ESC-TR-95-097

Technical Report
1030

An All-Neighbor Classification Rule Based on Correlated Distance Combination

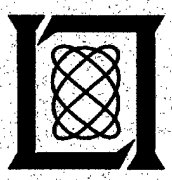
T.P. Wallace

5 November 1996

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS



Prepared for the Department of the Air Force under 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 the U.S. Air Force Space Command, Department of the Air Force, under Contract F19628-95-0002.

The ESC Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

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

FOR THE COMMANDER

Gary Tutungian
Administrative Contracting Officer
Contracted Support Management

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

AN ALL-NEIGHBOR CLASSIFICATION RULE BASED ON CORRELATED DISTANCE COMBINATION

T.P. WALLACE
Group 93

TECHNICAL REPORT 1030

5 NOVEMBER 1996

LEXINGTON MASSACHUSETTS

ABSTRACT

This report describes a new method of classifying data vectors by involving a two-step process. First, a data-specific step produces a "distance" qualitatively describing the similarity of the vector under analysis to each vector in a database representing a particular class. Second, the evidence represented by the vector of statistically correlated "distances" is combined into an overall numerical confidence that the vector under test belongs to the same class as the database vectors. In addition, the supporting evidence is available in the form of the individual distances.

This "all-neighbor" method has several advantages over competing formalisms such as neural networks or the *k*-nearest-neighbor classification method. It can deal with data vectors of varying dimension, as long as the distance measure is capable of comparing them in some fashion. Even more importantly, it can deal with distance vectors of varying dimension, a common situation when dealing with a heterogeneous reference database. It produces a numeric confidence rather than just a simple classification. Further, it uses all the information contained in the distance vector, and it facilitates adjustment of false alarm rates. The method is applied to several different data types to demonstrate its generality.

TABLE OF CONTENTS

	Abstract		iii
	List of Illustrations		vii
1.	INTRODUCTION		1
2.	SINGLE-DISTANCE STATISTICS		5
	2.1	Narrowband Radar Signature Description	5
	2.2	Estimating Single-Distance Statistics	7
3.	MULTIPLE-DISTANCE COMBINATION		11
	3.1	Overview	11
	3.2	Assessing Correlations	11
	3.3	Analytic Model Development	13
4.	SYSTEM PERFORMANCE ASSESSMENT AND OPTIMIZATION		17
	4.1	System Overview	17
	4.2	Performance Characterization	17
	4.3	System Optimization	21
5.	MULTIPLE STATISTICAL MODELS		25
	5.1	Possible Statistical Groups	25
	5.2	Handling Multiple Likelihood Functions	27
6.	GENERAL APPLICABILITY OF METHOD		29
	6.1	C-band Narrowband Radar Example	29
	6.2	Wideband Radar Example	32
7.	SUMMARY AND FUTURE WORK		37
	REFERENCES		39

LIST OF ILLUSTRATIONS

Figure No.		Page
1	Angle definitions.	5
2	Representative signatures.	6
3	Single-distance statistics.	8
4	Estimated S_n likelihood ratios.	12
5	Analysis of a_n for various n.	13
6	Analytic S_n curves.	15
7	System overview.	17
8	System performance before optimization.	18
9	System optimization.	21
10	System performance after optimization.	22
11	Look angle histogram.	25
12	Crossover elevation histogram.	26
13	C-band signatures.	29
14	C-band estimated S_n likelihood ratios.	30
15	C-band a_n for various n.	31
16	C-band system performance after optimization.	31
17	Orion 1 range profile (3-axis stable).	32
18	Galaxy 6 range profile (spinner).	33
19	Single-distance statistics.	34
20	Estimated RP S_n likelihood ratios.	35
21	Analysis of RP a_n for various n.	35
22	RP system performance before optimization.	36

1. INTRODUCTION

There exists a substantial group of pattern recognition problems in which complicated signals are measured, representing multiple classes. For example, consider the problem of categorizing the state of an internal combustion engine by examining an audio recording of its operation. If the analyst has access to the detailed design of the engine, it is possible that the various sounds may be decomposed into frequencies representing motions or vibrations of specific engine parts, which could then be studied at some level. But what if such knowledge is lacking, and instead a database of engine recordings is available, representing possibly a range of engine states from the set (new, worn, damaged)?

Or consider monitoring an earth-orbiting spacecraft using a narrowband radar. Such a radar records an instantaneous total radar cross-section (RCS) value that fluctuates with time as the spacecraft moves with respect to the radar. Again, if the structure of the spacecraft (as deployed) is exactly known, and the orbit and motion of the spacecraft are precisely known, in principle one can predict the signature using electromagnetic prediction methods. If the radar wavelength is large enough compared to the size of the spacecraft, one may even expect a tolerable amount of processing on a large computer. However, in reality one or more of these criteria is usually not satisfied, so the only recourse may be to base one's assessment on a collected database of historic signatures of the same spacecraft or class of spacecraft.

This type of problem has often historically been attacked using nonparametric methods, such as a nearest-neighbor (NN) classifier [1] or neural network, [2] which selects a single class as the most likely. The NN method has been demonstrated as an effective pattern recognition technique in many experiments. An oft-cited reference [3] proved that the probability of error of this method is bounded by twice the Bayes error. However, this is only true asymptotically, in the infinite sample case, so this result is rarely applicable in practice. In practice, an undersampled situation usually exists in which it may be helpful to consider more than the nearest neighbor of a vector. And of course one may be able to do better than twice the Bayes error. The k-nearest-neighbor method [1,4] and its variants [5,6] are sometimes used to give the classifier more information, although both theoretical and experimental work [3,7,8] suggest that this is approach is sometimes less effective than the simple NN.

The neural network is architecturally different from the NN approach, but the performance has proved to be similar in many experiments. It has been shown [9–11] that in the limit, the performance of neural networks approaches the Bayesian limit, which sounds twice as good as the NN situation. However, both these results are for an idealized case of infinite data, and most reported experiments comparing the two methods [8,12–14] show the NN methods to perform slightly better in most cases. The training phase of network construction evidently creates boundaries in hyperspace that are similar to those defined by the NN method. The output is usually comparable to the output of the NN algorithms as well, consisting of a simple classification.

There are difficulties encountered in applying these methods to certain real problems. The confidence in a classification is usually unavailable, except as a statistical performance over an entire ensemble of data. It is always desirable to be able to explain to the user the rationale behind the system's assessment, as is done in some rule-based systems. Explanatory evidence tends to be absent or lacking with these methods, although an NN system can exhibit the nearest neighbor itself as evidence to support the classification. In

either case the user can be provided with an assessment of the overall performance of the classifier based on certain experiments or training runs. It would be better to report the confidence in each classification, however, since certain input patterns can in principle be determined not to match anything in the database well enough for high confidence in the class assignment, while others may generate extremely good matches.

Another common complication in this type of problem involves varying dimensions of the data vectors. For example, what if most of the engine recordings last 10 s, but a small group of potentially valuable examples last only 3 s? What if radar satellite signatures inhabit a total look angle range from 20° 160°, with some short, some long, and some not even overlapping in look angle? The NN and neural network methods normally require input vectors of constant dimension, so some method must be found to meet this requirement.

One additional problem is that there is no guarantee that all the possible classes of the system under analysis are represented by vectors in the database. In fact, many malfunctions probably are not represented. In some cases it may be impossible or dangerous to collect data on malfunctioning systems. Work has been reported [15] on general methods of deciding that a vector does not belong to any class represented in the database or that it is ambiguous. While these methods may indeed be useful, they are mainly applicable to the standard case of vectors in n-space.

An interesting paper by Denoeux [16] describes a method of combining the Dempster-Shafer formalism with the k-NN algorithm. This technique also automatically generates an assessment of the uncertainty in the classification, as expected for Dempster-Shafer methods. This is definitely in the spirit of the current method, but k is fixed, so not all the available information is used. Although large k may be used with this method, in the case of a heterogeneous database in which different input vectors have different numbers of comparable database vectors, it is difficult to select k a priori without sometimes having fewer than k available distances.

In light of the preceding discussion, one can itemize some desirable properties of a system for assessing the state of an object, given a sample vector and a database of historic vectors. None of the methods in the literature seems to have all the following properties, which assume a system based on some kind of direct (NN) or indirect (neural net) comparison of vectors.

1. The comparison algorithm should be able to handle input vectors of different dimension. (Some minimum size may be necessary.) This rules out using a simple feature vector in n-dimensional space, since some of the dimensions may be missing. In fact, while A may be comparable with B, and B with C, it is possible that there is no way to compare A with C. One way to deal with this problem is to work with distances between vector pairs, where possible, rather than with the original vectors.

2. If a large subset of the database is comparable with the new vector, then all the evidence generated from each possible comparison should be used by the system. The NN approach would not satisfy this requirement, but even k-NN methods fail to utilize all the potential evidence. It follows that the system must be able to assess a new vector based on comparisons with a database subset of varying size. It is expected that the confidence in a given

class assignment would be lower on the average when fewer database vectors are available for comparison.

3. To support individual assessments and explanatory evidence, the methods should be either statistical or "fuzzy," generating a continuum of confidences rather than just dividing a pattern space into disjoint regions.

4. The system should not assume that all the possible classes are represented in the database. A "none-of-the-above" hypothesis should exist, or individual hypotheses of particular class membership (or not) should be tested. This requirement is compatible with the desired numerical confidence assessment of point (3).

It is easy to write down this wish list, but it is difficult to develop methods that satisfy each point. A statistical approach is obviously one way to proceed, but the individual distances between pattern vectors are correlated, and the dimension of the necessary joint probability density functions (PDFs) might be as large as dozens or hundreds. Given a database of some hundreds or thousands of vectors, it appears insurmountable to take a Bayesian approach. Even if a few thousand vectors were comparable and available, and hence perhaps ten million distances could be computed and analyzed, it appears intractable to estimate a joint density that might accurately represent the likelihood of a distance vector of several hundred or a thousand values.

In the method about to be described, this problem is approached by first computing a numeric distance between vectors, wherever possible, which is designed to capture their similarity. This is the problem-specific part of the method, and will not be dwelled upon in this report. Nonparametric methods are then used to estimate single-distance PDFs. This procedure is necessary since nonlinear preprocessing is typically used prior to computation of any distances, and sometimes combinatorial methods are used in the distance measure itself, making it difficult to predict the form of the PDFs. The nonparametric estimates show quite a bit of complexity, which further discourages analytic efforts.

Rather than classifying the unknown vector into one of the database classes, the hypothesis that the vector belongs to each possible class is tested. To obtain distances representing the null hypothesis, two methods have been used. The first is to simulate them, as described in Section 2. If simulation is possible, large numbers of simulated distances may be generated to largely solve the problem. If simulation is not possible, a large group of interclass distances may be used to represent the null hypothesis. Given these two groups of distances, a likelihood function giving the probability of class membership may be estimated. Often there is some a priori information about the class of the unknown vector. For example, in the satellite monitoring situation it is fairly certain which satellite is being tracked, so a single assessment of the confidence that the satellite is nominal is produced. If the situation is less certain, multiple confidences of membership in various classes may be produced.

But the single-distance problem is really not the heart of the matter. The difficulty is in dealing with multiple correlated distances. In the course of this research a new technique was developed, combining the evidence for n distances, given fixed n. Further analysis of this statistic for varying n produced a useful analytic expression for evidence combination, which is a function of n. Methods of optimizing the parameters of this expression based on large-scale classification experiments have been used with good success. The result is not only an overall confidence of class assignment, but also an easy recapitulation of the individual

evidence going into the overall assessment, which is very useful to convince a skeptical user that the confidences are correct.

This method is believed to have general applicability. Initially it was applied to narrowband radar signatures of low-earth-orbit (LEO) satellites with both low and high intervector correlations. More recently it has been applied to wideband radar range profiles of geosynchronous satellites, which represent completely different physics, as well as different satellite configurations, operations, and orbits. The technique has worked well in both cases. Work is under way to apply the method to photometric data.

