ESC-TR-97-137

Technical Report
1051

ATR Performance Modeling and Estimation

D.E. Dudgeon

7 December 1998

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the OSD/DDR&E
under Air Force Contract F19628-95-C-0002.

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This work was sponsored by OSD/DDR&E under Air Force Contract F19628-95-C-0002.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Tutungian

Gary Tutungian
Administrative Contracting Officer
Contracted Support Management

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

ATR PERFORMANCE MODELING AND ESTIMATION

D.E. DUDGEON
Group 401

TECHNICAL REPORT 1051

7 DECEMBER 1998

LEXINGTON MASSACHUSETTS

ABSTRACT

The purpose of this document is to assess the state of the art of performance modeling and estimation for synthetic aperture radar (SAR) automatic target recognition (ATR) algorithms. The study underlying this report is part of the Lincoln Laboratory effort under the OSD ATR Program. The intent is not to produce an exhaustive, detailed, voluminous report describing all ongoing efforts, but rather to capture in a succinct yet essentially complete way the approaches currently in use for modeling the performance of SAR ATR algorithms.

To provide context for this document's assessment and recommendations, a brief discussion of the breadth of ATR problems and some of the resulting technical challenges will be conducted. This discussion will generally be couched in terms of the SAR ground surveillance problem of detecting and recognizing various military vehicles, though many of the statements are easily extended to other sensors and other applications. Basic performance metrics will be defined, and a discussion will ensue of the kinds of questions it would be useful to have addressed by a performance model.

The principal approaches to modeling the performance of SAR ATR algorithms and estimating performance characteristics under a variety of conditions will then be outlined and assessed. The document concludes by recommending certain actions to encourage progress in the development of SAR ATR performance modeling and evaluation tools and methodologies.

EXECUTIVE SUMMARY

The purpose of this document is to assess the state of the art of performance modeling and estimation for synthetic aperture radar (SAR) automatic target recognition (ATR) algorithms. The study underlying this report is part of the Lincoln Laboratory effort under the OSD ATR Program.

There is no single "ATR problem." Even in the smaller domain of SAR ATR, there is a spectrum of problems, from straightforward to challenging. Performance modeling and estimation tools and methodologies can support the development of future military ATR systems by permitting systems designers and analysts to understand the trade-offs among sensor capabilities, computational burden, and performance. Currently such understanding is derived from empirical studies; sensors are built, data is collected and analyzed, and ATR algorithms are developed and tested in the laboratory. This is an expensive and time-consuming process.

The degree of difficulty of SAR target detection and recognition problems is generally determined by the amount of variability that the SAR signatures of the targets, backgrounds, and other objects can exhibit. An ATR algorithm must represent and be able to "reason" about these sources of variability to be effective. This is one of the principal challenges facing ATR algorithm developers today. One way to constrain the combinatorics of this variability is through context, prior information, or data from other sensors or sources. The effective use of context is another principal challenge for ATR developers.

ATR performance is generally measured in terms of a probability of detection, a false alarm rate, and a "confusion matrix," which represents the fraction of time that one target type is mistaken for another. Other metrics for specific applications can of course be useful, and they are usually built upon the foundation of these three. Performance can then be modeled by developing tightly bounded estimates of these metrics as a function of various parameters of the problem (e.g., variability of target signatures, number of target types, complexity of the background, density of other man-made objects, etc.).

To highlight some trade-offs that a performance model should help quantify, some typical system design questions are posed and their importance discussed. The essence of these questions asks how much information is available from the sensor data and other sources and how much of it can be exploited by an ATR algorithm.

Currently there are two principal approaches to the problem of SAR ATR performance modeling, Bayesian probability analysis and information theory, but they should not be regarded as competing with each other. In many ways, they are related. Both stand on the solid mathematical foundation of probability theory, which they use to represent and manipulate uncertainty in their calculations.

In the Bayesian approach, probability distributions are used to represent the variability in target and background signatures. Conditional probability densities play a central role, as does Bayes' Law for relating them. Although the theory is well understood, it has limitations when applied to the problem of performance modeling. To apply the theory, assumptions are generally made (such as the use of Gaussian distributions and independence of information sources) to ensure mathematical tractability, but these assumptions are not always good enough to represent the problem with the desired accuracy.

If the complexity needed to make accurate performance estimates leads to intractable analyses, then one is confronted with turning to Monte Carlo simulation. Given the decreasing cost of computation, this may be a more productive approach as time goes on. However, it is easy to devolve from a Monte Carlo simulation of a performance model to simply running ATR algorithms over large data sets to evaluate performance. The problem with the latter is that it may not lend insight into how an ATR will perform in a situation not represented in the test data.

The information theory approach casts the recognition problem as a communication process. The information about a target's identity is there in the real world, but to get to it, one must observe it through a sensor and its associated processing, just as one must receive a message through a communications channel that may be noisy and lossy. Information theory brings in the concept of entropy and measures of relative information to try to quantify how information and thus performance is lost along the processing chain. As with the Bayesian approach, information theory relies on estimating or assuming various probability distributions to represent uncertainty, and it can suffer from the same pitfalls when assumptions do not match reality closely enough.

In spite of the common foundation based on probability theory, SAR ATR performance modeling approaches exhibit considerable diversity. This diversity stems from choices made in modeling the target signatures, the background signatures, and the confuser signatures.

Based on this assessment of SAR ATR performance modeling and estimation, we make the following recommendations:

- Develop a standard set of performance modeling questions to be addressed by a performance model.
- Develop a standard nested set of ATR problems of increasing complexity to drive development of performance models.
- Continue to collect representative SAR data and analyze it in the context of performance model development.
- Improve the characterization of background and confuser objects.
- Codify, characterize, and assess the utility of contextual information.
- Develop performance models that address the utility of multi-sensor and multi-look integration of information.
- Sponsor and organize an annual meeting on performance modeling and estimation and develop the Virtual Distributed Laboratory (VDL) site into an electronic exchange for ATR ideas and findings.

TABLE OF CONTENTS

1. INTRODUCTION

The purpose of this document is to assess the state of the art of performance modeling and estimation for synthetic aperture radar (SAR) automatic target recognition (ATR) algorithms. The study underlying this report is part of the Lincoln Laboratory effort under the OSD ATR Program. The intent is not to produce an exhaustive, detailed, voluminous report describing all ongoing efforts, but rather to capture in a succinct yet essentially complete way the approaches currently in use for modeling the performance of SAR ATR algorithms.

To provide context for this document's assessment and recommendations, a brief discussion of the breadth of ATR problems and some of the resulting technical challenges will be conducted. This discussion will generally be couched in terms of the SAR ground surveillance problem of detecting and recognizing various military vehicles, though many of the statements are easily extended to other sensors and other applications. Basic performance metrics will be defined, and a discussion will ensue of the kinds of questions it would be useful to have addressed by a performance model.

The principal approaches to modeling the performance of SAR ATR algorithms and estimating performance characteristics under a variety of conditions will then be outlined and assessed. The document concludes by recommending certain actions to encourage progress in the development of SAR ATR performance modeling and evaluation tools and methodologies.

1.1 A SPECTRUM OF ATR PROBLEMS

People sometimes speak of the ATR problem, as if it were a single problem, for example, by asking, "When will the ATR problem be solved?" There is of course not a single ATR problem, but a spectrum of military and civilian problems to which recognition technology can be applied. Some of these problems are being solved with today's technology, but many others are difficult, challenging, and beyond the scope of our current algorithmic capabilities.

In general, there are two basic problems that arise in target recognition, and they are usually analyzed separately. The first is the detection problem: determining if the signature of any target of interest is present in the sensor data. The second is the classification problem: can a target signature be reliably distinguished from the signatures of other targets? In practice, this latter problem also encompasses distinguishing target signatures from those resulting from the background or other non-target objects.

A brief word about terminology: Often in the ATR literature, distinctions are made among the definitions of classification, recognition, and identification. Typically, classification is meant to imply distinguishing between broad classes of targets (e.g., wheeled vs. tracked vehicles, or tanks vs. trucks), recognition is meant to imply distinguishing between target types (e.g., M1A1 vs. T-72 vs. M109), and identification is meant to imply distinguishing individual vehicles (e.g., T-72 serial number A08 vs. T-72 serial number A12). Sometimes the term "discrimination" is used to mean separating target signatures from those that arise from background or other non-target objects. These definitions are not universal, however. In this document, these terms will be used more or less interchangeably to mean recognition unless it is important to make a finer distinction.

Generally, problems in which the targets to be detected have known signatures and are embedded in a background that is reasonably modeled as a stationary random process can be solved. The quality of the solution also depends on the sensor characteristics; there must be adequate contrast between the target signature and the background, and there must be enough sensor resolution so that a certain minimum number (usually on the order of 50–100) of pixels subtend the target signature. For such problems, the template correlation approach (which is akin to matched filtering) can give good results.

In addition to the problem of detecting targets, one often wants to recognize targets by type. In SAR ground surveillance, the problem is usually stated in terms of detecting military vehicles (e.g., tanks, trucks, self-propelled artillery, etc.) and recognizing them by type (e.g., M1A1, T-72, M109, M35, HMMWV, SCUD, etc.). The ability to do this depends on knowing the various target signatures, there being significant differences in those signatures, and having a sensor with enough sensitivity and resolution to be able to sense those differences.

