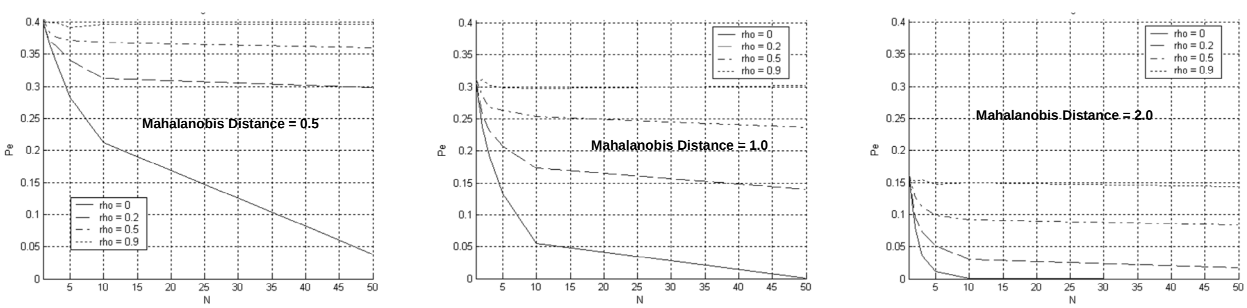

Figure 3 Probability of error as a function of number of sources and source performance

For the two class problem of Figure 3, there are two scalar features. The features for one class (say non-target) have a zero nominal value. The other class (say target) has a nominal value equal to the Mahalanobis distance. Gaussian noise is added to each with zero mean and covariance with 1's on diagonal and rho for all off-diagonal values. Fusion is performed using optimum Bayesian fusion[5, Section 4.2] at min Pe threshold setting with equal priors. High correlation implies less fusion gain, fewer sources implies less fusion gain, and poor source performance implies more gradual, but ultimately more gain.

4.5.4 Domingos 1996[14, 15], J. Friedman 1996[16] and Rish 2001[17]

This work builds on the empirical observation that Naïve Bayes classifiers often perform well in practice, even when the application is known to lack independence. There are two angles to this story. First, optimal performance in terms of the decisions made (called zero-one loss) rather than the computation of the probabilities involved may result even in the presence of certain kinds of strong dependencies. Domingos proves several specific cases of dependencies where Naïve Bayes has minimal zero-one loss. Rish's work suggests that Naïve Bayes does well in the presence of functional dependencies or for low entropy distributions (a strongly dominant class). The second angle is that when the models are

[‡‡] As is used in Drakopoulos'1991 and elsewhere.

not known exactly (e.g., must be approximated from data) the simple Naïve Bayes classifier may compensate for higher bias with lower variance to achieve an overall favorable performance.

4.5.5 N. Friedman 1997[18]

Develops and tests a Bayes classifier with a Bayes Net dependency model. Tests with the UCI[19] data set showed improvements over the Naïve Bayes approach.

4.5.6 Summary

While acknowledging the issue of whether Bayes classifier trends are indicative of Bayesian DLF trends, there are some interesting, if contradictory in some respects, results presented. There are the results of N. Friedman that show benefit from modeling dependences on the UCI data. Shi also presents examples where the benefit of accurate models (i.e., modeling dependencies) is greater than the benefit of Naïve Bayes over the single best source. The Shi examples involve cases where one source is much better than the other source, so the benefit of Naïve Bayes fusion over the single best source is perhaps unusually small.

On the other hand, the results of Domingos, J. Friedman and Rish show that good (or even optimal) performance is possible without modeling dependencies, using again the UCI data set for a demonstration.

Dependencies in the source data is expected to adversely affect the fused performance in two ways. First, because of the dependencies, there is simply less information being provided by the extra sources. Drakopoulis'1991 shows a strong loss in performance as the correlation increases. Others show that some types of dependencies do not cause losses in performance. Second, accurately and tractably modeling the dependencies may not be possible, so inaccuracies in the models may further degrade performance. A small sampling of FLF work was considered in this section in hopes of gaining some insight into the FITE DLF issues related to dependencies in the data.

4.6 Promising papers for further consideration

The survey of DLF through 1992 in Dasarathy'94[6] Chapter 5 reviews several papers, and there are no doubt many since, that deserve consideration for FITE's purposes. We mention a few of those here, using Dasarathy's indices, to indicate the direction of continued work in this survey; not only in the papers of interest, but also in the topics of interest. The Dasarathy'1994 survey represents the field as being mostly concerned with the fusion of geographically separated sources where communication issues are a primary factor. A major thrust of the DLF work of that time, beginning with R81-1, related to adapting the parameters of the fusion contributors to improve the fused system's performance. The paper R87-1 discusses the potential benefits associated with this problem. OC sensitivity and the impact of knowing OCs is part of R88-8. Dependent sources are part of R89-3 and Bayes fusion without independence assumptions in R89-9. R89-13 reports "Linear degradation … when there is data correlation …". R92-2 extends the uncorrelated baseline of R86-1 for correlated sources. R88-6 suggests some issues with serial fusion? R91-8 considers the consequences of inaccuracies in the probability models of a ML fuser.

5. SUMMARY AND CONCLUSIONS

5.1 Baseline analysis

From the baseline DLF problem, with excursions and inferences, we tentatively conclude that DLF is better at rejecting incorrect decisions than in adding correct decisions, DLF gain varies with operating point, more sources is better, higher source performance (e.g., Pid) is better, and larger ATR library is better (for a constant Pid), uncertainty in operating point corresponds to uncertainty in performance, a better source ATR has many mild attractors rather than a few large attractors, and fusion will be more beneficial when there are no dominant sources.