2. SINGLE-DISTANCE STATISTICS

2.1 NARROWBAND RADAR SIGNATURE DESCRIPTION

The method will be presented by working through an entire example, using the narrowband radar signature problem that inspired its development. Previously published descriptions [17] of methods for comparing narrowband radar signatures contain details of the low-level feature extraction steps, so they are not pursued here. A brief description of the data and processing will help motivate the statistical discussion.

The objects of interest are LEO satellites with reasonably circular orbits with apogee/perigee that do not vary too much over their lifetime, for example, remaining within 700 to 750 km. In this case, the orbital geometry as seen by the radar can be characterized by the crossover (maximum) elevation angle of the pass. The signature can be plotted as a function of look angle, the angle between the satellite velocity vector and the radar (viewed from the satellite), as shown in Figure 1. Using look angle gives repeatability of signatures for stable satellites that maintain one or several configurations.

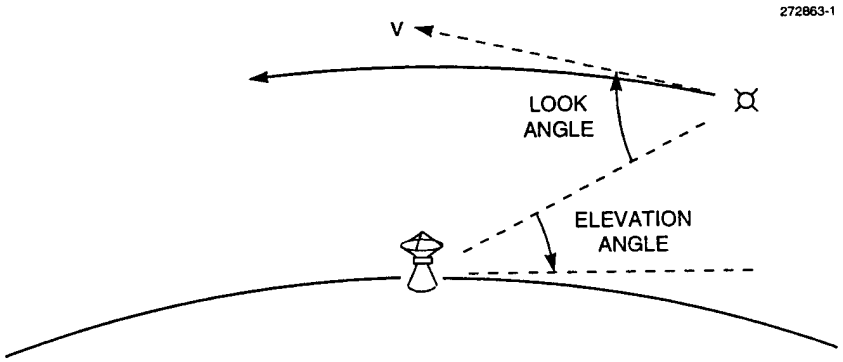


Figure 1. Angle definitions.

The resulting signatures look something like Figure 2, which illustrates several features of the data. The signatures shown are from stable satellites of unknown configuration, but similar pass geometry, which renders them comparable. The abscissa is look angle, while the ordinate is dBsm, i.e., decibels above and below one square meter radar cross section. First, the lengths of the signatures differ. Second, sometimes glitches and dropouts occur in portions of the signature. Third, sometimes substantial similarity occurs in part of the signature, while the rest of it does not match well at all.

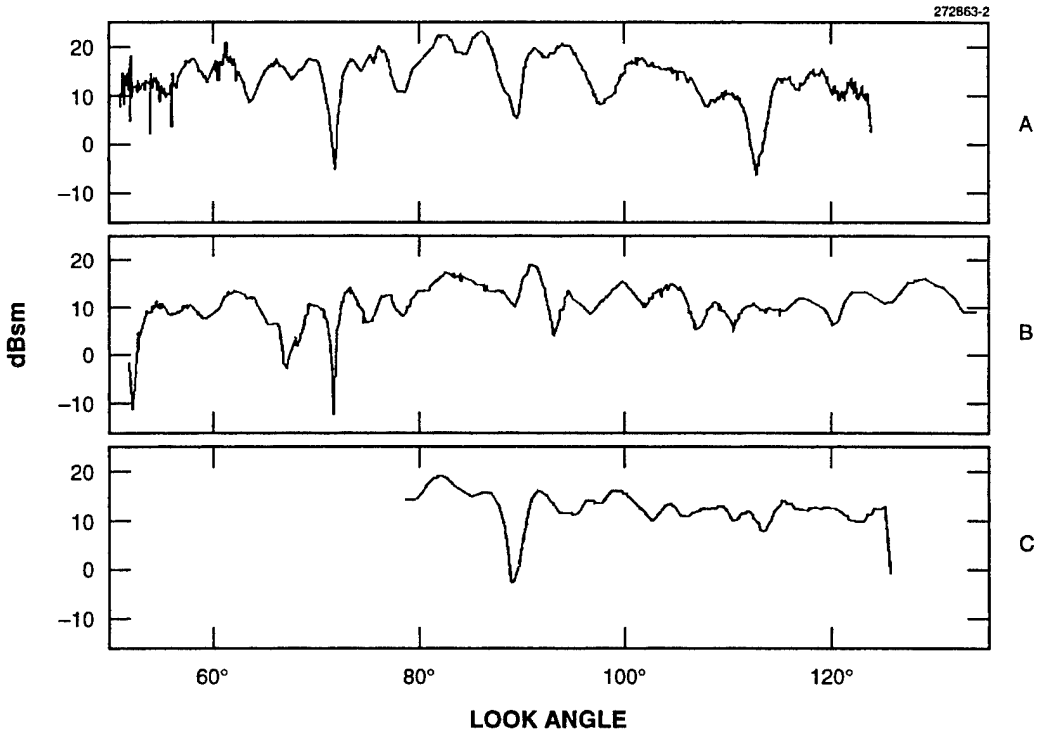


Figure 2. Representative signatures.

Preprocessing to clean up the data is quite important. Conventional linear and nonlinear filtering is performed on the signatures initially. Since the peaks in radar data are known to sometimes fluctuate greatly, they are processed logarithmically (as displayed in Figure 2). Since the nulls are known to be nonrepeatable, nonlinear processing reduces their relative magnitude by processing them linearly, essentially flattening out the signature around zero.

The basic approach to distance calculation from this type of data is to compare two signatures along their range of common look angle, whatever that may be. Within that range, the signatures are chopped up into overlapping segments of about 10°, which are cross-correlated to obtain measures of segment similarity. The correlations and especially the shifts that the correlation process identifies as best aligning the segments are used to compute an overall distance.

The available common look angle ranges from about 30° to 120°. Furthermore, some signatures representing stable satellites, with a similarity that is desirable to capture, show similarity only over part of the signature. For these reasons, the entire region of common look angle is not used to compute a distance, but rather medium-sized chunks of about 35° are used, resulting in multiple "partial" distances to describe the match of each pair of signatures. These chunks are overlapped about 50% to avoid edge effects.

To summarize, after preprocessing two signatures, their common look angle is divided into overlapping segments of approximately 35°. Each of these segments is then divided into small overlapping segments of about 10°, and a partial distance is computed based on cross-correlations of the small segments. A subset of the segments may be used (with distance penalty) so that dropouts or totally bad segments do not destroy the distance.

This procedure was performed on the signatures of Figure 2. The resulting log distance vector between A and B, d(A,B) was –4.79, –2.12, 2.05, where a distance below about 0 indicates similarity, and the –4.79 indicates great similarity. Indeed, the correspondence between the leftmost portions of the signature pair is noticeable, but most of the rest of the signatures look dissimilar. Likewise, d(A,C) is –2.06, a single distance due to the much smaller common look angle. Finally, d(B,C) is 0.80, again reflecting no similarity at all.

Rather than a single distance describing the difference between two signatures, this algorithm outputs distance vectors of varying lengths. The individual distances are of course correlated, since their corresponding vector segments overlap about 50%, among other reasons, seeming to make problems even worse, but the plan is to process many correlated distances anyway downstream. Rather than take any steps to merge distances here, these multiple quantities are treated as separate, correlated pieces of evidence about the status of the unknown signature.

2.2 ESTIMATING SINGLE-DISTANCE STATISTICS

Assume that nothing is known about the satellite class under analysis except that it normally is stable with respect to the earth, and it probably has a limited number of physical configurations. Given a new signature of this class, the question is whether the satellite is still stable (hypothesis H_1) or perhaps has lost stability and is oriented in an unusual way (H_0). Rapid tumbling will normally be evident from the periodic nature of the signature, but the slow tumble case is much more difficult to detect. The database contains numerous signatures representing various physical configurations of the satellite. In this problem, the issue is not determining the configuration of the satellite, just in testing the hypothesis that the satellite is stable.

Given a database of several hundred or even several thousand signatures, potentially many thousands of signature comparisons are suitable for analysis. The basic assumption is that the distance statistics will be such that a stable satellite will have smaller distances than an unstable satellite. The database can be used to determine the single-distance statistics for a stable satellite, but there are two problems with the statistics for the unstable case. First, there are likely to be many fewer examples of known unstable satellites represented in the database. The reasons for this are varied, but it is probably not very useful to collect such data, nor do most satellite classes have many objects that spend much time in an unstable mode. Second, false alarm rates are being measured here, and in reality many *more* distances are needed to get a good estimate of false alarm rates in the important small-distance region.

The solution is to simulate the unstable case by comparing stable signatures from the database in which the geometries are completely different. This simulation can be done by insisting that the crossover elevations of the signatures compared differ by more than 10°, for example, rather than less than 3° (for UHF signatures). The look angles can also be shifted by various larger amounts than the distance measure

can tolerate, simulating a yaw-like instability. Reversing the look angles for one signature simulates flying that satellite backwards, doubling the potential simulated unstable distances.

Given these sample vectors for the hypotheses H_0 and H_1, Bayesian hypothesis testing theory [18] is applicable, using the PDFs $p(R|H_1)$, where R is the measurement, which is in this case a distance. Given a priori probabilities and a cost function, a weighted ratio of these PDFs produces a likelihood ratio, defining a method for choosing between the two hypotheses. In this two-hypothesis case, assume equal a priori probabilities, and a trivial cost function weighting any type of error equally. Then the solution reduces to a simple ratio of the PDFs.

The distance measure previously described is a highly nonlinear calculation, and computing a histogram of the distances illustrates that fact. The top two curves of Figure 3 represent the stable and simulated unstable cases, each of which is the logarithm of a log-distance histogram, after normalization to unit area.

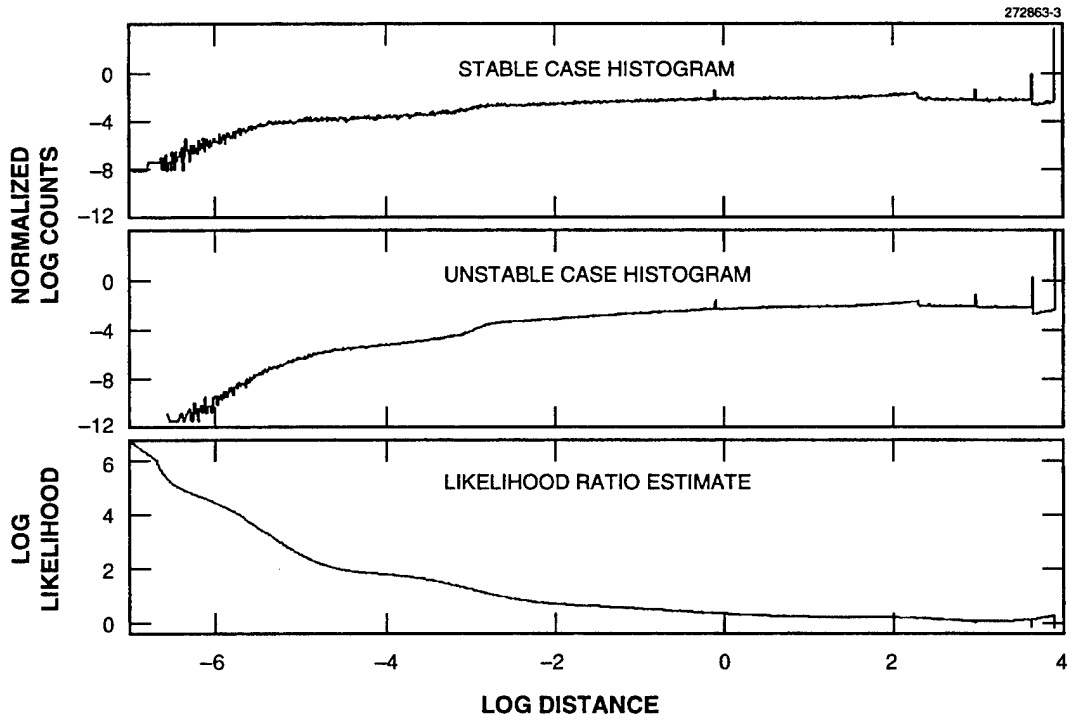


Figure 3. Single-distance statistics.