ATR algorithms perform less well when confronted with targets in situations where they can exhibit significant variability in their signatures or are embedded in backgrounds that are not spatially homogeneous and contain many other objects of human manufacture (sometimes referred to as "confusers"). Target signature variability in the SAR ground surveillance application can result from a number of things. The signature will vary as the target presents different aspects to the sensor. Since one does not generally know target location and orientation (often called target "pose") in advance, one must be prepared to detect and recognize targets at all reasonable poses. Targets such as tanks, self-propelled artillery, and missile transporter-erector-launchers (TELs), have large parts that can articulate, altering their shapes and radar signatures. Most military vehicles have smaller parts, such as hatches, doors, and vents, which can be opened. Many can also have items such as toolboxes, fuel drums, cables, shovels, and wheel chocks that can be attached or not. Military trucks may be empty or loaded, and may or may not be covered with a canvas roof. Finally, stationary targets may be partially or completely occluded, camouflaged, or parked close together or close to other objects such as buildings, and some vehicles may be in motion at the moment of observation. All of these changes in configuration or situation, target "state" if you will, result in variations in the target signature. Measuring, storing, and modeling all these variations in combination is an onerous task indeed.

The space of variation for the background is even larger and more difficult to model. The background may include vehicles and other objects that are not targets of interest and for which the ATR algorithm has no template or model. It would be useful algorithmically to be able to ignore background variation, but it does influence target detectability and to some extent interacts with the targets' SAR signatures. Ignoring background variation could lead to reduced performance.

The spectrum of ATR problems can range from finding a known target signature in a well-characterized background to detecting and recognizing a set of highly variable and possibly occluded target signatures in a complex background littered with many man-made objects.

1.2 THE NEED FOR PERFORMANCE MODELING

The term "performance modeling" as used in this document means the ability to predict or estimate the bounds or limits of various performance metrics of ATR algorithms in a given situation. It has also sometimes been called "ATR theory." The goal is not necessarily to predict how a particular algorithm will perform on a particular data set, but rather to understand what, if any, fundamental limits there are to performance in various situations.

As a grossly oversimplified example, consider two targets which are not distinguishable in the data taken by a sensor with given characteristics. It would be fruitless to invest time and money trying to develop an advanced algorithm to separate the two targets given only that sensor data. A measure of the separability of the two targets would help to determine this before making an unwise investment. If it were operationally important to be able to separate the two targets, then this case argues for the development of a new sensor whose data will exhibit some degree of separability.

A good performance modeling methodology would be useful in addressing a variety of questions about ATR performance limits. Answers to such questions would be extremely useful to designers and analysts of future ATR systems, allowing them to understand and make informed decisions about system trade-offs.

In the sections below, we shall first review the principal challenges facing SAR ATR algorithms. Then we shall review some basic performance metrics that form the foundation for analyzing the performance of SAR ATR algorithms. After that, some typical ATR system design questions will be discussed to motivate researchers who hope to develop useful performance modeling and estimation tools. Following that, we review some current approaches to performance modeling and give our assessment. We then conclude with a set of recommendations.

2. THE CHALLENGES OF AUTOMATIC TARGET RECOGNITION

The principal challenge of automatic target recognition is coping with the variability of the signatures of targets, confusers, and backgrounds. This issue is briefly outlined in the paragraphs below as part of the background for the following discussion on performance modeling issues. Not only does signature variability represent a serious challenge to developers of robust ATR algorithms and systems, it also represents a challenge to developers of ATR performance modeling and estimation tools that must be sophisticated enough to capture the complexity of the problem and its attempted solutions.

2.1 TARGET, BACKGROUND, AND CONFUSER REPRESENTATION

The generic ATR problem consists of distinguishing among target types, background, and confusing objects. How each of these classes is represented in the digital world of computation determines in part how successful an ATR algorithm can be. A representation must capture what is known about each class and yet be flexible enough to encompass the signature variability observed in the real world.

2.1.1 Target Signature Variability

In SAR imagery, there are two principal contributors to target signature variability. The first is due to geometry and is easily understood; the second is due to the coherent nature of the radiation being used to illuminate and image the scene.

SAR imaging represents a projection from the 3-D world to a 2-D image. Consequently, as a target or its major parts change their orientation with respect to the sensor line of sight, the shape they project into the imaging plane (often called the slant plane in SAR parlance) will change. (Though the projection is different, this is similar in concept to ordinary optical imaging, where the 2-D shape that an object exhibits in the image depends on its 3-D orientation with respect to the imager's line of sight.)

An ATR algorithm must be capable of representing this geometric variation. Typically in template-based algorithms, this is done by having a set of templates spanning the target orientations and major articulations. In model-based algorithms, 3-D target models with moving parts are used to drive predictions of 2-D signatures based on orientation and articulation hypotheses. Other algorithmic approaches might use other means to capture this geometric variability.

Another cause of signature variation results from the use of coherent radiation (radio waves in this case) to illuminate and image the scene. Reflecting surfaces that fall in the same resolution cell of the SAR image will interfere with one another, causing a grainy appearance known as speckle noise. The pattern of constructive and destructive interference (which looks like a pattern of small bright and dark spots overlaying the image) is remarkably sensitive to the angle of incidence of the incoming radar beam. Changes in this angle of a few degrees, which will not alter the geometry much, will nonetheless cause significant changes in the speckle pattern. A simple algorithm that tries to match a measured target signature to a representation that does not incorporate the variability of the speckle pattern will be susceptible to speckle-induced mismatches. Generally ATR algorithms represent speckle variability by a random process and account for it statistically.

In addition to these two principal causes of target signature variability, there is often variability due to configuration differences, some intentionally implemented by an enemy to alter his signature. Examples include the presence or absence of equipment attached to the body of a T-72 tank, such as rear fuel drums, towing cables, toolboxes, and shovels. Radar scattering camouflage netting is one way of intentionally changing the SAR signature of a target.

2.1.2 Background Signature Variability

The SAR signature of the background can vary dramatically simply because the background itself can vary dramatically. In many ATR algorithms, the background is represented as a random process to capture the effects of this variability. This works reasonably well when the background is stationary, that is, of a consistent quality that doesn't vary statistically from place to place locally around a suspected target. However, for complex backgrounds with many man-made artifacts, such stochastic representations are generally inadequate.

2.1.3 Confusing Objects

Objects that are easily confused with targets of interest, sometimes called "confusers" for short, may need to be represented in an ATR algorithm to help distinguish them from targets. Representing a confuser can be problematic because one may not know enough about a confuser to create an adequate model, and one may not have enough signature samples from the confuser to create an adequate database.

2.1.4 Sensor Effects

SAR sensors themselves can introduce some variability into what might otherwise be a reproducible signature (say from a calibrated test target). Receiver noise and other sources of noise enter the radar signal processing and image formation chain to introduce uncertainties in the resulting SAR images. Generally though, these uncertainties are well modeled probabilistically and are small relative to those introduced by target geometrical variability and speckle.

The calibration (or lack thereof) of the synthetic aperture radar can contribute to signature variability. Ideally, the radar would be calibrated to sense the absolute radar cross-section (i.e., the radar reflectivity) of objects in its field of view. The SAR image is then a map of radar cross-section as a function of azimuth and depression angles. In practice, the radar may not be calibrated or its calibration may drift during a mission. In this case, the SAR image represents the relative reflectivity of objects in its field of view. Measured target signatures may then differ by an unknown gain factor from database signatures or modeled signatures. An ATR algorithm must be aware if the radar is not calibrated so that it can appropriately compare signatures with disparate gains.

2.2 USE OF CONTEXT

Another challenge for automatic target recognition is to exploit information from sources other than the sensor data. This information is generally referred to as "context," and it may take many forms. Currently most ATR algorithms use the sensor data and some form of target signature database (either sets of signature templates or models for predicting signatures) as their sole sources of information. More advanced algorithms attempt to make use of information such as terrain maps, road locations, or even descriptions derived from previous surveillance missions.

2.2.1 Information from Other Sources

Terrain maps and other maps can provide useful contextual information for an ATR algorithm, particularly in the case of challenging backgrounds. To be able to use map information effectively, however, the algorithm must be able to register the map information with the measured sensor data accurately. Map information can take many forms, including images previously collected by the same or other sensors. Maps allow the ATR algorithm to form a "preconceived notion" of what to expect so that it can detect deviations from that expectation.

Operational information, such as "SCUDs like to hide near tree lines" or "SAM sites consist of a radar, a command vehicle, and two or more missile batteries," may influence the strategies that ATR algorithms use to find targets. Prior information that a particular target or set of targets was found earlier in a certain area could be used to focus a search. Knowledge that certain types of vehicles are incapable of traversing certain kinds of terrain may also help reduce the set of hypotheses that need to be considered in recognizing a possible target.

There can be many such sources of contextual information that an ATR algorithm, as part of a future intelligence, surveillance, and reconnaissance system, could and should exploit to improve recognition performance.