5.2 Dependencies

5.2.1 Dependencies hurt performance, even when modeled well

For all the cases considered in the Drakopoulis'1991 study, by the point of correlation equal to 0.2 there was noticeable performance loss. In the Shi study, when there were accurate models, the benefit of modeling the dependencies (over Naïve Bayes fusion) was greater than the benefit of Naïve Bayes fusion (over the single best source). There are the results of N. Friedman that show benefit from modeling dependences on the UCI data. Finally, in the Bayesian classifier example, high correlation implies less fusion gain in a straight forward way.

5.2.2 Not modeling dependencies does not further diminish performance

Optimal performance in terms of the decisions made (called zero-one loss), rather than the computation of the probabilities involved, may result even in the presence of certain kinds of strong dependencies. Rish work suggests that Naïve Bayes does well in the presence of functional dependencies or for low entropy distributions. When the models are not known exactly (e.g., must be approximated from data) the simple Naïve Bayes classifier may compensate higher bias with lower variance to achieve an overall favorable performance. The results of Domingos, J. Friedman and Rish show that good (or even optimal) performance is possible without modeling dependencies, using the UCI data set for a demonstration.

Note that while this suggests that modeling dependencies may have limited benefit, it does not suggest that additional conditioning OCs in the conditional probability models will not have substantial benefit. The additional conditioning variables may reduce the dependencies themselves, which are believed to have a major impact.

5.3 Other

There is work being conducted, particularly by the US Navy and Army, with parallels to the FITE framework, although the reporting of test results from such work is limited. Distributed fusion is not driven by performance, so centralized fusion is an appropriate focus for FITE. For multi-ATR fusion, the degree of improvement may depend on the participating ATRs having different OC sensitivities. The machine learning literature is especially good at considering the limitations of imperfect (learned in their case) models.

Finally, it is noted that, even with perfect models, the DLF gain may be quite modest and it may be fairly easy (utilizing the baseline analysis with excursions and inferences) to check whether the best possible performance is good enough for a given application.

ACKNOWLEDGEMENTS

The FITE program resulted from AFRL management's (Dr. M. Bryant and Ms. L. Westerkamp) recognition of the fundamental character of decision-level fusion and their willingness to allocate resources to allow its exploration. This paper had many contributions from the rest of the FITE team.

REFERENCES

1. Kahler, B., Blasch, E., and Goodwon, L., "Operating Condition Modeling for ATR Fusion Assessment", SPIE Defense and Security 2007, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications Conference, Orlando, FL, April 2007
2. Hall, D. L. and S. A. H. McMullen, *Mathematical Techniques in Multisensor Data Fusion*, 2nd Ed., Artech House, 2004
3. Duda, R. O., P. E. Hart and D. G. Stork, *Pattern Classification, 2nd ed.*, Wiley Interscience, 2001
4. Morgan, D. R. and T. D. Ross, "A Bayesian Framework for ATR Decision-Level Fusion Experiments," *SPIE Defense and Security 2007, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications Conference*, Orlando, FL, April 2007
5. Melsa, James L. and David L. Cohn, *Decision and Estimation Theory*, McGraw-Hill, New York, 1978
6. Dasarathy, B. V., *Decision Fusion*, IEEE Computer Society Press, Los Alamos, CA, 1994
7. Drakopoulis, E. and C. C. Lee, "Optimum Multisensor Fusion of Correlated Local Decisions", *IEEE Trans. Aerospace and Electronic Systems*, Vol. AES-27, No.4, pp.593-605, July 1991
8. Butler, C. A., J. E. Baker, J. A. Crowe, D. P. Kierstead, and C. A. Mauro, "Combat air identification fusion algorithm (CAIFA), *SPIE Defense and Security 2003, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications Conference*, Orlando, FL, April 2003
9. Chong, C., and S. Mori, "Distributed fusion and communication management for target identification", *Proceedings of the 8th International Conference on Information Fusion*, Philadelphia, PA, July 2005
10. Dietterich, T. G., "Ensemble learning", *The Handbook of Brain Theory and Neural Networks*, Second Edition, 2002

11. Rizvi, S. A., and N. M. Nasrabadi, “Fusion of FLIR automatic target recognition algorithms”, *Information Fusion*, 4, pp.247-258, 2003
12. Lynch, R. S., and P. K. Willett, “Use of Bayesian data reduction for the fusion of legacy classifiers”, *Information Fusion*, 4, pp.23-34, 2003
13. Shi, X., and R. Manduchi, “A study on Bayes feature fusion for image classification”, *2003 Conference on Computer Vision and Pattern Recognition Workshop*, Volume 8, p.95, 2003
14. Domingos, P. and M. Pazzani, “Beyond independence: Conditions for the optimality of the simple Bayesian classifier”, *International Conference on Machine Learning*, pp.105-112, 1996
15. Domingos, P. and M. Pazzani, “On the Optimality of the Simple Bayesian Classifier under Zero-One Loss”, Machine Learning, Vol.29, pp.103-130, 1997
16. Friedman, J. H., “On Bias, Variance, 0/1—Loss, and the Curse-of-Dimensionality”, *Journal Data Mining and Knowledge Discovery*, Vol. 1, No. 1, pp.55-77, March, 1997
17. Rish, I., J. Hellerstein, and T. S. Jayram, “An analysis of data characteristics that affect naive Bayes performance”, IBM Technical Report RC21993, 2001
18. Friedman, N., D. Geiger, and M. Goldszmidt, “Bayesian Network Classifier”, *Machine Learning*, 29, pp.131-163, 1997
19. Newman, D.J., S. Hettich, C.L. Blake and C.J. Merz, UCI Repository of machine learning databases, http://www.ics.uci.edu/~mlearn/MLRepository.html, Irvine, CA, University of California, Department of Information and Computer Science, 1998