The discontinuities and spikes are related to several facts. First, the segments must reach their best alignment by shifting look angle within certain limits, and a certain minimum number of segments is needed. If that minimum number does not have the requisite shift, the comparison is considered a failure, and an arbitrary value of 50 is assigned to the distance (log distance of 3.91), explaining the large spike at 3.91.

Second, if 35° is divided into 50% overlapping segments of 10°, 6 segments are obtained. The smallest distances occur when all 6 segments have legal and similar best shifts. When only 5 segments are used, a discontinuity occurs, and similarly for 4, etc.

The difference of these log curves is a likelihood ratio, estimating the likelihood that a single distance represents a stable satellite, Now the data samples suffer from unknown correlations, so it is difficult to predict exactly how many samples will be needed to obtain a given level of accuracy. The solution is simply to use all available stable case distances (a little less than half a million in this instance) and simulate a larger number of unstable distances (a little under two million).

It is clearly necessary to smooth these curves. The discontinuities could cause bias problems if simple filtering is used. Methods of dealing with this problem may be taken directly from spectral estimation [19,20] and involve "trend-removal" or "prewhitening" in which spikes, steps, and possibly polynomial trends are estimated and removed from the data to prevent bias of the other, smaller frequency components. Such methods are used in the current scheme, and the result is a likelihood ratio curve such as shown at the bottom of Figure 3.

A couple of features of this likelihood function estimate may be noted. The spikes were obtained by removing the corresponding spikes from the histograms and processing them separately. The important small distance region is extrapolated out beyond the data (straight line) at a rate hopefully representative of the true function. The rest of the function is based on the difference of smoothed versions of the histograms.

The likelihood function shows that a very small distance generates a rather large likelihood; for example a distance of –5 gives a log likelihood of 2.4, so the odds that this single distance represents the stable case are about $e^{2.4}$:1 or 11:1. A very good match between two signatures is not likely to be coincidental.

On the other hand, a total match failure (distance 3.91) only generates log likelihood of –0.32, giving odds stable of 0.73:1. This is because there are many ways that two signatures can fail to match. For example, signatures from a stable satellite might represent different physical configurations of the satellite, or the state vector describing the sensor viewing might be inaccurate. Clearly, it is essential to combine many such likelihoods to obtain much higher (and lower) confidences.

3. MULTIPLE-DISTANCE COMBINATION

3.1 OVERVIEW

Assume there exist a number of correlated distances, each of which represents evidence that a given vector matches a vector in the database of good vectors. The question is whether to accept H_1, object stable, or H_0, object unstable. Assume that there is some (unknown) intrinsic dimensionality in the data. In the narrowband radar signature example, this might approximate the number of physical configurations of a satellite.

Conceptually, if the number of distances is much less than the intrinsic dimensionality of the data, the correlations between distances should be less; if this number is larger, they should be greater. From this viewpoint, consider the following two-step process for determining an overall likelihood given n multiple correlated distances.

First, compute the n individual single-distance log likelihoods and sum them *as if they represented independent statistics*. The resulting evidence combination S_n will overestimate the total log likelihood, but will be corrected by multiplying by a factor K_n less than unity and depending on n. The overall log likelihood Λ_n will be

$$\Lambda_n = K_n S_n \quad . \tag{1}$$

This severe dimensionality reduction is suboptimal, but it has certain nice properties. All the distances are used, their relative contributions are equally weighted, and the particular pieces of evidence contributing most to the overall assessment are easy to identify. Varying n can be handled, as long as K_n is computed for each n.

To see that this method is suboptimal in general, consider the fact that several large likelihoods combined with several small likelihoods may give the same S_n (and Λ_n) as all medium likelihoods. There is perhaps no reason to assume that the former case should be assessed differently than the latter, but this is possible, and in general will sometimes occur.

3.2 ASSESSING CORRELATIONS

In the single distance case, a nonparametric approach was used to estimate the likelihood functions. No a priori functional forms were used or even hypothesized. Taking this approach further, n can be fixed, and then the statistic S_n is analyzed in the same way the individual distance statistics were analyzed.

A large number of S_n samples must be obtained to represent the two hypotheses. To do this, the same intradatabase distances are used as discussed earlier in estimating the single-distance likelihoods. However, they are now taken in groups of n, so that multiple S_n estimates are obtained. For example, if a certain vector has 12 distances in the database for $n = 5$, then the first 5 are used to make one S_5 sample, the second 5 to make another, and the last two are discarded.

This procedure is used for the distances representing both H_0 and H_1, whereupon the PDFs are estimated using histograms, and then a likelihood ratio Λ_n is formed just as before. While previously rather ugly, spiky, nonlinear PDFs resulted, these S_n histograms are much smoother, as would be expected.

One key point about this procedure is that the data set is reused for all values of n. For example, taking the example 12 distances, if n were 6 rather than 5, the first 6 distances would be taken to form one value of S_6 and the last 6 to form another. These estimates are obviously highly correlated due to this high data overlap, which would ordinarily not be good. However, the hope is to find an analytic expression that usefully approximates these likelihood functions, so how they vary with n is important. Correlated errors may actually make it easier to see this variation, although the absolute errors may be greater.

Figure 4 shows some likelihood functions Λ_n estimated using UHF (430 MHz) narrowband signature distances from a particular class of satellite. For various logarithmically spaced values of n, estimated values of Λ_n are shown. These curves appear to be nearly linear for positive S_n, with varying slopes. The slopes themselves decrease with n, suggesting an intrinsic data dimensionality such that larger numbers of distances are more highly correlated. The curves go reasonably close to the origin; it seems logical that summing inconclusive evidence should produce inconclusive results.

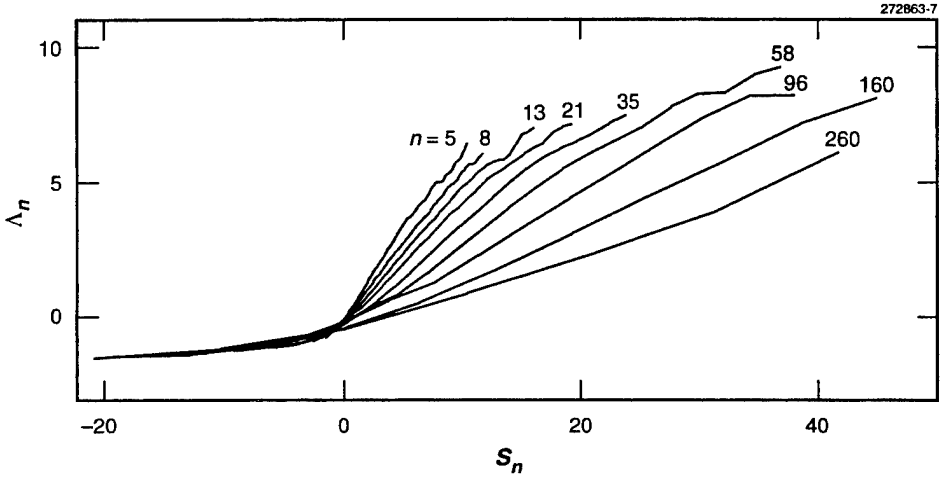


Figure 4. Estimated S_n likelihood ratios.

It is not easy to see major correlated fluctuations in these particular curves, but this effect has been observed in other cases. It is clear that the curves seem to maintain quite uniform spacing, which is likely the result of correlated estimation errors. The large S_n region is not estimated very accurately, since there is a shortage of H_0 data here. Similarly, the negative S_n region has few H_1 samples, and the curves are not very accurate.

A system has been defined that can be used to determine the overall likelihood of the hypotheses given n distances, if n is one of the values of the curves. If not, in principle, curves for all n could be calculated, but this would take quite a bit of time and storage. Interpolation between the curves was done successfully in one system, but that approach has certain problems. Interpolation defines a function of two variables, $F(S_n,n)$, but this will contain certain small fluctuations, in which, for example, increasing S_n might result in decreasing or constant Λ_n.The curve in Figure 4 for $n = 13$ shows this type of effect.

A potentially more serious problem can occur if a system is fielded using this type of statistical model, necessarily containing some maximum n based on the available database at the time of statistical modeling. Even at that time, some vectors being classified will possess more than n associated distances, but not enough for estimating S_n. This problem is exacerbated by a gradually increasing database for systems that automatically do database updates. In this case, extrapolating between the last available cases (160 and 260 in Figure 4) could be very inaccurate, possibly even producing a negative slope. In this case, a very large S_n could generate a negative Λ_n!

A much better approach would be to use an analytic model that:

1. Never has decreasing Λ_n with increasing S_n for any n
2. Provides a smooth analytic function without fluctuations
3. Reasonably fits the existing curves for specified values of n
4. Has few enough parameters to be easily computable and possibly amenable to optimization techniques.

The development of such a model is the subject of Section 3.3.

3.3 ANALYTIC MODEL DEVELOPMENT

It seems reasonable that the curves of Figure 4 might possess a constant slope for fixed n, indicating that S_n should be discounted by this fraction to obtain an overall likelihood Λ_n. In that spirit, Figure 5 shows several steps of heuristic analysis of the slopes of Figure 4.

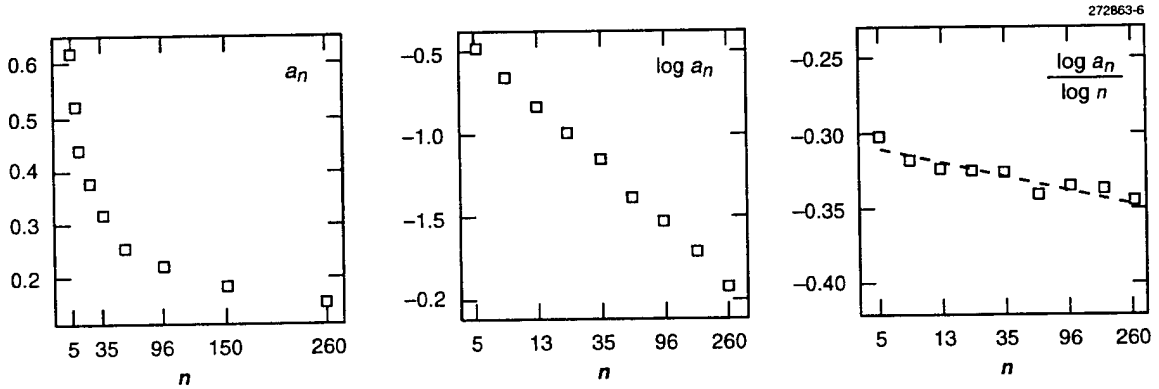


Figure 5. Analysis of a_n for various n.

In the left graph, the slopes $a_n \Lambda_n/S_n$ are estimated by simply using the ratio of the largest y and x from Figure 4 for each curve. The graph is on a linear scale; the annotation follows the specific values of n rather than being uniform. The center graph plots the log (base e) of the slopes in the first graph, and a logarithmic scale is used for the n axis. A nearly linear variation in slope with n can clearly be seen. (There is perhaps a slightly steeper slope for large n.)

In the right graph that slope is divided by log$n$. All the slopes appear within a narrow range of about –0.30 to –0.35, and there is a gradual decrease, on the average. The approximately linear variation of the center graph would produce a one-parameter model, which is certainly convenient. Instead, a line is fit to the right plot, ending up with a two-parameter model, still quite convenient and fitting the data even more accurately.