2.2.2 Integration of Information

The availability of contextual information, including information from other sensors, will not help solve the recognition problem if it cannot be integrated properly to provide consistent, robust, and correct recognition decisions. The challenge of how to integrate information from disparate sources is an active research area, since the information can potentially be combined at multiple levels, the sensor data level, the feature level, the target level, the situational level, and perhaps others. In some cases, it is straightforward to see how to combine information, especially in the case where two sources are generating a similar type of information, like map data. In other cases, it may not be so obvious.

3. SOME BASIC PERFORMANCE METRICS

In this section, we summarize some basic performance metrics in widespread use in the ATR R&D community. This summary is not intended to be a complete list. Other metrics may also be useful for particular ATR applications, but the metrics discussed below generally comprise a common foundation for performance analysis.

3.1 DETECTION AND FALSE ALARM

To recognize a target of interest, an ATR algorithm must first detect the presence of the target signature in the sensor data. Of course, detection is not always perfect; some targets may not be detected (missed targets) and other objects or parts of the background may mislead the detector, causing it to signal the presence of a target when in fact a target is not there (false alarms). The performance of the detector is generally measured by the probability of detection (P_d) and the probability of false alarm (P_{fa}) or related metrics.

The classical probability of detection is straightforward to define. Given that a target signature is present in the sensor data, the probability of detection is the probability that the detection algorithm will report a detection. Generally, this probability is derived empirically, running data through the algorithm and reporting the fraction of target signatures detected. Sometimes the probability of miss, that is, the probability that given the presence of a target the detector will *not* report a detection, which is equal to $1\text{-}P_d$, will be used to characterize the performance of the detector.

The probability of false alarm is defined as the probability the detector will report a detection given that no target is present. In practice, this probability is often derived empirically by processing data with no target signatures through the detector and counting how many detections are erroneously reported. There is one problem with this definition: it is often difficult to quantify how many discrete opportunities there are to signal a detection erroneously in a given data set, making it difficult to normalize the actual number of false alarms to obtain a ratio that resembles a probability. This problem is often circumvented by quoting a false alarm rate (FAR), e.g., the number of false alarms per unit time or per unit area or per other unit that measures the extent of the data set, rather than a probability of false alarm.

3.2 THE ROC CURVE

ATR detection performance is often summarized as a ROC curve. Here ROC stands for "receiver operating characteristic" (Van Trees, 1968), a term dating back to the early days of radar detection theory. It is a parametric plot of the probability of detection (along the vertical axis) and probability of false alarm or false alarm rate (along the horizontal axis). In its classical form, the ROC curve is swept out by varying the detection threshold, which trades detectability for false alarm rejection capability.

The shape of the curve depends on the signal-to-noise ratio (SNR). In the classic detection problem, the SNR is a measure of the separability between the target-present probability density function and the no-target-present case. Each ROC curve is comprised of operating points with the *same* SNR.

In some ATR applications, the plotting of a ROC curve can be problematic for two reasons. First, there may be a number of parameters analogous in some sense to a threshold that is varied to sweep the ROC curve. Usually all parameters are fixed except for one, and it is used to sweep the curve, or a combined, single parameter is used. Second, it may be difficult to define a parameter analogous to the SNR, because the background against which the target is to be detected is too complex to be characterized simply by a noise power estimate. When ROC curves are produced empirically, one would ideally use a data set with background imagery that is consistent in its characterization. This may be difficult to do in practice, because it may be difficult to characterize the background simply or because of limited available data.

3.3 RECOGNITION AND THE CONFUSION MATRIX

Once the ATR algorithm thinks a target has been detected, it then proceeds to attempt to recognize the target signature. Generally, an ATR algorithm has a database of some sort to remember characteristics of the target signatures for the set of targets it is intended to recognize. Depending on the specific algorithm, this database may take the form of a set of models, a set of templates, or a set of parameters (such as the weights in a neural network). In most real situations, in addition to target signatures the ATR algorithm will be confronted with the signatures of other objects not characterized in its database.

The fundamental measure of recognition performance is a conditional probability. It is the probability that the algorithm will recognize a target of being of type I given that the signature actually comes from a target of type J, Prob($I|J$). If I=J, then the algorithm has made a correct recognition decision. If not, the algorithm has made an error.

These conditional probabilities can be assembled into a matrix, often called a confusion matrix because it shows the likelihood of confusing one target with another. If there were N targets whose signatures are characterized in the algorithm database, then the confusion matrix would have size NxN. (There is no convention on how the conditional probabilities for a given target are organized, by row or by column. Usually confusion matrices are labeled to avoid any ambiguity.)

Often confusion matrices are augmented to make other derived metrics more explicit. For example, it is common to have a "none of the above" or an "unknown" target type. The existence of such a non-target type allows the recognition stage of the ATR algorithm to reject false alarms that may have leaked through the detection stage. However, it is also possible that the recognition stage will fail to recognize one of the target signatures in its database and declare the target to be unknown erroneously. This possibility is indicated in the confusion matrix as a set of conditional probabilities: Prob(unknown|J).

The off-diagonal terms of the confusion matrix can be summed by row or column to give a measure of how often a given target is misclassified or how often other targets are misclassified as the given target. The on-diagonal terms can also be averaged (usually in a weighted manner) to give an overall probability of correct recognition.

In practice, the confusion matrix is often derived empirically. If this is done, it is useful to record the conditional probabilities as ratios (e.g., so many instances of deciding target I given the total number of targets of type J) so that it is apparent how much data went into the estimation of each conditional probability.

4. SOME IMPORTANT SYSTEM DESIGN QUESTIONS

ATR system performance and the numerical values of the metrics designed to assess it vary with the quality of the sensor data, the difficulty of the environment being sensed, and the separability of target signatures in some feature space, both from each other as well as from those of potentially confounding objects. In addition, it depends on the potency of the ATR algorithm being used to extract information from the data. Because of the cost, both in time and money, of building sensors and collecting data and ground truth on targets and backgrounds, ATR system designers need tools that can help them make design trade-offs before the systems are built. Here is where a good performance modeling methodology can help.

Below are some questions that an ATR systems designer might ask and that a performance model might address. This list is not intended to be exhaustive; the reason for its inclusion in this report is to stimulate the imaginations of developers of performance modeling techniques. A useful performance model would be able to address questions such as these.

Are We Getting All the Information Out of the Data? Given the characteristics of a certain sensor and the target signatures it is capable of measuring, one would like to know the best possible detection and recognition performance that can be expected. This is indeed a difficult question to address, because the answer depends on so many things relating to the characteristics of the sensor, the targets, and the background. The answer is valuable because it permits one to assess how close to optimum performance any particular algorithm's performance is. If the best existing algorithm falls well short of optimum performance, then an investment in further algorithm development, particular different approaches, is probably warranted. If, on the other hand, the best algorithm is achieving close to optimum performance but that performance is not good enough operationally, then one could argue that an investment needs to be made in better sensor designs.

How Much Would it Help if We Could Estimate Orientation Better? Targets generally have some set of variables associated with them that can be considered as representing their state. Orientation with respect to the sensor is one of the more obvious ones, but there may be others, such as turret orientation with respect to the body, gun elevation, hatches and doors opened or closed, and the presence or absence of other features that depend on the target type. In some cases the effect on the target signature of a change in one of these variables may be miniscule and may be intentionally ignored, relying on the robustness of the ATR algorithm to recognize the target. In other cases, such as target orientation or turret orientation, the effect on the signature is so profound that it is best to estimate the value of the variable as part of the recognition process. This leads to questions such as: How much performance improvement would result if we could estimate orientation better? Should I try to estimate the value of a variable or try for robustness with respect to its effects?

What's the Added Value of a More Advanced Sensor? Given that the modeled optimum performance is not good enough for a particular application or mission, how would various sensor improvements raise the ceiling on performance? For example, would an increase in sensor resolution by a factor of two permit the probability of detection to reach 0.9? In a suburban environment? How much improvement in target separability, say, between a T-72 tank and an M1 tank, could one expect if a polarimetric SAR sensor were used? Such analyses could be invaluable in driving the design of future sensors.

What's the Added Value of Multiple Sensors? Of Multiple Looks? In a future battlespace with a number of tactical surveillance sensors of different types, it may be fruitful to coordinate data collection on areas of interest. A system designer could ask how much potential improvement would there be in ATR performance if data from multiple sensors could be processed before making decisions. A related question is: How much potential improvement would there be if the same sensor looked at the area more than once, presumably from different vantage points?

What's the Added Value of Context or Additional Information from Other Sources? In the past, ATR algorithm developers have concentrated primarily on detecting and recognizing targets primarily based on the presence of a target signature in the sensor data in comparison with the algorithm's data base of signatures. For reasons mentioned earlier, this is a difficult problem. In practice, other sources of information may also be brought to bear on the recognition problem. For example, when trained image analysts process imagery, they generally know something about the location, the terrain, perhaps the weather, perhaps the recent history of the site, whether targets of interest were detected there previously, and so on. Using such contextual information may help to increase the likelihood that certain types of targets will be found there and reduce the likelihood of others.

Incorporating context into an ATR algorithm is not a solved problem. It would be illuminating to understand how much potential ATR performance improvement would result if context were to be incorporated. Then the payoff of investing effort into designing an ATR system that could obtain and use certain types of contextual information could be quantified.

5. CURRENT APPROACHES TO SAR ATR PERFORMANCE MODELING

There are two principal approaches to the problem of SAR ATR performance modeling in use today, Bayesian probability analysis and information theory, but they should not be regarded as competing with each other. In many ways, they are related. In the subsections below, these approaches are described and their contributions and limitations are discussed.

5.1 THE BAYESIAN APPROACH

The Bayesian approach is based on using probability to model uncertainty in the knowledge of target signatures, background conditions, confuser object signatures, and parameters. This approach has one dominant advantage: the concepts of probability and probabilistic reasoning have a strong mathematical foundation, are generally well understood, and lead to results that are intuitively satisfying. As with most mathematical analyses of real world situations, the Bayesian approach makes some fundamental assumptions, and resulting derivations make other assumptions and approximations along the way. In cases where the analysis fails to mimic reality, it is not the mathematical reasoning *per se* that is incorrect; it is usually the breakdown of an assumption.

A complete review of the Bayesian theory is well beyond the scope of this report, and there are many texts that could be consulted for a detailed review (e.g., Van Trees, 1968). For the sake of context, the fundamental ideas are briefly mentioned below.

5.1.1 Classical Detection Theory

In classical detection theory, the space of target signatures is represented by a probability density function. Often the probability density function is one-dimensional and the independent variable represents some feature of the target signature, such as energy. Similarly, there is a probability density function to represent the signature of the background. Ideally, the feature of the target signature is chosen so that the two probability density functions have minimal overlap. Classically, if the feature value computed for some portion of the sensor data exceeds some threshold, then a target is declared. If the threshold is not exceeded, then no declaration is made. Given this paradigm, it is conceptually straightforward to compute the probability of detection and probability of false alarm (and consequently the ROC curve) as a function of the two probability densities and the threshold.

The situation becomes computationally more complex when the probability densities for the signatures of the targets and the background are functions of several variables. Now, instead of a simple threshold, a surface in multidimensional space must be specified. It may simply be a hyperplane, set of hyperplane segments, or even a more complicated surface used to separate the two probability density functions as much as possible. Given this separating surface, it is still conceptually straightforward to integrate the two probability density functions to obtain the probability of detection and the probability of false alarm, but it can be computationally difficult.

Clearly, classical detection theory depends on being able to obtain, from measurements or first principles, the necessary probability density functions. Often these are assumed for mathematical convenience to be Gaussian densities. If the real world of target and background signatures from a given sensor (or sensors) is not adequately modeled by such a density, then the predictions from classical detection theory can be misleading. For example, real world densities often have heavier tails than the Gaussian density, which drops off very quickly as its argument grows. This can cause one to underestimate the false alarm rate.

5.1.2 Classical Maximum *a posteriori* Estimation

Maximum *a posteriori* estimation is often used to try to distinguish among different target types. If one cares to regard "background" as a target type, then maximum *a posteriori* estimation can be used as a detector.

In classical maximum *a posteriori* estimation, one computes the conditional probability that a certain target type is present given the sensor data. To do this one uses Bayes' Rule to combine certain other probabilities, including the probability of observing the sensor data given that the target was indeed present. In texts this is often written as

$$p(\mathbf{x} \mid \mathbf{s}) = \frac{p(\mathbf{s} \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{s})} = \frac{p(\mathbf{s} \mid \mathbf{x})p(\mathbf{x})}{\sum_{\mathbf{x}} p(\mathbf{s} \mid \mathbf{x})p(\mathbf{x})}$$

where **x** represents the target type and **s** represents the sensor data. Such equations can be written for each target type. To recognize a target type, the equations are used to generate the probability of each possible target type given the sensor data, and the target type yielding the highest value is chosen. Since the denominator on the right side of each equation is the same, it is often ignored when comparing them.

Often a threshold is also used to ensure that the winning target type has a certain minimum probability. It is also possible to use more sophisticated decision logic, for example, insisting that the winning target type enjoy a comfortable margin of victory over the second-place finisher. This is one way to try to reject false alarms or confusers that have been erroneously detected.

There is a well-developed literature (Devroye, 1996) that analyzes pattern recognition from the probabilistic viewpoint, developing, for example, bounds on the

Bayes' error of a two-class classifier given various distance or divergence metrics between posterior densities. In addition, the theory of multiple hypothesis testing is well developed, and methods for probabilistic reasoning as new data become available to refute or support a hypothesis have been developed (Pearl, 1988). However, in many cases it is analytically intractable to find solutions and construct the appropriate metrics, such as the confusion matrix. One is usually forced into numerical analyses and Monte Carlo simulation.

5.1.3 Examples of the Approach

In addition to texts (such as Van Trees, 1968) on classical detection and estimation theory, Fukunaga's book on statistical pattern recognition (Fukunaga, 1972) and its subsequent revision (Fukunaga, 1990) develop much of the foundation for today's research in ATR performance modeling. Most derivations attempt to derive formulas for the Bayes' error under various assumptions, such as assuming Gaussian conditional densities. For example, Fukunaga and Krile published a method to calculate the Bayes' error for two multivariate Gaussian distributions (Fukunaga, 1969). A correspondence item on the probability of error bounds for the M-class problem was also published around this time (Lainiotis, 1971). Research in this vein has continued over the last thirty years, including for example a paper by Garber and Djouadi on the bounds of the Bayes' error for the multiple hypothesis problem (Garber, 1988). Williams and Clark (Williams, 1996) advocated the use of multivariate statistical tools, such as intrinsic dimensionality of data, to help estimate Bayes' error. They used Xpatch synthetic SAR imagery to demonstrate their hypotheses. (Devroye, 1996) sums up much of the theoretical work in this area.

As the general theory of statistical pattern recognition was developing, there were attempts to apply its results to problems in radar signal processing. Clearly, much of the early work on detection and estimation was motivated by radar processing, typically with ground-based radars attempting to detect and track aircraft. As measurements from real radar systems became available, researchers realized that Gaussian models were not adequate for analyzing many of the recognition tasks. For example, a Bayes classifier may be optimal for the case where target and clutter distributions are known and the radar is calibrated, but a sub-optimal linear classifier is more robust to errors in the assumed distributions and does not require the radar to be calibrated (Novak, 1984).

A comparative study specifically for ATR performance evaluation (Williams, 1997) found that nonparametric error estimation techniques are superior to those which use the Gaussian form of the Bhattacharyya metric to select features and to estimate upper bounds on ATR performance. To demonstrate this conclusion, Williams used ADTS data (33 GHz, 1-foot resolution) from a Lincoln Laboratory airborne sensor to analyze the distinguishability of grassy and treed areas in SAR imagery.

Currently, researchers are attempting to develop and analyze more faithful models of real SAR measurements and real ATR algorithms. For example, as part of the MSTAR program, Irving and Washburn analyzed the performance of a Bayesian classifier with orientation uncertainty (Irving, 1996A), using a Gaussian mixture model for the underlying statistics. They found that their results were more accurate when they modeled a sum of lognormal random variables as another lognormal density, rather than invoking the Central Limit Theorem to justify a Gaussian assumption. They validated this result with Monte Carlo simulation. Their analysis resulted in loose performance bounds (as indicated by a ROC curve) when compared to the Lincoln Laboratory ATR algorithm (see The Lincoln Laboratory Journal, 1993) driven by real SAR imagery (Advanced Detection Technology Sensor (ADTS), 33 GHz, 1-foot resolution, HH-polarization, 56 square kilometers of Stockbridge, NY). Their conclusion was that their random process background model "failed to capture the hostile characteristics of actual, heterogeneous clutter (populated with treelines, shadow regions, and man-made discrete objects...)."

Irving, Washburn, and Grimson (Irving 1996B) developed a technique for bounding the ROC performance of a detector that uses patterns of peaks in SAR imagery to detect and discriminate targets from clutter. The analysis assumed sophisticated models for the peak locations based on 2-D Poisson random processes and used a generalized likelihood ratio test to discriminate between targets and clutter. They were able to produce parameterized ROC curves showing the relative merits of using an increased number of predictable peaks per target and reducing the uncertainty in peak location. They concluded that it would not be possible to reach their particular goal of $P_d = 0.9$ at a false alarm rate of 0.001 per square kilometer using only peak location information.

Monte Carlo estimation techniques were used by Horowitz (Horowitz, 1997) to analyze the classification error between two targets at known orientations and locations and to assess ATR performance as a function of SAR resolution. He used a Rayleigh fade model to generate an ensemble of target signature examples from laboratory-made target signatures using a laser (as a scaled radar) and scaled target physical models. Differences between two targets were characterized by a scatterer-to-background power ratio (SBPR), which is simply related to the pixel-by-pixel contrast between the two target images. The relative utility to performance of corresponding pixels with differing contrast was also analyzed, as was the potential payoff of interferometric SAR data.

Earlier, the potential advantages of an enhanced sensor were also examined by DeGraaf (DeGraaf, 1988) for the case of polarization. He analyzed the performance of a variety of polarimetric detectors, including the likelihood ratio test, the polarimetric matched filter, and a polarimetric span detector, using simple statistical clutter and target models. He found that "Based on these theoretical predictions, we conclude that fully polarimetric SAR systems are substantially more capable of detecting man-made targets in clutter than comparable scalar systems."