Mathematically, an analytic expression K_n is sought, which approximates the available a_n, can be computed for any n, and can be inserted into Equation (1) to compute the total likelihood Λ_n. K_n is considered a factor less than unity, which compensates for correlation. Fitting a line to the points of the right graph of Figure 5 gives the expression:

$$\frac{\log a_n}{\log n} = m\log n + b \quad , \tag{2}$$

from which it immediately follows that

$$\Lambda_n = e^{[(m\log n + b)\log n]} S_n \quad . \tag{3}$$

In the example of Figure 5, $m = -0.009$ and $b = -0.297$.

Figure 6 shows the resulting analytic curves for the narrowband radar signature example. These curves can of course be plotted for any value of (n,S_n), but shown are approximately the same values as in the estimated curves of Figure 4. Comparing this plot with the estimated plot, the curve spacing is slightly different, the curves pass through the origin, and the results are much smoother. The essential behavior does seem to be captured, however.

The merit of the system is not reflected in how well these two sets of curves match, anyway. The original estimated curves are not beyond reproach, as they are estimated using limited, correlated data. The real test is how well a real system based on this model performs; there may also be an opportunity to optimize the parameters m and b to improve system performance. These issues are discussed in Section 4.

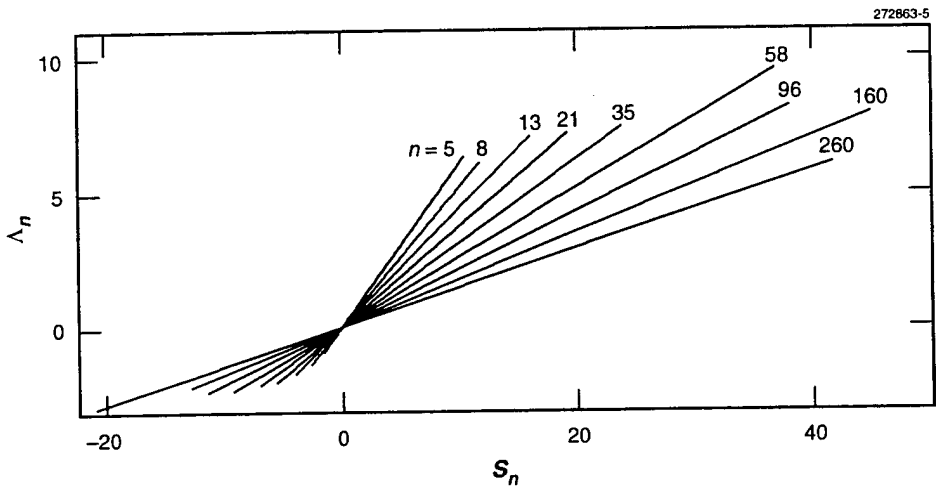


Figure 6. Analytic S_n curves.

4. SYSTEM PERFORMANCE ASSESSMENT AND OPTIMIZATION

4.1 SYSTEM OVERVIEW

It is helpful at this point to recapitulate the steps used in an actual system, as opposed to the steps needed to build the statistical models that go into it. The system itself is rather simpler in operation than in design and can be diagrammed as shown in Figure 7.

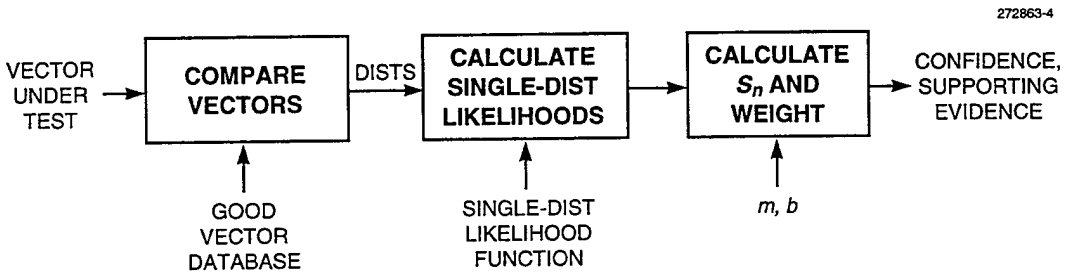


Figure 7. System overview.

The first processing step is to compare a vector under test with appropriate vectors in the database—a major effort and the data-specific part of the algorithm. It is up to the designer to produce an appropriate means of comparing two vectors of this data type, as was done in the narrowband radar signature example. The output of the first box is one or more distances, quantifying similarity between the vector under test and each comparable database vector. The next step is to refer to the single-distance likelihood function, converting each distance into a likelihood. This is generally a simple table lookup, possibly including interpolation.

The final step is to sum the individual likelihoods to form the quantity S_n and then to correct its value using Equation (3) and the provided values of m and b. If the resulting log likelihood Λ_n is, say 3, the odds of hypothesis H_1 are e^3:1 or 20:1, and the reported confidence in this hypothesis is 20/(20 + 1) or about 95%. The supporting evidence consists of the individual distances, which can be sorted by maximum likelihood and then displayed along with the vectors themselves.

4.2 PERFORMANCE CHARACTERIZATION

The best way to characterize the performance of this type of system is to conduct an experiment, classifying a number of good as well as "bad" vectors. The resulting confidences give a good idea of how much information is being extracted from the data. They do not reveal how much information is contained in the data, but they can be useful for optimization and to inform users of the expected performance prior to obtaining experience with the system.

The good vectors may be obtained easily using the "leave one out" method of classifying each one against the rest of the database. The effect of a single vector on the statistical model that is being used is negligible, so the results are not very biased. As always, the bad vectors are a more difficult proposition. In the narrowband radar signature example, simulated unstable vectors that involve elevation differences and/or look angle shifts could be used in an attempt to simulate performance against a satellite that is unstable and slowly tumbling. Instead, signatures will be used of satellites of other classes as the bad case, representing the case in which a satellite is misidentified or a new launch is of an unexpected type. As long as the other classes selected are not extremely similar to the class being monitored, they would be expected to fail the test for hypothesis H_1, stable satellite of the other class.

As an example, 2101 good vectors and 979 bad vectors were classified, which were really from four objects, one each from four other satellite classes. Computing the confidences produced the distribution shown at left in Figure 8. The confidences were divided into the 6 bins shown, each representing the fraction of the total confidences that fell into some confidence range. For example, confidences of 90% to 99% were reported for over 14% of the good signatures but less than 3% of the bad. To the right of Figure 8 is a graphical depiction of the constant K_n, which is used to correct S_n in the confidence-generation code. The values of m and b are those previously obtained by fitting to the curves of Figure 5, $m = -0.009$ and $b = -0.297$.

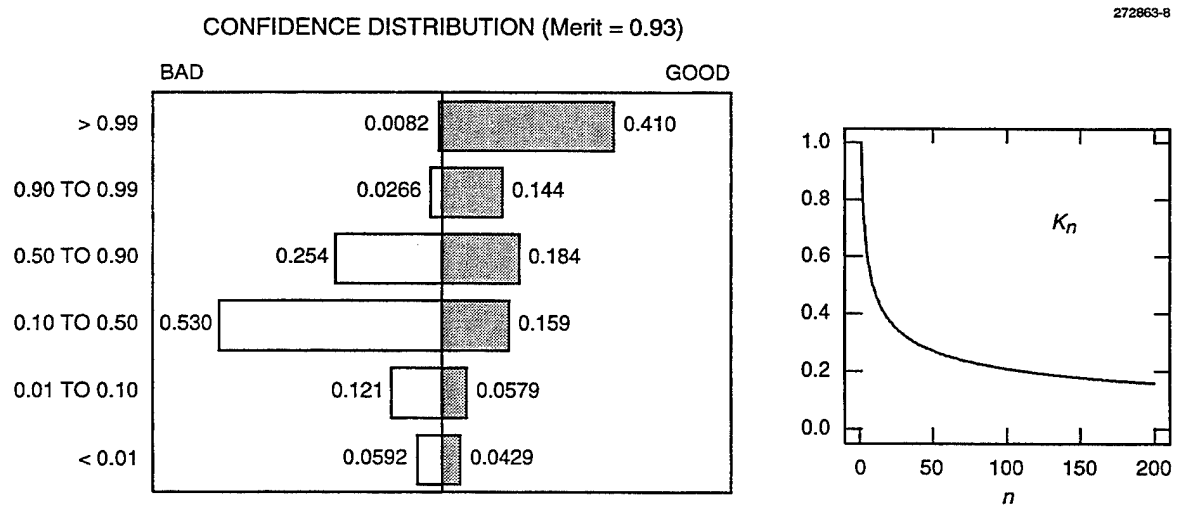


Figure 8. System performance before optimization.

This confidence bar chart may be immediately used to assess whether system performance is adequate to be operationally useful to someone tasked with the job of satellite monitoring. It seems that in over 50% of the cases, a good signature results in a confidence report of greater than 90%, while less than 3% of the bad cases generate such a report. These percentages appear useful, if not much information can be extracted from the signature by other means.

It is very difficult to say whether a report of > 99% confidence should occur 20%, 40%, or 60% of the time, given a good vector. On the other hand, false alarm requirements are known. For example, a very low confidence of less than 1% seems to occur with good data over 4% of the time. This rate is clearly a problem; one would like to calibrate the system so that this occurs 1% of the time or less. Similarly, the fraction of the good confidences located in the 1% to 10% range should be 9% or less. The same thing is true on the other side of the graph. The fraction of the bad confidences indicating > 99% good confidence should be less than 1%, which it is, and the 90% to 99% case should be less than 9%, which it is.

A figure of merit can be formulated, describing the quality of the system using the given statistics and data. Consider the good half of the graph first. The idea is that for log likelihoods small in absolute value, this measure is proportional to the log likelihood. A positive log likelihood boosts the merit, and a negative one diminishes it. However, for very large absolute likelihoods, a law of diminishing returns sets in; the system is not considered much better if it reports a confidence in H_1 for a good signature of 0.99999 than if it reports 0.999. Hence it is desirable to flatten out the incremental merit for very large likelihoods. In the small likelihood region at the bottom of the chart, negative likelihoods are already being added, which is the proper thing to do, but false alarms may be further reduced by providing a larger scaling factor for those negative likelihoods producing very low confidences.

The bad half of the chart can be handled by similar means, but it has been found useful to further correct for false alarms in the high-confidence region, essentially waiting until all vectors have contributed their share to the merit, and then subtracting an additional penalty based on false alarm considerations.

Let λ_i represent the system-calculated log likelihood that the ith vector satisfies H_1, based on some number n_i of distances, not shown notationally. Let $m_i^{(g)}$ ($m_i^{(b)}$) represent the individual merits of the assessments for the good (bad) vectors, numbering N_g (N_b). Let P represent a penalty for exceeding false alarm limits for the bad vectors. Define an overall system merit:

$$M = \frac{1}{N_g} \sum_{\text{good}} m_i^{(g)}(\lambda_i) + \frac{1}{N_b} \sum_{\text{bad}} m_i^{(b)}(\lambda_i) - P \quad . \tag{4}$$

Dividing the λ_i into intervals corresponding to Figure 8, and noting that the confidence corresponding to a given log likelihood is $C_i = e^{\lambda_i}/(1 + e^{\lambda_i})$ the confidences (0.01, 0.10, 0.50, 0.90, 0.99) correspond to log likelihoods (–4.6, –2.2, 0, 2.2, 4.6). With this in mind, define

$$
\begin{aligned}
m_i^{(g)}(\lambda_i) &= 4.6 + 0.05(\lambda_i - 4.6) && 0.99 \le C_i \\
&= \lambda_i && 0.90 \le C_i < 0.99 \\
&= 0.7\lambda_i && 0.50 \le C_i < 0.90 \\
&= \lambda_i && 0.10 \le C_i < 0.50 \\
&= 3.0\lambda_i && 0.01 \le C_i < 0.10 \\
&= 5.0\lambda_i && 0.00 \le C_i < 0.01
\end{aligned}
\qquad (5)
$$

Similarly,

$$
\begin{aligned}
m_i^{(b)}(\lambda_i) &= -5.0\lambda_i && 0.99 \le C_i \\
&= -3.0\lambda_i && 0.90 \le C_i < 0.99 \\
&= -\lambda_i && 0.50 \le C_i < 0.90 \\
&= -\lambda_i && 0.10 \le C_i < 0.50 \\
&= -\lambda_i && 0.01 \le C_i < 0.10 \\
&= 4.6 - 0.1(\lambda_i + 4.6) && 0.00 \le C_i < 0.01
\end{aligned}
\qquad (6)
$$

Denote the lower and upper limits of the kth confidence interval by l_k and u_k, respectively. Let the fraction of the total λ_i falling in the kth interval be f_k. A false alarm penalty p_k is invoked whenever a bad vector has f_k larger than expected in the high confidence intervals [0.90, 0.99), [0.99, 1.00]:

$$
P = \sum_{p_k > 0} p_k \quad , \qquad (7)
$$

where

$$
P_k = \frac{j_k}{u_k - l_k} - 1 \quad . \qquad (8)
$$

The result of this calculation is a numeric characterization of the merit of the system, which can be used to improve performance by optimizing system parameters.