The work of Shapiro and his students attempts to improve theoretical results (in the sense of making them better reflect reality) for SAR and other sensors. They advocate a first-principles approach to incorporate the physics of the sensor and its interaction with the target. Recently they reported a target detection theory for stripmap SAR (Yeang 1998) that is used to compare a multiresolution detection processor to a whitening-filter SAR processor. The comparison is based on deriving the signal-to-noise-and-clutter ratio (SNCR) from models for the radar signal, background scattering, and receiver noise and uses complex Gaussian distributions at the signal level. Using a Neyman-Pearson formulation, the probability of detection for a fixed false alarm rate is a function of the derived SNCR. They found that the whitening-filter processor was superior, but it is interesting to note that the relative performance of these two detectors depends on the clutter-to-noise ratio. In the clutter-dominant case, the multiresolution processor is far from optimum and the whitening-filter processor has an even bigger advantage.

Grenander and Srivastava (Grenander, 1997) and other colleagues at Johns Hopkins University and Washington University (St. Louis) (Miller, 1997) are attempting to formulate a general approach to ATR and performance estimation based on Bayesian principles that can easily incorporate multi-target and multi-sensor situations. They use "CAD models of targets with attached texture, reflectivity, surface material descriptions. The variability in targets is modeled by group actions (rotation and translation) on rigid templates and, hence, target inference reduces to optimization over these groups." Their analysis includes the use of Hilbert-Schmidt (minimum mean square error) bounds for the target recognition error.

As part of DARPA's MSTAR project on model-driven SAR ATR technology, predicted target signatures are generated for comparison to features extracted from SAR imagery. Because of potential uncertainties in target location, orientation, and configuration, the predicted signatures (in some cases) include a probability density function for the amplitude of each pixel (Keydel, 1998). This information is available to the matching algorithm within MSTAR that assesses how well a prediction based on a target hypothesis matches the reality of the SAR image. Although this is not performance estimation *per se*, it nonetheless requires an accurate analysis of the causes of uncertainty in the signature modeling process. As computational capability and modeling capability grow, we may see a trend develop in which ATR algorithms attempt to model their performance in the operational environment and adjust their parameters accordingly.

In fact, it is worth noting that research by David Doria at Hughes (Doria, 1997) has been progressing in this direction for a FLIR-based ATR algorithm. The user specifies a required false alarm rate, and the algorithm adjusts its parameters, thresholds, and density of extracted features to try to realize that false alarm rate given the current data. This requires the algorithm to have some model of its own false alarm performance.

The literature of Bayesian analysis and statistical pattern recognition in general, as well as its application to SAR and other ATR problems, is extensive. The examples cited above give an interesting but far from complete snapshot of the field today. Occasionally a journal will devote a special issue to ATR (e.g., Lincoln Laboratory Journal, 1993, IEEE Transactions on Image Processing, 1997), and for the last several years there has been an annual conference on Algorithms for Synthetic Aperture Radar, including sessions on performance estimation, as part of the SPIE AeroSense Meeting in Orlando every April. Proceedings from these meetings are also useful for assessing current SAR ATR research activities.

5.1.4 Limitations of the Bayesian Approach

The Bayesian approach requires a complete model of all the uncertainties associated with every variable upon which the final conditional probability density depends. This is both the strength of the approach as well as the source of its weaknesses. Because it is a complete probabilistic model, one can be sure that the answers obtained by using the Bayesian approach are mathematically correct. However, because the approach requires a complete model, it means that assumptions must often be made in the absence of information. The classical example of this is the assumption that a variable is equally distributed over all its possible values when there is complete ignorance of the value or the probability density of the variable.

For example, if we have no knowledge of the orientation of the gun barrel of a tank with respect to its body, we generally assume *a priori* that it is equally likely to be found at any angle. In reality, there maybe several preferred angles (e.g., straight ahead, straight behind, 10 degrees from straight ahead to permit access to the driver's hatch, within 45 degrees of straight ahead in most tactical formations, etc.). However, if we have reason to hypothesize a non-uniform prior probability density for the orientation of the gun barrel, this information is easily incorporated into the Bayesian formality.

In the paragraphs below, we mention some typical places where the probabilistic model used by the Bayesian approach must be completed by intelligent guessing.

Conditional Probability Density Estimation. Clearly, the use of maximum *a posteriori* estimation to determine the likelihood of the presence of one target over another depends on the accuracy of the conditional probability densities involved in Bayes' formula. The variable **s** representing the target signature is a vector of values, potentially a large number of values if the sensor has a resolution high enough to resolve individual features on the target. (This is sometimes called "putting a lot of pixels on target.") Even if **s** is taken to be a smaller set of feature values, rather than the pixel values themselves, its dimensionality can be high, especially if one wants to reliably distinguish targets from one another.

The non-parametric estimation of high-dimensional probability densities from training samples is extremely difficult. Some researchers have claimed "that attempting to estimate [probability density functions] nonparametrically above 5 dimensions is difficult and above 20 dimensions is futile" (Baggenstoss, 1997, after Scott, 1992). Generally, some sort of kernel is used in estimating the density nonparametrically. In a high-dimensional space, this kernel either causes the collapse of the estimate to small regions in the neighborhood of the training samples, or it smoothes all the training samples together so that the details of the underlying density are lost.

Parametric estimation may be possible if there is an understanding of the form of the underlying conditional probability density function, but in general, the accurate estimation of such densities is difficult. One usually compromises the accuracy of such estimates for mathematical tractability so that one can get an answer at all.

Another issue with the estimation of the conditional probability density $p(\mathbf{s}|\mathbf{x})$ for the target signature given the target type is that it may also be a function of other variables that either need to be made explicit or need to be integrated out. For example, a target signature will depend on the type of target as well as the target's orientation with respect to the sensor. Now the conditional density becomes $p(\mathbf{s}|\mathbf{x},\theta)$. It must be multiplied by the prior density $p(\mathbf{x},\theta)$, which represents the fraction of the time we'd expect to see target type $\mathbf{x}$ at orientation θ. When used in Bayes' formula, these densities yield the conditional posterior density $p(\mathbf{x}, \theta|\mathbf{s})$. This may be more information than we really want. If we just want to estimate the target type and don't care about its orientation, then we can integrate over orientation to obtain $p(\mathbf{x}|\mathbf{s})$.

Some variables may be better modeled as parameters with unknown values to be estimated rather than random variables for which a probability density must be assigned. This approach leads one to the topic of maximum-likelihood estimation and its set of related techniques (Van Trees, 1968).

The Problem of Priors. The other term in the numerator of Bayes' formula is the *a priori* probability $p(\mathbf{x})$ of the unknown target type $\mathbf{x}$. This term can bias the computation of $p(\mathbf{x}|\mathbf{s})$ so that target types that appear more often are more likely to be chosen. Because for most problems there is no legitimate reason for expecting one target type to be present more often than another, $p(\mathbf{x})$ is often assumed to be $1/N$, where N is the number of target types for the problem. For certain ground surveillance problems, it may be possible to estimate this prior probability better, because one may have an idea of the proportion of military vehicle types in a given situation from doctrine or other contextual knowledge. Generally though, the problem of priors is that they are unknowable, so one must guess at a reasonable expression for $p(\mathbf{x})$ and hope (or analyze) that the resulting answer is somewhat insensitive to the exact guess used.

For example, consider again the case mentioned above where a target may be of different types and have an arbitrary orientation. To get to our final estimate, we need to make estimates of $p(\mathbf{s}|\mathbf{x},\theta)$ and $p(\mathbf{x},\theta)$. We may be willing to assume that target type and orientation are independent, so that $p(\mathbf{x},\theta)=p(\mathbf{x})p(\theta)$, and we may be willing to assume that the prior density for orientation is uniform, but the point is that we have to make some definitive assumption to be able to apply the Bayesian approach.

The Assumption of Independence. To make expressions mathematically tractable, one often assumes that various random variables contributing to the posterior density are independent of one another. This allows the joint density for those variables to be written as a product of their individual densities. The assumption of independence is often justified, but in some cases, it is not, and it may lead to inaccurate estimates of performance. Generally, independence between two random variables can be used as the limiting case where the value of one variable has no correlation with and conveys no information about the value of the other.

Bayes Error Estimation. Even in the simple case of a two-class recognition problem, estimation of the optimal performance can be challenging. The Bayes error is the probability of making a mistake when using the optimum Bayes decision rule. It depends strongly on the conditional probability densities, as one might guess, and it can be interpreted as the degree of difficulty of the discrimination problem at hand.

It is straightforward to compute the Bayes error given the conditional and prior probability density functions. Generally, one usually has available not these densities but a set of training data from which the densities must be estimated and the optimum decision rule chosen. In this case, it can be shown theoretically that it is impossible to estimate the Bayes error universally well (Devroye, 1996; Irving, 1996A). For any fixed, finite set of training data, there exists a probability density for the data that is sufficiently complex to make estimating the Bayes error from that training set "hopelessly inadequate" (Irving, 1996). A related result shows that for a finite set of training data, the only estimator of the Bayes error guaranteed to bound it from below is the trivial estimator zero.

In spite of these difficulties, it is often possible to estimate an upper bound for the Bayes error. The work of Vapnik and Chervonenkis showed that it is possible to estimate the Bayes error for cases where the decision rule is constrained, for example, to be a linear discriminant (Devroye, 1996; Irving, 1996A).

5.1.5 Alternatives and Extensions

Because the Bayesian approach requires a complete model, which rarely exists in the real world except as an approximation, there have been other approaches that attempt to mitigate the completeness requirement. Generally these have been developed in the context of reasoning systems for "intelligent" algorithms, but some of these ideas may have use in the estimation of ATR performance as well. Here we briefly mention a few of these approaches.

The concept of maximum entropy has been used with success in the field of spectrum estimation and related areas of statistical inference. The basic idea is to assume that any unknown variable is as random as it can be, consistent with the known constraints of the problem. This can be considered as a justification for assuming that an unknown prior density is uniform over the possible values of the underlying variable.

The Dempster-Shafer theory of evidence (Shafer, 1976; Pearl, 1988) is an attempt to introduce the concepts of belief and plausibility into a reasoning system. Belief can be interpreted as the degree to which a particular logical inference can be believed or "proven;" plausibility represents the degree to which the inference is consistent with the know facts. Together they can be considered as a range of probabilities. For example, the hypothesis that a target is a tank may be plausible if one observes a strong scatterer in a SAR image. The belief that the target is a tank may still be low, because one scatterer is not much evidence to support the tank hypothesis. The plausibility of many other hypotheses may be high also.

The Dempster-Shafer theory also includes the concept of unresolved belief. A compound event (such as M1A1 *or* T-72) can be given a measure of belief in addition to the belief given to the individual elements of the event (M1A1 and T-72 in this case).

Probabilistic logic (Nilsson, 1986; Pearl, 1988) can be used in principle to develop constraints on the probabilities of hypotheses whose priors are unknown. Logical relations are used to bound the probabilities so as to not violate either the logic or the rules of probability theory.

Fuzzy set theory has an extensive literature by Zadeh and others (Yager, 1987). Its fundamental concept is different from the ideas underlying basic probability theory. In probability theory, one has the notion of events that occur or not with some uncertainty, measured by probability. The notion of a fuzzy set is one in which membership may have degrees, represented by a continuous variable, rather than a binary variable (an object is either a member or not). One of the concepts that derives from this body of study is the notion of possibility theory (Zadeh, 1978) which tries to measure consistency between a hypothesis and the underlying data.

5.2 INFORMATION THEORY APPROACH

Some researchers have attempted to apply the results of information theory to recognition problems. The use of information theoretic ideas is not inconsistent with the Bayesian approach to recognition discussed above. Quite the contrary—information theory depends heavily on probability theory and Bayesian analysis, but adds to it other interpretations and other analysis tools. In the subsections below, the fundamental ideas are outlined and their application to automatic target recognition is discussed.

5.2.1 Recognition as a Communication Process

Several researchers including (Briles, 1993) have posed the recognition problem as a communication problem. Instead of a target sitting in a background that is imaged by a sensor to produce sensor data, one imagines a message (the ideal target signature), which is to be distinguished from other messages (other ideal target signatures and perhaps that of the background as well). All of the messages are subjected to an imperfect channel, which is characterized by a capacity. The channel conveys the now corrupted messages to a receiver, where decisions must be made about the content of the original message.

5.2.2 Fundamental Ideas

The basic ideas of information theory and their elaboration are contained in several excellent texts on the subject, including (Cover, 1991). The notion of information is quantified by defining it in terms of the negative logarithm of the probability of occurrence of a symbol. (Symbols are used to construct the messages to be communicated.) The logarithm is conventionally taken to have base two, so that information is measured in *bits*. (The term "bit" was originally coined as shorthand for binary digit; in this context, a bit can be thought of as representing a binary decision.)

The entropy, or randomness, of a source is defined as the average information conveyed by the symbols emanating from that source. It is generally symbolized by H and given by the equation

$$H = -\sum_x p(x) \cdot \log_2 p(x)$$

where x is a discrete random variable and $p(x)$ is its probability mass function. The entropy can also be interpreted as the expected value of $-\log p(x)$. A source has its maximum entropy if all its symbols are equally likely.

The concepts of conditional entropy (the entropy of a random variable whose probability mass function is conditioned upon some event) and joint entropy (the entropy of several random variables with a joint probability mass function) are defined in an analogous fashion.

The relative entropy is defined as the Kullback-Leibler distance between two probability mass functions (see Cover, 1991). It can be written as

$$D(p,q) = \sum_x p(x) \cdot \log \frac{p(x)}{q(x)}$$

where p and q represent the two probability mass functions. Note that $D(p,q)$ is not a true distance function in the mathematical sense since it is not symmetric, but it does represent in a single number the extent of the difference between the two probability mass functions.

The concept of mutual information can now be defined. It is the reduction in entropy that occurs when one is presented with additional information. Formally, it can be written as the entropy of a probability mass function minus the conditional entropy of the probability mass function. If the conditioning information (e.g., the value of a related random variable) is significant, then the conditional probability mass function will be significantly "narrower" (i.e., more restrictive) than the unconditional, and the reduction in entropy will likewise be significant. The mutual information between the original random variable and the conditioning random variable will be large in this case. It turns out that the mutual information can also be written as the Kullback-Leibler distance between the joint distribution of the two random variables and the product of the two marginal distributions.

One of the seminal events of information theory was the discovery by Shannon that an imperfect channel could still carry information losslessly (in the limit) as long as its channel capacity was not exceeded. Channel capacity is measured in terms of an information rate in bits per second. If one wishes to transmit more information over the channel, then one must accept that there will be distortions or errors in the data being transmitted. (Compression of data is one way to reduce the number of bits to be transmitted, but that doesn't reduce the information rate, just the redundancy in the data.) The trade-off between transmission rate and the distortion suffered by the data being transmitted can be formalized in a rate distortion theory (Cover, 1991).

5.2.3 Relation to Bayesian Approach

It is apparent that the fundamental ideas of information theory depend heavily on probability. It is no surprise therefore that many of these ideas and the analysis tools derived from them can be related to Bayesian concepts. Mutual information plays a central role in many of the information-theoretic analyses. For example, there should be a large degree of mutual information between ideal target signatures and the target recognition decisions at the output of an ATR. If there isn't, one can ask where the information is being lost. Was it not there in the first place (target signatures too similar)? Was it lost in the channel (the sensing process)? Was it lost in the receiver (the ATR algorithm)?

The so-called theory of types (Cover, 1991) is sometimes used in information theory. In this context, a type is defined as an empirical probability distribution representing the relative proportion of occurrence of each value of a random variable in a given data set. One theorem dictates that the number of types is polynomial in n, where n is the length of the sequences in the data set. Since the number of possible sequences is exponential in n, then at least one type has exponentially many sequences. Results from the theory of types can be applied to the problem of the estimation of the Bayes error in a hypothesis-testing situation.

5.2.4 Examples of the Application of Information Theory to ATR

By making the analogy between a target recognition problem and a communication problem, a Bayesian rate distortion function (Briles, 1993) can be defined as the minimum of the mutual information between the source (target type prior probabilities) and the channel output (target type probabilities conditioned on the sensor data) under conditions where the Bayes risk (average probability of error) is bounded from above. A theorem relates this Bayes rate distortion function and the conventional rate distortion function. A further theorem relates mutual information to the probability of making a recognition error. Briles states that "knowledge of the source-observation mutual information can be used to determine how well a correlation identifier will perform." Briles tests his approach by using Monte Carlo integration to evaluate an integral needed to plot the Bayes rate-distortion function. He compares this to a plot of information rate *vs.* probability of recognition error obtained from 15 ultra-wideband radar range profiles (3 targets at 5 positions each) with added Gaussian white noise at 13 varying strengths. Each range profile consists of approximately 400 uncorrelated samples. The numbers for the measured data plus added noise lie above the derived Bayes rate-distortion curve, indicating that the derivation somewhat underestimates the information rate for a given distortion for these data.

In a series of papers, Garber and Zelnio attempted to make some simple estimates of ATR performance using radar range profiles (Garber, 1997). In the context of a real-aperture radar producing range profiles of airborne targets, they developed the concept of sensor capacity, analogous to the information theoretic channel capacity, to describe a sensor's ability to keep different target signatures distinct. Using target constraints and the sensor capacity, they estimate the intrinsic separability of target signatures, which in turn drives the probability of recognition error. In essence they regard the targets as providing a potentially rich source of information, but to utilize this information to make recognition decisions, one must have a sensor with enough capacity to convey the information to a processing algorithm. To test their estimates of recognition error, Garber and Zelnio compared them to probabilities of classification error computed using a data set of 1512 radar returns for three target aircraft. They found that their error estimates were "quite optimistic" when compared with results from the measured data.

Mutual information can be regarded as a similarity measure between two random variables, and thus it can be used in place of other common metrics to measure degrees of difference. For example, Viola has used the mutual information to register images with one another, including disparate types of images such as a 3-D model with an intensity image (Viola, 1995). For Gaussian probability densities, it is possible to show that maximizing information, minimizing entropy, and minimizing variance are equivalent. For non-Gaussian densities of course, they are somewhat different. The suggestion is that the operation of maximizing mutual information may lead to better performance for non-Gaussian densities than maximizing a likelihood ratio. For establishing performance bounds, mutual information provides another metric with which to measure the difference between two probability density functions.