4.3 SYSTEM OPTIMIZATION

4.3.1 Example

In Figure 8 the merit calculated using such a method is reported as 0.93. The values of m and b can be optimized by changing them and then recomputing all the confidences to obtain a new merit value. There are many optimization techniques suitable for maximizing this merit. In this situation, each function calculation is very expensive and may take minutes or even hours for very large databases. A simple scheme was used that takes linear steps tailored to the problem to find a local maximum, followed by binary search to refine the value. This method was applied to the two parameters alternately.

The optimization proceeded as shown in Figure 9. After a brief excursion into disaster, the search rapidly identified a local maximum with about twice the initial merit. It took 24 iterations to verify the maximum of 1.91 within the limits of search termination, but after 7 iterations a value of 1.89 was attained. The computation time on an HP-735 workstation was 97 min, or about 4 min per iteration.

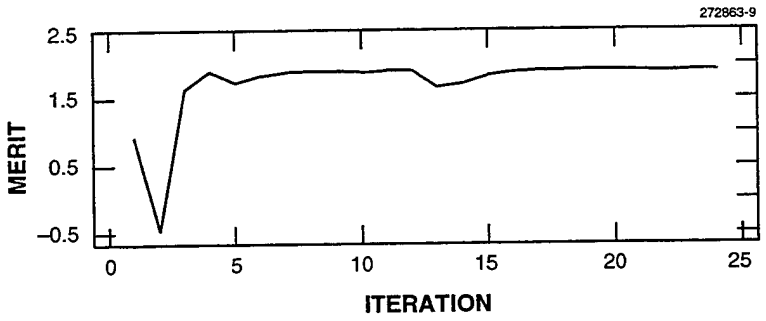


Figure 9. System optimization.

The resulting confidence histogram is shown in Figure 10. Note that the false alarm problem is now corrected. The chance of obtaining a confidence over 90% in the good case has dropped from 55% to 44%, but the false alarm rate in this region for the bad case has dropped from 3.5% to 1%. Similarly, the chance of obtaining a confidence below 10% in the bad case has dropped from 18% to 7.7%, but the "miss rate" for the good case has dropped from 10% to 4.2%, which is more significant. In general, system errors (in which a bad vector is identified as good with high confidence and vice versa) are a much bigger problem than correct assessments that are less conclusive. In the latter case, the user just waits for more data, while in the former the user may be confused.

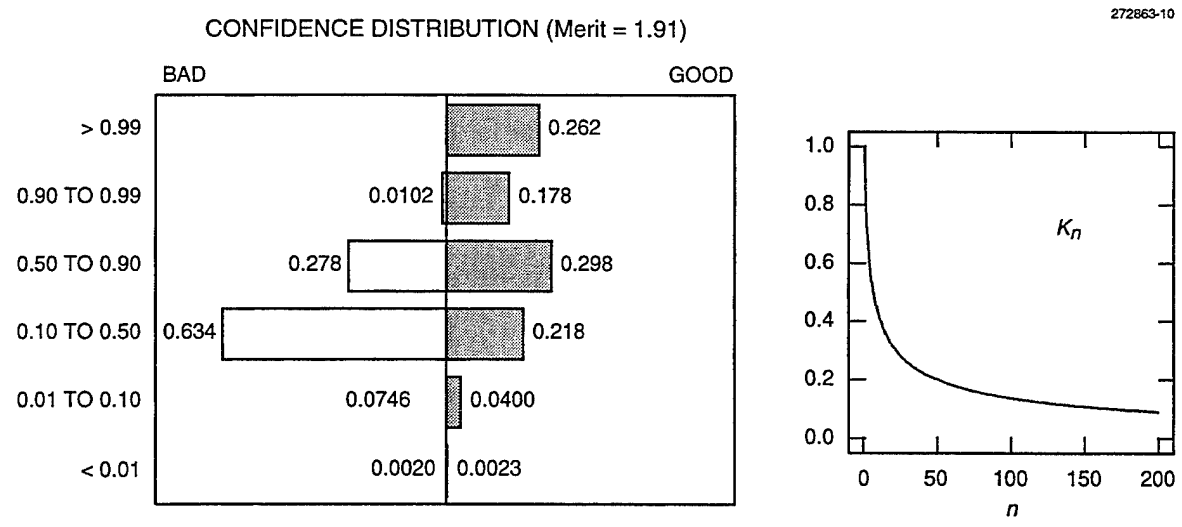


Figure 10. System performance after optimization.

It is clear that the K_n curve drops off more rapidly in the optimized case. The parameters found by optimization were $m = -0.0302$ and $b = -0.2944$. Large likelihoods are being discounted more, so in general the very high and low confidences decrease and the middle confidences increase. Note that the fraction of each case above and below 50% does not change; the correction functions pass through the origin, so the sign of the log likelihoods is preserved.

4.3.2 Parameter Limits

Parameters m and b must have their values constrained to ensure that the exponential of Equation (3) remains between one and zero. It is easy to see that m must be negative in order for this to be true for large n. Given a negative m, b can be slightly positive, up to m log2. These limits should rarely be challenged when fitting to experimental curves such as those of Figure 4. However, optimization, particularly when working with small amounts of data, may exceed these limits.

It can be very expensive in both CPU time and memory to calculate S_n curves such as those shown in Figure 4. A viable alternative is to go directly to the optimization step, once the single-distance likelihood function has been calculated. If reasonable starting values are used, this alternative has proven effective.

4.3.3 Optimizing Other Parameters

A method was shown of optimizing the very important parameters m and *bright* are at the heart of the method for combining correlated evidence. At the end of this process the optimized merit can be viewed as

a numeric representation of the amount of information the system is extracting from the data. This merit can be used to assess or optimize any other parameter in the algorithm in the same way.

For example, certain decisions are made in the estimation of the single-distance likelihood function with regard to smoothing, etc. These decisions can be assessed by simply trying several different values, and comparing the merits found after optimizing m and b in each case. The m and b values may be different if the likelihood estimation changes, but that is not important. The important thing is whether the final optimized merit is significantly different. The same method is applicable to the definition of the actual distance measure itself; the effect of varying distance measure parameters can be easily assessed.

Substantial computation may be associated with these optimization techniques, particularly if a full-fledged optimization loop is used to vary a front-end parameter such as used in the distance measure. Calculations may be performed for all the good distances, simulated bad distances, likelihood functions, and optimized m and b parameters for each iteration. This process can easily take many hours per iteration. On the plus side, however, it is a fully defined algorithm that requires no human intervention. In some cases, such optimizations have run for days or weeks, when the CPU time was available.

5. MULTIPLE STATISTICAL MODELS

5.1 POSSIBLE STATISTICAL GROUPS

To this point, the situation looks rather homogeneous; a large group of vectors is compared one-by-one with appropriate database subsets. All that is known about each vector is that it is from a particular satellite type, so all the vectors are thrown into one statistical class, yielding a single likelihood function, and single subsequent parametric method of evidence combination. In practice, things are often not this uniform, so multiple statistical classes may be required. Since the division into statistical classes is obviously very application-dependent, the signature monitoring example will be revisited.

Figure 11 shows the common look angle extent of the 2101 UHF narrowband radar signatures classified in the optimization experiment of Section 4. A minimum of 30° of look angle is required for processing; those vectors with less are rejected. At the low end, a single distance is computed between each signature and its potential matching signatures from the database. At the high end, several overlapping partial distances are often produced. It is possible that dividing the signatures into multiple categories based on look angle extent could improve performance. The likelihood function and (m,b) pair used for each group might be designed to better compensate for a possibly slightly higher distance correlation due to the increased overlap with increased available look angle.

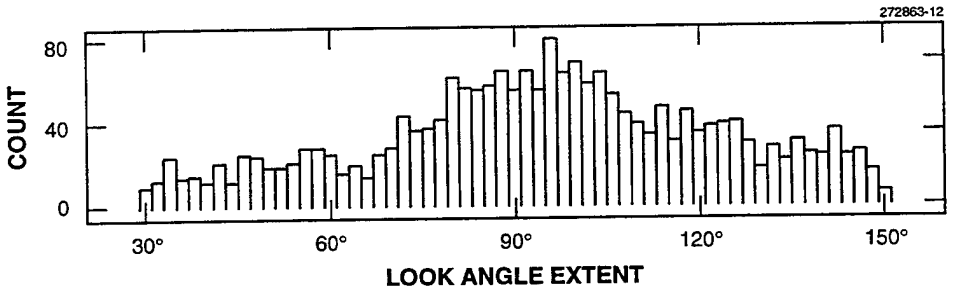


Figure 11. Look angle histogram.

Or consider Figure 12, which shows the maximum elevation at crossover for the example 2101 signatures. This view of the database is even more interesting, since two signatures must have crossover elevations within 3° or so to be compared. The bars are 1° wide so that when a signature with crossover elevation of around 10° is compared with the database, it may find as many as 400 or 500 possible matching signatures. Clearly, the large n region of the classification system is being exercised. On the other hand, a signature over 50° or so in crossover is probably going to be compared with about 20 to 30 other signatures, a smaller value of n.

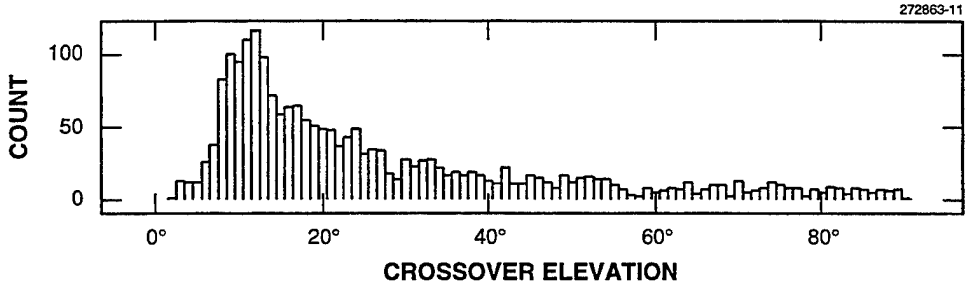


Figure 12. Crossover elevation histogram.

There is reason to suspect that the statistics of the distances might depend on crossover elevation to some extent. For example, consider a satellite with a large, relatively flat surface that always nearly faces the earth. When viewed from a low elevation, this structure may be viewed edge-on, and it may have small RCS, so the signature may be dominated by other returns. At high elevation, that same structure is viewed broadside and may give large lobes that dominate the signature. Depending on the characteristics of this particular structure relative to the rest of the satellite, dividing the signature distances into two groups based on crossover elevation might improve statistical modeling, and hence overall system performance.