5.2.5 Limitations of the Information Theory Approach

Because information theory depends so heavily on probability and statistics, it suffers from many of the same limitations as the Bayesian approach. Prior probabilities must be assumed, conditional probabilities must be estimated. If one assumes a particular form for a probability density function so that one may estimate its parameters from measured data, then there is the possibility that the assumed form may not be correct. If one wishes to estimate a density nonparametrically from data, then one will need enough data to estimate any subtleties in the density.

5.3 SUMMARY OF REPRESENTATIVE APPROACHES

In an attempt to summarize the foregoing discussion of the various approaches, a table is included below. Not all references mentioned in the text appear in the table. The entries go through the Bayesian approaches in chronological order and end with the two information theory approaches.

Reference	Estimates	Approach	Data	Comments
Fukunaga 1969	Bayes error	Assumes 2 multivariate Gaussian dist.		
Garber 1988	Bayes error	Assumes M multivariate Gaussian distributions		Reduces M-class problem to pairwise errors
Devroye 1996	Bayes error	Nonparametric estimation of distribution	Assumes training set	Detailed discussion of nonparametric theory
Irving 1996B	ROC bounds	2-D Poisson distribution of peaks	ADTS SAR images of Stockbridge	Clutter model too optimistic
Williams 1997	Bayes error	Nonparametric estimation of distribution	ADTS SAR images of trees and grass	Nonparametric more robust than Gaussian
Doria 1997	False alarm rate	Bayesian from first principles	FLIR	Models false alarm rate to make ATR adaptive
Grenander 1997	Hilbert-Schmidt bounds	Bayesian	Generated from CAD models	General theoretical approach for multi-target multi-sensor
Horowitz 1997	Recognition error, contrast, SBPR	Bayesian, Monte Carlo evaluation	From scaled models, plus Rayleigh fades	Assumes known pose
Yeang 1998	Bayes error, SNCR	Bayesian from first principles		Whitening filter superior to multiresolution processor
Briles 1993	Bayes rate-distortion function, mutual information	Information theory, Monte Carlo integration	15 ultra-wideband radar range profiles (3 objects x 5 positions) plus added white Gaussian noise	Defines a Bayes rate-distortion function between target and signature spaces
Garber 1997	Recognition error	Information theory	1512 radar range profiles of 3 aircraft	Optimistic performance estimates

6. ASSESSMENT

The problem of detecting and recognizing military vehicles using airborne synthetic aperture radar is difficult. The primary source of this difficulty is the degree to which the signatures of the targets, other confusing objects, and the background can vary.

There are several types of activities that go on in the SAR ATR research community, although the lines among them are often indistinct. Much of the work and much of the interest of researchers involves the development of new SAR ATR algorithms in an attempt to extend the domain of application of ATR technology to increasingly more difficult and realistic problems. Other activities support this work, such as SAR data collection, SAR data analysis, synthetic data generation, formal algorithm evaluation methodology and tools, and of course performance estimation and modeling.

To date, most SAR ATR algorithm development research has concentrated on recognizing target signatures using only information derived from the SAR imagery. In concert, performance estimation for SAR ATR has been focused on characterizing the uncertainty and unknowns in target signatures. Little regard has been paid to the problem of characterizing the background, other than by modeling it as a random process with a particular probability density function. Almost no analysis has been made of how much performance could be boosted by the proper use of contextual information.

In spite of a common foundation based on probability theory, SAR ATR performance modeling approaches exhibit considerable diversity. This diversity stems from choices made in modeling the target signatures, the background signatures, and the confuser signatures. For example, one may attempt to model the statistics of the target pixel values, the statistics of the locations of the principal peaks in the SAR image of a target, or the statistics of some other feature or set of features thought to be useful in separating targets from each other and from clutter. In Section 2, the representation of target, background, and confuser signatures was discussed; these representations drive the statistical models used to characterize their variability, and the statistical models in turn drive the details of the performance modeling approach.

If the complexity needed to make accurate performance estimates leads to intractable analyses, then one is confronted with turning to Monte Carlo simulation. Given the decreasing cost of computation, this may be a more productive approach as time goes on. However, it is easy to devolve from Monte Carlo simulation of a performance model to simply running ATR algorithms over large data sets to evaluate performance. The problem with the latter is that it may not lend insight into how an ATR will perform in a situation not represented in the test data.

7. RECOMMENDATIONS FOR FUTURE RESEARCH

In this section, we articulate a set of actions that we feel will further the development of SAR ATR systems that perform well enough to be considered for integration into future military intelligence, reconnaissance, and surveillance systems. We welcome discussion of these points in the hopes that a consensus of action will be reached between DoD funding authorities and the R&D community.

Given the interest among researchers in developing algorithms, it takes some discipline to address performance modeling and evaluation without lapsing into writing recognition code, especially if numerical techniques are necessary to analyze performance. It would be helpful if the community could develop a more or less standard set of performance modeling questions to be addressed by a performance model. The questions raised in Section 4 could be taken as a starting point. An important component of this action is to involve those who understand real military needs and who can extrapolate them into the future. Operational organizations understand their current capabilities and shortcomings in light of their assigned missions, but they may not fully appreciate the possibilities that the emerging ATR technology is opening.

To complement the set of performance questions discussed above, a standard set of SAR ATR problem domains would be useful, for both developing more accurate performance estimates as well as evaluating ATR algorithms. The motivation here is that there exists a wide spectrum of recognition problems, from stable target signatures in a well-characterized background to realistically variable signatures in a complex urban background. Just as ATR algorithms of increasing complexity will be required to handle the more difficult end of the spectrum, we expect that performance models of increasing complexity will be needed to estimate performance in the more difficult cases. A nested set of problems of increasing difficulty would be a useful gauge for developing and understanding the limitations of performance models.

One of the keys to better performance models is a thorough understanding of SAR data. This implies, of course, that significant amounts of representative data be taken so that one can analyze the data and understand its variability. We recognize that data collection is an ongoing activity, and we certainly support its continuation. However, collection must be followed by analysis with an eye toward improved performance models. It would be ideal if adequate data were available not only to DoD contractors and labs, but also to universities and other research institutes.

In most current performance models, the background is characterized as a random process. Various studies have suggested various distributions with estimates of the appropriate parameters. This is adequate for targets placed on a homogeneous background such as a grassy field. It becomes less than adequate for complicated natural backgrounds (mixes of fields, forest, rocks, varying vegetation, and lakes) and possibly hopeless for highly structured, man-made backgrounds. Characterizing the background is more difficult than characterizing known targets, because backgrounds can exhibit wider variability. We recommend that the problem of background characterization be addressed more thoroughly, making use of the data and analyses mentioned above, to enable performance modeling of more realistic situations. A similar argument may be made for the modeling of confuser objects as well, which of course contribute to the complexity of some backgrounds.

If one looks down the road to an operational system using recognition technology, the ATR algorithm will be operating in some context, supporting a military mission, perhaps confirming the locations of suspected targets or searching new areas for evidence of military activity. Contextual information has the potential to improve the performance of the ATR algorithm, for example, by eliminating some hypotheses, reducing the possibility of others, and generally constraining the recognition decisions to be sensible. However, to be able to use such contextual information effectively, it must be represented in a form amenable to supporting the required computations. The term "context" is probably too broad to be useful in helping to design ATR algorithms; it must be broken down into well-defined pieces of information potentially available to the algorithm. When that is done, algorithm designers can begin to understand how (and how much) such contextual information can be used to improve target recognition. We recommend that potentially useful contextual information be codified, and that research into its characterization and utility for recognition algorithms be supported.

One can look forward to the day when a battlefield commander routinely has at his disposal surveillance information from several types of platforms. Recognition algorithms must be developed to integrate this information into sensible recognition hypotheses. A system designer must be able to address the question of how much impact on detection and recognition performance is to be expected from the availability of data from a second (or third or fourth...) sensor. One could similarly ask about the potential utility of multiple looks at a suspected target from the same sensor taken at various time intervals. We recommend that performance models be developed that are capable of estimating the amount of new information available in second or third looks at a suspected target from the same or different sensors so that increments in detection and recognition performance can be accurately assessed.

Finally, the problem of automatic target recognition and developing practical ATR systems, as mentioned earlier, is actually a spectrum of problems, involving different tactics for different applications and situations. Despite progress to date, many of these problems remain to be addressed and solved. Given this state, it is imperative that researchers have access to each other's findings and can interact to establish fruitful lines of inquiry. Over the past several years, the SPIE AeroSense Conference on Algorithms for Synthetic Aperture Radar and the ATRWG (Automatic Target Recognition Working

Group) Conference have emerged as forums for these interactions. In addition, various IEEE conferences have sessions on recognition technology for various applications. Such meetings are of course useful for exchanging ideas and results; however, we feel it would also be useful to have a series of meetings focused on ATR performance modeling and estimation. Therefore we recommend that OSD sponsor and organize an annual meeting on this topic, perhaps in conjunction with ATRWG, and perhaps six months after the annual SPIE meeting. In addition, the VDL (Virtual Distributed Laboratory) ATR site at the Air Force Research Laboratory should also be used as a communications nexus for electronic discussions of developments in recognition technology and a repository for papers on ATR and performance estimation.