Other minor factors could be considered, such as which sensor produced a signature, or whether the crossover elevation difference is very small or nearer to the 3° limit. But the most important factor by far involves using a priori knowledge about the satellite configuration. If it is known that a satellite is in certain configurations at certain times, the distances can be divided into two categories, same configuration and different configuration.

For example, many sun-synchronous satellites, such as the French SPOT series or the Canadian RADARSAT, maintain their orbital plane at a relatively fixed orientation to the sun. They typically use a solar panel with a single degree of freedom, which rotates once during each orbit of about 100 min. It is easy to predict the position of such a panel and characterize the configuration of the satellite at a given time with one parameter, a solar panel angle. It is logical to expect that when the solar panel configurations are nearly the same, the signature similarities would be much greater, and different statistics would be appropriate.

In summary, the larger the database, the more data subsets can be evaluated. If the database size is marginal, it is too hard to estimate the statistics of subsets, and it is best to just combine everything together. As more data becomes available, this decision can be reevaluated.

5.2 HANDLING MULTIPLE LIKELIHOOD FUNCTIONS

Assume the ability to divide the distances computed between a vector under analysis and the database into more than one group. Clearly, the first step is to estimate single-distance PDFs and likelihood functions for each group. The question, then, is how to generalize the system depicted in Figure 7 to handle this model.

This generalization has been done in two different ways, and in one case it was done both ways for the same data set. First, if there is a marginal amount of data, yet confidence is high that the data grouping is reasonable, the single-distance PDFs can be estimated and verified that they are distinct and reasonable. For example, the same-configuration case should have a larger number of small distances. The likelihood functions can then be computed.

A small data set is insufficient to estimate separate values of m and b for the two cases. What can be done is to just form a single hybrid S_n by summing the likelihood for each distance by looking it up in the appropriate likelihood function. The value of n is then the sum of all the distances, and it is corrected by Equation (3) using a single (m,b) pair. The individual distances are assessed more accurately, but the correction for correlation seems a little crude. We call this method early combination.

With a somewhat larger database, separate (m,b) values can be calculated for each statistical group, say m, b, m' and b'. Distances are split into two groups, ending up with two log likelihoods, representing assessments of the unknown vector in terms of multiple disjoint sets of database vectors. The problem is how to combine these assessments; again, these are correlated, so simple addition of log likelihoods should overestimate the total likelihood.

One possibility is to weight them in some fashion and then add. This weight might depend on the fraction of distances that went into each assessment, for example. Experimentally, the best technique is to simultaneously optimize the four parameters m, b, m' and b', where the overall confidence is obtained using simple addition of the separate log assessments. In this way the optimization process tries to adjust the weights to properly assess the vectors. It will simultaneously try to correct for the simple addition overestimate. We call this method late combination.

For one class of satellites with a physical configuration that can be represented by a single angle, the distances were divided into "same" and "different" configuration groups, and both early and late combinations were tested. The numeric merit value indicated that the late combination system extracted significantly more information from the data. The early system was usable, but the late system was clearly better.

In another case, an extremely large database was divided into 16 distance classes. This division was based on 4 independent tests: same/different object, close/far crossover elevation, same/different configuration, and short/long common look angle. This processing used 8 different likelihood functions to compute the S_n values, and then combined 2 such values with the aid say m, b, m' and b' values.

Multiple hypotheses about the input vector can be tested when multiple statistical classes are identified. For example, in the previous large-database situation in which 16 classes were identified, the object, elevations, and look angle situations are known, but the physical configuration of the unknown satellite is in general not known. This sounds like a problem: which PDFs should be invoked to properly obtain the individual distance likelihoods?

Actually, this is an opportunity to extract more information from the system than just whether the satellite is stable and in some normal configuration. Several hypotheses were tested in which H_k represents the hypothesis that the satellite is stable and in configuration k. The hypothesis that generates the largest likelihood is accepted, if it exceeds some absolute threshold (indicating good confidence stable) as well as some threshold relative to the next-largest likelihood (indicating an unambiguous configuration determination). Whenever the database signatures' configuration can be validated (by requiring internal consistency, for example), this type of a system can be constructed.

6. GENERAL APPLICABILITY OF METHOD

6.1 C-BAND NARROWBAND RADAR EXAMPLE

Thus far, the method of evidence combination presented has been based on a single data set, that depicted in Figure 5. In fact, this method has been used on a wide variety of data, some of which is presented in this section. Both the sensor and the satellite class is varied, illustrating the range of problems to which this method has been applied.

The first example considers another narrowband radar data set, but this time the radar frequency is C-band (5.6 GHz) rather than UHF. The character of the data shown in Figure 13 is significantly different from the UHF data previously considered (Figure 2). In addition, a satellite class with much greater radar signature repeatability is considered. The similarities between the two signatures of Figure 13 are great; they were taken on the same object three days apart.

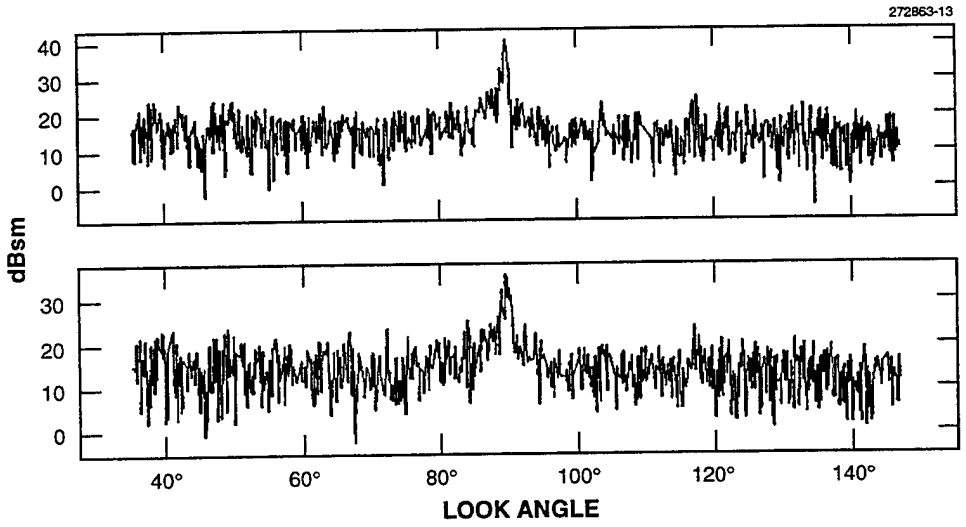


Figure 13. C-band signatures.

It seems that algorithms for automatic signature comparison that are designed and tuned for UHF data would need substantial revision to deal with the rapid variation of the C-band data, but such is not the case. The same signature comparison code performs usefully, outputting partial distances characterizing local signature similarity. Slightly different parameters are ordinarily used, but even the same parameters give reasonable performance.

When the single-distance PDFs (not shown) for this class of object are estimated, they show much greater difference between the H_0 and the H_1 cases than the previous example. This is because the satellite

has fewer configurations and affects UHF data from this satellite class as well as C-band. As a result, the single-distance likelihood functions show greater absolute values of log likelihood for a given distance. In this particular case, several statistical groupings are used, employing the early combination method of looking up each distance in an appropriate likelihood function, summing the total likelihoods, and then correcting with the analytic model.

Proceeding in the usual way, Figure 14 shows the estimated values of Λ_n for this C-band case and may be compared with Figure 4. There are some similarities and some differences. The summed likelihood abscissa includes likelihoods of over twice the previous case, which is directly attributable to the greater likelihoods found in the single-distance density (not shown).

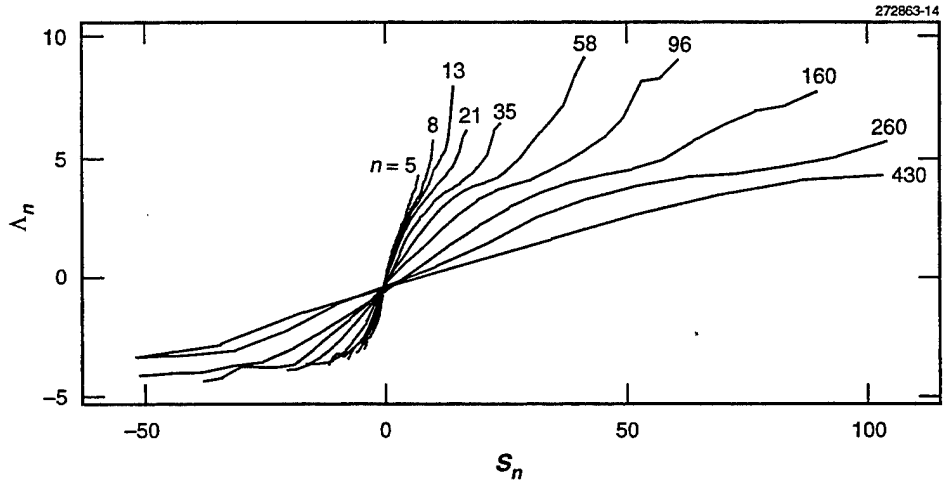


Figure 14. C-band estimated S_n likelihood ratios.

The linear region extends into the small negative likelihood region more so than in the previous case, lending greater support to the idea of using a single correction factor for all likelihood sums. There are obvious correlations in the curve fluctuations, especially for values of 35, 58, and 96, which are caused by the correlations discussed previously.

The big question, however, is how well the analytic model will fit this different data set. Figure 15 shows this analysis and can be directly compared with Figure 5. The overall trends are again similar. While the center plot of Figure 5 is almost linear, suggesting that a constant slope might be a useful model, Figure 15 shows more variation. The right plot of Figure 15 shows a reasonable fit to a straight line, and indeed a system constructed using these statistics shows good performance.

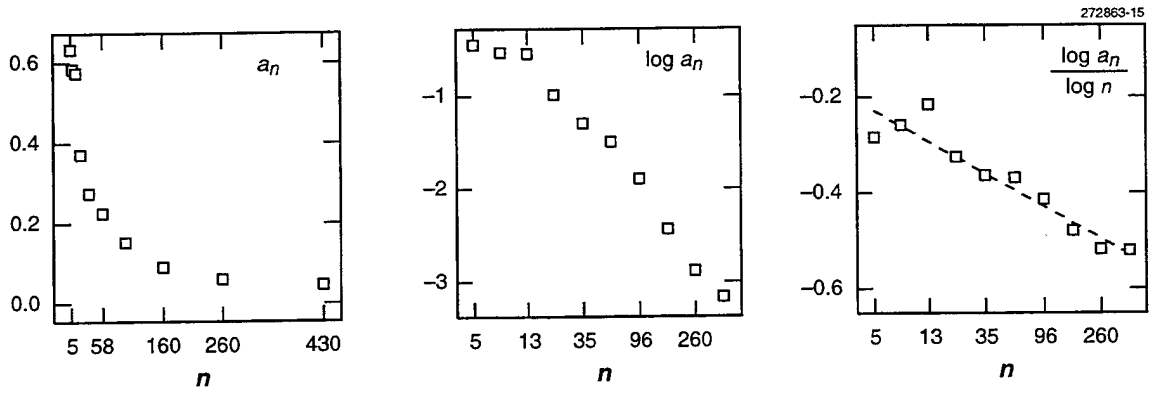


Figure 15. C-band a_n for various n.

Figure 16 shows the performance of this system after parameter optimization in the same format as Figure 10. Note that the system error rate is low, and a high percentage of good signatures generate a high confidence. The system merit is more than twice that of the example of Figure 10. The K_n curve shows that generally smaller corrections are being applied to S_n than previously.

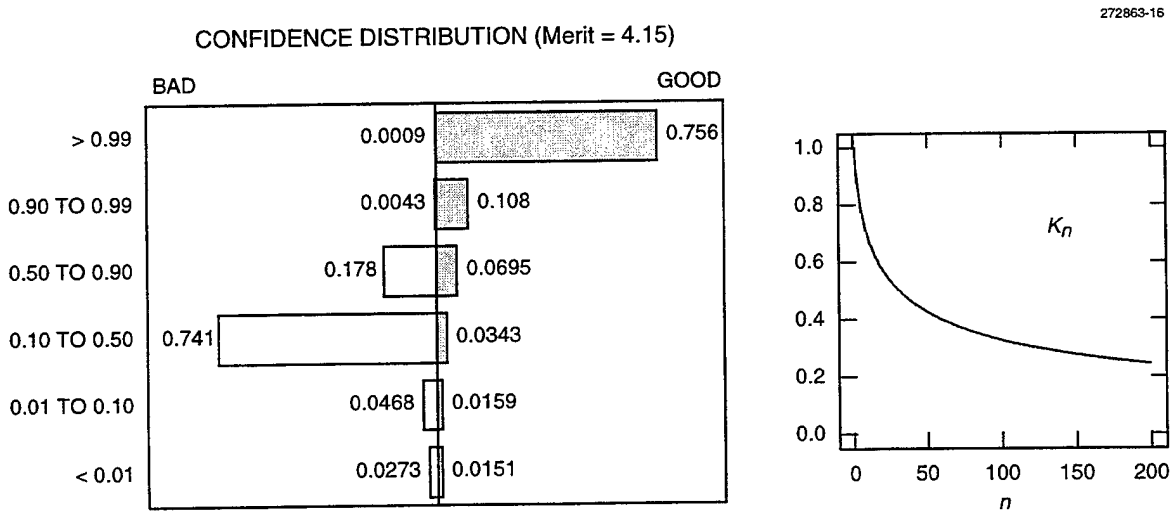


Figure 16. C-band system performance after optimization.

6.2 WIDEBAND RADAR EXAMPLE

The final example is significantly different, involving a wideband radar that can measure high-resolution range profiles (RPs). Techniques for aircraft classification using radar RPs have been described [21,14] recently, containing numerous RP plots. RPs are one-dimensional functions that sometimes look vaguely like signatures, but represent instantaneous RCS as a function of range rather than total cross section as a function of time. The range is swept out by each radar pulse in nanoseconds, while the signatures represent several minutes of data as an LEO satellite passes the sensor.

In addition, the targets are again changing. Rather than LEO satellites, geostationary-earth-orbit (GEO) satellites (in orbital planes nominally passing through the equator) are used. Since these do not move with respect to the radar, range profiles can be collected to help identify and monitor them.

While LEO satellites may have complicated motions of instruments and solar panels, not to mention the motion of the entire satellite relative to the radar, GEO satellites usually have minimal motion of instruments with essentially no motion of the satellite relative to the radar.

Most GEO satellites can be classified into two groups. One group has one or two solar panels pointing north/south and nominally rotating once per day, to follow the sun. There will normally be some daily variation in the RPs due to solar panel motion. Figure 17 shows an example RP from a typical satellite of this class. The initial peaks usually represent returns from earth-pointed instruments such as antennas. The large peak usually represents a return from the satellite main body. The solar panels usually affect the RP just beyond the large peak. In fact, if they face the radar, they can sometimes produce a very large specular peak, larger even than that of the main body.

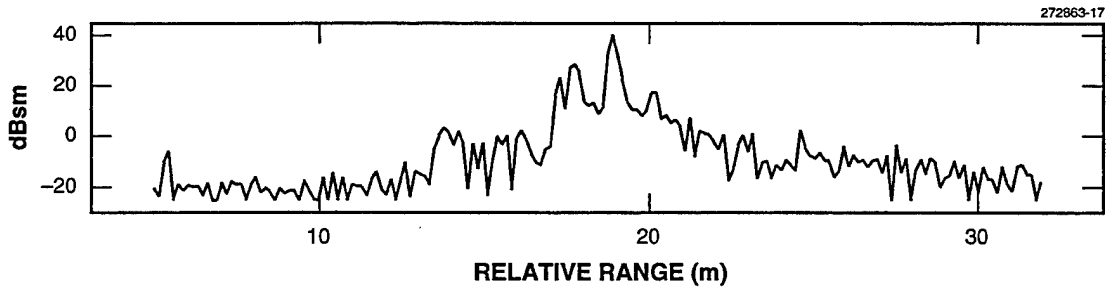


Figure 17. Orion 1 range profile (3-axis stable).

The other group of geosynchronous satellites spin, collecting solar energy with cells arranged on the outside of the drum. Figure 18 is an example of this class. In general, there should be much less variation from RP to RP in this class; any variation may be attributed to station-keeping differences and instrument pointing changes.

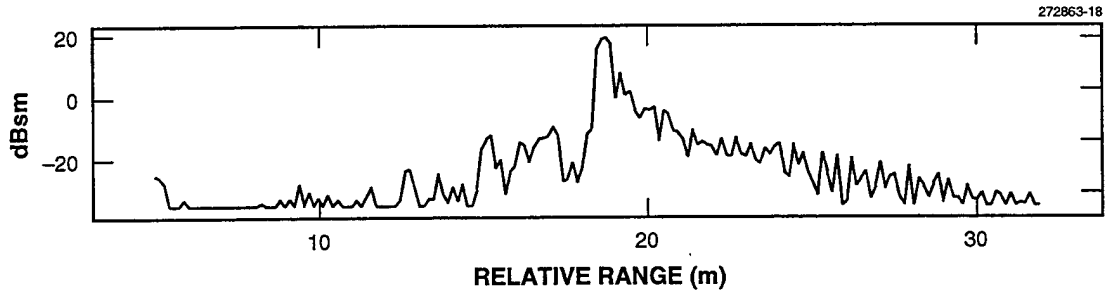


Figure 18. Galaxy 6 range profile (spinner).

The algorithms used to compare two RPs are not discussed here, but the comparison is much simpler than that of the narrowband radar case and is based on cross-correlation of entire RPs after preprocessing. The result is a single distance, describing the similarity between two RPs. Once distances are computed, the statistical algorithms described previously can be applied to assess the data. This example will be used to demonstrate the generality of the method.

In the previous narrowband radar examples, about 2500 signatures from the database were used. In this wideband case, only 338 total RPs are available to work with, which gives a total of 56953 distances of all kinds. Again, the interest is in monitoring the stability of a satellite and identifying it if necessary. If an RP can be formed, the satellite is relatively stable, but there is uncertainty whether it might be slowly tumbling, or if it is the wrong object. The following approach is adopted:

1. For H_0, use all distances between satellites of different classes. This simulates the case where RPs do not match due to bad identification (ID). Based on previous experiments, this level of mismatch may also be useful in case the correct satellite is in an unexpected orientation.

2. For H_1, use all distances between similar satellites with very similar configurations, i.e., spinners to spinners or 3-axis stable to 3-axis stable in which the Greenwich Mean Time (GMT) is within 1 hour, and hence the solar panel angle is expected to be within roughly 15°.

3. For H_2, use the remaining distances (3-axis stable to 3-axis stable) but with GMT difference greater than 1 hour. These should show similarity, but not as great as those in H_1.

The top graph of Figure 19 shows the PDF estimates obtained by dividing the distances into these three groups. The estimates indeed show that H_1 has the smallest associated distances, followed by H_2, and then H_0. The differences of these log curves give the log-likelihood curves shown in the bottom graph. To interpret a single distance in this system, simply determine whether the two RPs involved are comparable spinners, or 3-axis stable satellites with GMT within 1 hour (and comparable position in the

geosynchronous belt). If so, use the solid curve, allowing the possibility of obtaining fairly large confidences from a single comparison. If not, the dotted curve must be used, and less information is obtained. The reason to bother with the dotted curve is that many more distances become available, and their combined likelihood may give a good assessment of the RP.

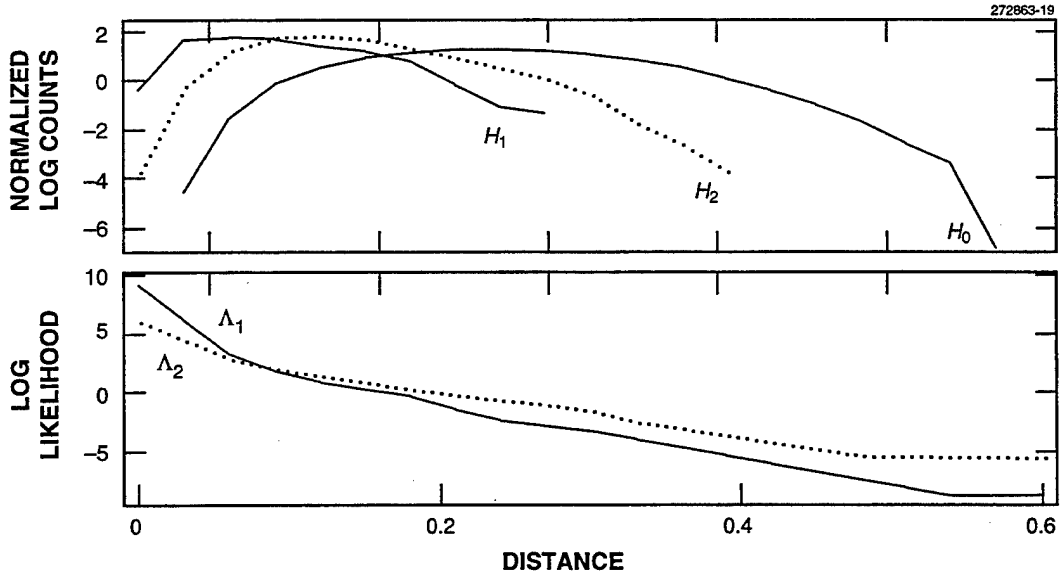


Figure 19. Single-distance statistics.

Figure 19 may be compared directly with Figure 3. Note that the forms of the PDFs are completely different, as are the likelihood functions. In particular, much larger negative likelihoods are obtainable in this example than in the narrowband radar examples. The question here is whether this method of combining likelihoods and correcting for correlations will be applicable.

There is not enough data supporting the analysis H_1 to do late combination in which the corrected, summed likelihoods are combined to form the overall likelihood. Instead, the early method is used in which the individual likelihoods are obtained from the appropriate curve, and then the heterogeneous likelihood sum is corrected using a single correction factor. Computing the experimental sums S_n as before gives the curves of Figure 20. These look qualitatively similar to the previous example shown in Figure 4. They are more symmetrical in the negative likelihood region, which would be expected since the single-distance PDFs are also more symmetrical. The values of n are less, reflecting the much smaller database.

Next these slopes were analyzed, producing Figure 21, in the same format as Figure 5. This figure certainly looks familiar; although the parameters are somewhat different, it shows similar behavior of the slopes of the S_n curves. Hence, even though the physics, targets, and single-distance statistics are completely different, the statistical model appears to detect the underlying dimensionality of the data and provides a useful fusion algorithm.

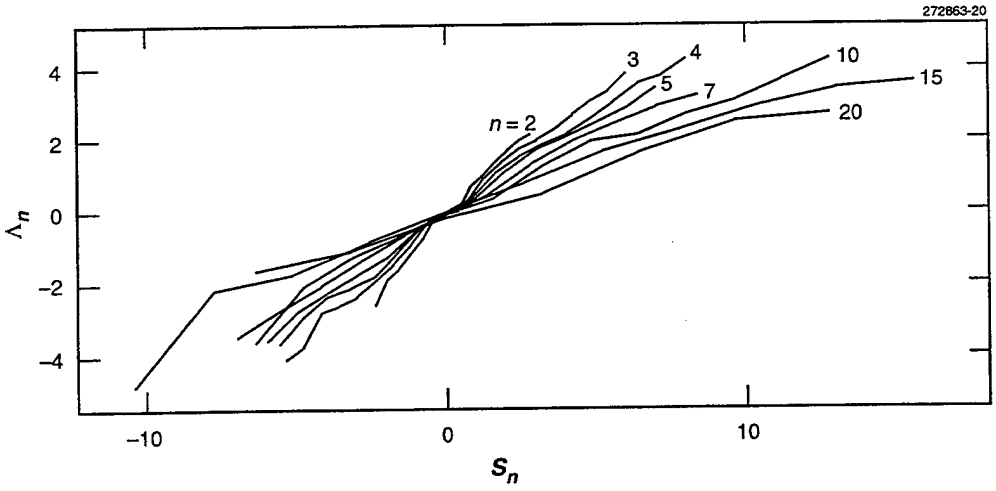


Figure 20. Estimated RP S_n likelihood ratios.