In summary, our recommendations are to:

- Develop a standard set of performance modeling questions to be addressed by a performance model.
- Develop a standard nested set of ATR problems of increasing complexity to drive development of performance models.
- Continue to collect representative SAR data and analyze it in the context of performance model development.
- Improve the characterization of background and confuser objects.
- Codify, characterize, and assess the utility of contextual information.
- Develop performance models that address the utility of multi-sensor and multi-look integration of information.
- Sponsor and organize an annual meeting on performance modeling and estimation and develop the VDL site into an electronic exchange for ATR ideas and findings.

REFERENCES

(Baggenstoss, 1997) Paul M. Baggenstoss, "Heuristic Classifier Performance Bounds in High Dimensional Settings," preprint, July 8, 1997.

(Briles, 1993) Scott D. Briles, "Information-Theoretic Performance Bounding of Bayesian Identifiers," *SPIE Automatic Object Recognition III*, Vol. 1960, pp. 255–266, 1993.

(Cover, 1991) Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory,* John Wiley and Sons, Inc., New York, 1991.

(DeGraaf, 1988) Stuart R. DeGraaf, "SAR Image Enhancement via Adaptive Polarization Synthesis and Polarimetric Detection Performance," *Proc. Polarimetric Technology Workshop,* U.S. Army Missile Command, Alabama, GACIAC PR 88-03, August 16–18, 1988.

(Devroye, 1996) Luc Devroye, Laszlo Gyorfi, Gabor Lugosi, *A Probabilistic Theory of Recognition*, Springer, New York, 1996.

(Doria, 1997) David M. Doria, David W. Webster, James D. Leonard, "Results of Adaptive FLIR ATR Tests," preprint dated October, 1997.

(Fukunaga, 1969) Keinosuke Fukunaga and Thomas F. Krile, "Calculation of Bayes' Recognition Error for Two Multivariate Gaussian Distributions," *IEEE Trans. Computers*, Vol. C-18, no. 3, pp. 220–229, March, 1969.

(Fukunaga, 1972) Keinosuke Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1972.

(Fukunaga, 1990) Keinosuke Fukunaga, *Introduction to Statistical Pattern Recognition (2nd ed.)*, Academic Press, New York, 1990.

(Garber, 1988) F.D. Garber and A. Djouadi, "Bounds on the Bayes Classification Error Based on Pairwise Risk Functions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 10, no. 2, pp. 281–288, March 1988.

(Garber, 1997) Fred Garber and Ed Zelnio, "On Some Simple Estimates of ATR Performance, and Initial Comparisons for a Small Data Set," *SPIE Algorithms for Synthetic Aperture Radar IV*, Vol. 3070, pp. 150–161, Orlando, FL, April 1997.

(Grenander, 1997) Ulf Grenander and Anuj Srivastava, "Performance Analysis of Bayesian Object Recognition," Research Monograph (Preliminary version), Division of Applied Mathematics, Brown University, Providence, RI, 1997.

(Horowitz, 1997) Larry L. Horowitz and Gary F. Brendel, "Fundamental SAR ATR Performance Predictions for Design Tradeoffs," *SPIE Algorithms for Synthetic Aperture Radar IV*, Vol. 3070, Orlando, FL, April 1997.

(IEEE Transactions on Image Processing 1997) Special Issue on Automatic Target Recognition, *IEEE Trans. Image Processing*, Vol. 6, no. 1, January 1997.

(Irving, 1996A) William W. Irving and Robert B. Washburn, Alphatech, Inc., "Performance Estimation Theory," WL-TR-97-1021, interim report for period 06/01/95–12/01/96, submitted to Wright Laboratory/AACA, September, 1996.

(Irving, 1996B) William W. Irving, Robert B. Washburn, and W. Eric L. Grimson, "Bounding ROC Performance of Peak-Based Target Detectors," Alphatech, Inc. Report TM-489, November 1996.

(Keydel, 1998) Eric R. Keydel, Stephen Stanhope, Wayne Williams, Vasik Rajlich, Russ Sieron, "Modeling Image and Feature Uncertainties in Model-Based SAR Prediction," ERIM International, Inc., Ann Arbor, MI, preprint.

(Lainiotis, 1971) D.G. Lainiotis and S.K. Park, "Probability of Error Bounds," *IEEE. Trans. Systems, Man, and Cybernetics*, pp. 175–178, April 1971.

(Lincoln Laboratory Journal 1993) Special Issue on Automatic Target Recognition, *Lincoln Laboratory Journal*, Vol. 6, no. 1, Spring 1993.

(Miller, 1997) Michael I. Miller, Ulf Grenander, Joseph A. O'Sullivan, Donald L. Snyder, "Automatic Target Recognition Organized via Jump-Diffusion Algorithms," Special Issue on Automatic Target Recognition, *IEEE Trans. Image Processing*, Vol. 6, no. 1, pp. 157–174, January, 1997.

(Nilsson, 1986) Nils J. Nilsson, "Probabilistic Logic," *Artificial Intelligence*, Vol. 28, no. 1, pp. 71–87, 1986.

(Novak, 1984) Leslie Novak, "On the Sensitivity of Bayes and Fisher Classifiers in Radar Target Detection," *IEEE Proc. 18th Asilomar Conference on Circuits, Systems and Computers*, pp. 367–372, Nov. 5–7, 1984.

(Pearl, 1988) Judea Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, Inc., San Francisco, 1988.

(Scott, 1992) D.W. Scott, *Multivariate Density Estimation*, John Wiley and Sons, Inc., New York, 1992.

(Shafer, 1976) G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.

(Van Trees, 1968) Harry L. Van Trees, *Detection, Estimation, and Modulation Theory (Part 1),* John Wiley and Sons, Inc., New York, 1968.

(Viola, 1995) Paula A. Viola, "Alignment by Maximization of Mutual Information," Massachusetts Institute of Technology, Artificial Intelligence Laboratory, A.I. Technical Report 1548, June, 1995.

(Williams, 1996) Arnold C. Williams and Brandy Clark, "Evaluation of SAR ATR," *SPIE Signal Processing, Sensor Fusion, and Target Recognition V*, Vol. 2755, pp. 36–45, Orlando, FL, April 1996.

(Williams, 1997) Arnold Williams, "A Comparative Study of the Gaussian Form of the Bhattacharyya Metric for ATR Performance Evaluation," *SPIE Algorithms for Synthetic Aperture Radar IV*, Vol. 3070, pp. 258–266, Orlando, FL, April 1997.

(Yager, 1987) R. Yager, S. Ovchinnikov, R.M. Tong, and H.T. Nguyen, eds., *Fuzzy Sets and Applications: Selected Papers by L.A. Zadeh*, John Wiley and Sons, Inc., New York, 1987

(Yeang, 1998) Chen-Pang Yeang and Jeffrey H. Shapiro, "Target Detection Theory for Stripmap SAR using Physics-Based Multiresolution Signatures," preprint, *SPIE Algorithms for Synthetic Aperture Radar V*, Vol. 3370, pp. 646–660, Orlando, FL, April 1998.

(Zadeh, 1978) Lofti A. Zadeh, "Fuzzy Sets as a Basis for a Theory of Possibility," *Fuzzy Sets Syst.,* Vol. 1, no. 1, pp. 3–28, 1978.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
7 December 1998

3. REPORT TYPE AND DATES COVERED
Technical Report

4. TITLE AND SUBTITLE
ATR Performance Modeling and Estimation

5. FUNDING NUMBERS
C — F19628-95-C-0002

6. AUTHOR(S)
Dan E. Dudgeon

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lincoln Laboratory, MIT
244 Wood Street
Lexington, MA 02420-9108

8. PERFORMING ORGANIZATION REPORT NUMBER
TR-1051

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
OSD/DDR&E
3030 Defense Pentagon
Washington, D.C. 20301-3030

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ESC-TR-97-137

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The purpose of this document is to assess the state of the art of performance modeling and estimation for synthetic aperture radar (SAR) automatic target recognition (ATR) algorithms. The study underlying this report is part of the Lincoln Laboratory effort under the OSD ATR Program. The intent is not to produce an exhaustive, detailed, voluminous report describing all ongoing efforts, but rather to capture in a succinct yet essentially complete way the approaches currently in use for modeling the performance of SAR ATR algorithms.

To provide context for this document's assessment and recommendations, a brief discussion of the breadth of ATR problems and some of the resulting technical challenges will be conducted. This discussion will generally be couched in terms of the SAR ground surveillance problem of detecting and recognizing various military vehicles, though many of the statements are easily extended to other sensors and other applications. Basic performance metrics will be defined, and a discussion will ensue of the kinds of questions it would be useful to have addressed by a performance model.

The principal approaches to modeling the performance of SAR ATR algorithms and estimating performance characteristics under a variety of conditions will then be outlined and assessed. The document concludes by recommending certain actions to encourage progress in the development of SAR ATR performance modeling and evaluation tools and methodologies.

14. SUBJECT TERMS

15. NUMBER OF PAGES
48

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Same as Report	Same as Report	Same as Report