The final test of this analysis is, of course, to perform an experiment and assess the information obtained. For this case, an ID problem was again chosen for the performance characterization. Examining the database of 338 RPs, 136 different objects are contained in it. However, four satellite classes have a total of 144 RPs. For the "good" assessments, the likelihood that each of these 144 RPs is a stable satellite of the correct type was assessed. For the "bad" assessments, the likelihood that each of these 144 is a stable satellite of one of the other three types was assessed. This experiment gives 144 good assessments and 432 bad assessments.

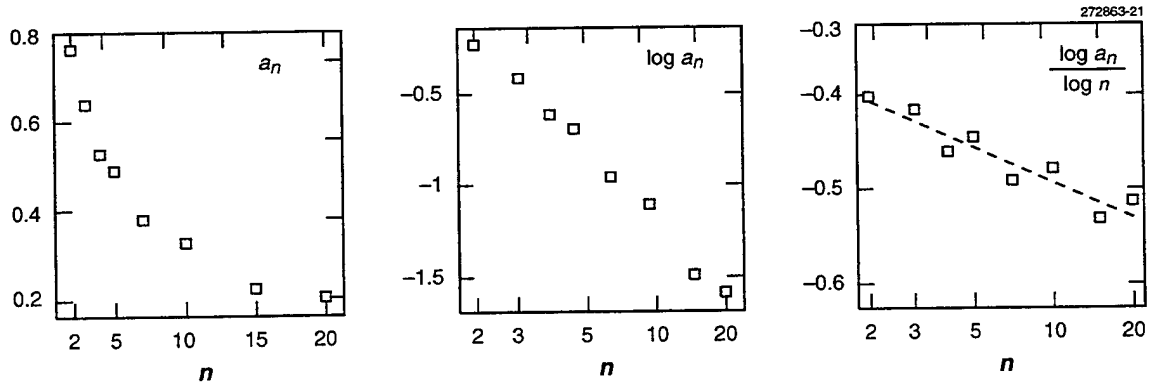


Figure 21. Analysis of RP a_n for various n.

Figure 22 shows the performance of the system on this experiment, using the model represented by the dashed line at right in Figure 21. The merit is 3.63, which compares favorably with the narrowband radar examples. False alarms are very low, and the confidences tend to be bunched towards the middle. Recall that the two-parameter (m,b) optimization process merely scales the log likelihoods, so the fraction of the confidences above or below 50% does not change. Since less than 3% of the bad cases are above 50%, and less than 12% of the good cases are below 12%, there is a lot of room for optimization by shifting the K_n curve up. The results of that optimization are not shown here. It is already clear that these methods appear applicable to this class of data and that the overall system can produce useful information.

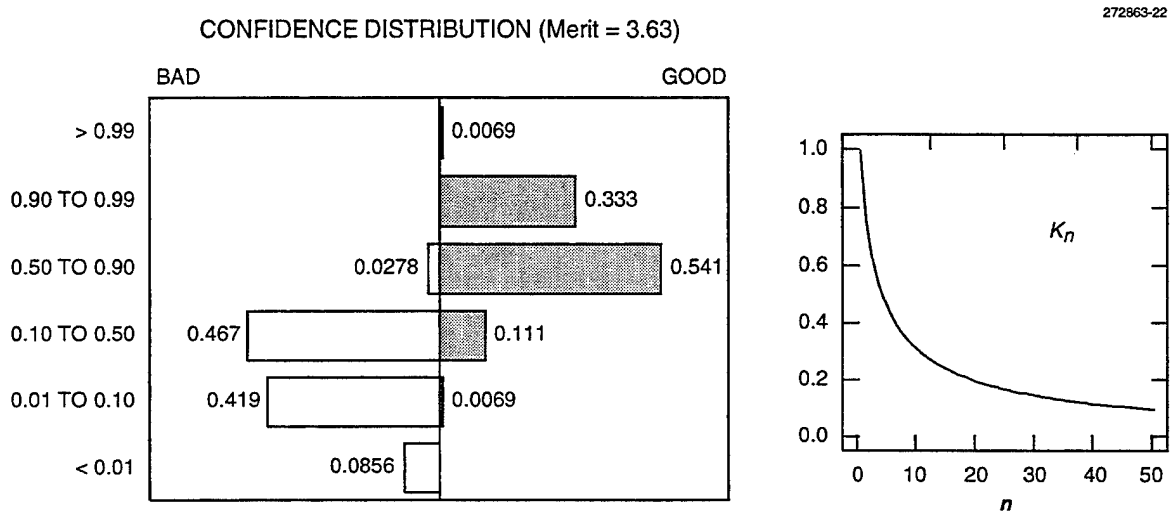


Figure 22. RP system performance before optimization.

7. SUMMARY AND FUTURE WORK

A new technique has been described for fusing correlated data. This technique has several advantages over competing methods such as neural networks or *k*-nearest-neighbor classifiers, including using data of varying dimension and databases of varying size without discarding any relevant information.

Once a data-specific method of vector comparison has been produced, this machinery can be applied in largely automatic fashion, producing an algorithm for assessing the confidence that a vector under test is "normal" as defined by the database. If the database can be divided into subsets representing multiple conditions, multiple hypothesis tests enable testing for the presence or absence of those conditions. System optimization is easily performed to adjust false alarm rates.

Experiments on several data types have demonstrated the wide applicability and robustness of the method. The investigation continues regarding the application of the method to new scenarios and data types. To date, this research has been data- and experiment-driven. It may be useful to invest more effort into analytic analysis of this model to determine the classes of input data for which it is best suited and to refine the algorithms.

It may become necessary to pursue data fusion beyond the vector comparisons described in this report. For example, in a satellite monitoring application, a priori knowledge or additional sources of information about satellite status might be available. Other pieces of information, such as specular locations extracted from the signatures themselves, might also represent useful evidence.

There is no immediate requirement for such a system, but the Bayesian belief network [22] formalism appears quite attractive. It seems that the "confidences" might fairly easily be interpreted as probabilities, providing good compatibility with the Bayesian formalism. Preliminary design sketches of such systems have been made, but no experiments have been done to date.

REFERENCES

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

2. R.P. Lippmann, "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine* **4**, 4–22 (April 1987).

3. T.M. Cover and P.E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Inf. Theory* ***13****, 21–27 (January 1967).*

4. K. Fukunaga and L.D. Hostetler, "Optimization of *k*-Nearest-Neighbor Density Estimates," *IEEE Trans. Inf. Theory* **19**, 320–326 (May 1973).

5. S.A. Dudani, "The Distance-Weighted *k*-Nearest-Neighbor Rule, *IEEE Trans. Sys. Man Cybern.* **6**, 325–327 (April 1976).

6. J.M. Keller, M.R. Gray, and J.A. Givens, "A Fuzzy *k*-Nearest Neighbor Algorithm," *IEEE Trans. Sys. Man Cybern.* **15**, 580–585 (July/August 1985).

7. T. Baily and A.K. Jain, "A Note on Distance-Weighted *k*-Nearest-Neighbor Rules," *IEEE Trans. Sys. Man Cybern.* **8**, 311–313 (April 1978).

8. W.F. Schmidt, D.F. Levelt, and R.P.W. Duin, "An Experimental Comparison of Neural Classifiers with 'Traditional' Classifiers," In *Pattern Recognition in Practice IV: Multiple Paradigms, Comparative Studies and Hybrid Systems*, L.N. Kanal and E.S. Gelsema, eds., Amsterdam: Elsevier Science (1994).

9. D.W. Ruck, S.K. Rogers, M. Kabrisky, M.E. Oxley, and B.W. Suter, "The Multilayer Perceptron as an Approximation to a Bayes Optimal Discriminant Function," *IEEE Trans. Neural Networks* **1**, 296–298 (December 1990).

10. F. Kanaya and S. Miiyake, "Bayes Statistical Behavior and Valid Generalization of Pattern Classifying Neural Networks," *IEEE Trans. Neural Networks* **2**, 471–475 (July 1991).

11. E. Barnard, "Comments on 'Bayes Statistical Behavior and Valid Generalization of Pattern Classifying Neural Networks'," *IEEE Trans. Neural Networks* **3**, 471–475 (September 1992).

12. S. Geva and J. Sitte, "Adaptive Nearest Neighbor Pattern Classification," *IEEE Trans. Neural Networks***23**, 318–322 (March 1991).

13. W.E. Weideman, M.T. Manry, H-C. Yau, and W. Gong, "Comparisons of a Neural Network and a Nearest-Neighbor Classifier via the Numeric Handprint Recognition Problem," *IEEE Trans. Neural Networks* **6**, 1524–1530 (November 1995).

14. M.A. Rubin, "Application of Fuzzy ARTMAP and ART-EMAP to Automatic Target Recognition Using Radar Range Profiles," *Neural Networks* **8**, 1109–1116 (1995).

15. B. Dubuisson and M. Masson, "A Statistical Decision Rule with Incomplete Knowledge About Classes," *Pattern Recognit.* **26**, 155–165 (1993).

16. T. Denoeux, "A *k*-Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory," *IEEE Trans. Sys. Man Cybern.* **25**, 804–813 (May 1995).

17. T.P. Wallace (Private Communication, 1996).

18. H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, New York: Wiley (1968).

19. M. Schwartz and L. Shaw, *Signal Processing: Discrete Spectral Analysis, Detection, and Estimation*, New York: McGraw-Hill (1975).

20. S.L. Marple, *Digital Spectral Analysis with Applications*, Englewood Cliffs, N.J.: Prentice-Hall (1975).

21. S. Hudson and D. Psaltis, "Correlation Filters for Aircraft Identification from Radar Range Profiles," *IEEE Trans. Aerosp. Elect. Sys.* **29**, 741–748 (July 1993).

22. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, Calif.: Morgan Kaufmann Publishers (1988).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	5 November 1996	Technical Report

4. TITLE AND SUBTITLE

An All-Neighbor Classification Rule Based on Correlated Distance Combination

5. FUNDING NUMBERS

C — F19628-95-C-0002
PR — 80
PE — 12424F,
31310F,
63223C

6. AUTHOR(S)

Timothy P. Wallace

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Lincoln Laboratory, MIT
244 Wood Street
Lexington, MA 02173-9108

8. PERFORMING ORGANIZATION REPORT NUMBER

TR-1030

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Air Force Space Command
Peterson AFB, CO 80914-1608

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ESC-TR-95-097

11. SUPPLEMENTARY NOTES

None

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

This report describes a new method of classifying data vectors by invoking a two-step process. First, a data-specific step produces a "distance" qualitatively describing the similarity of the vector under analysis to each vector in a database representing a particular class. Second, the evidence represented by the vector of statistically correlated "distance" is combined into an overal numerical confidence that the vector under test belongs to the same class as the database vectors. In addition, the supporting evidence is available in the form of the individual distances.

This "all-neighbor" method has several advantages over competing formalisms such as neural networks or the k-nearest-neighbor classification method. It can deal with data vectors of varying dimensions, as long as the distance measure is capable of comparing them in some fashion. Even more importantly, it can deal with distance vectors of varying dimension, such as are common when dealing with a heterogeneous reference database. It produces a numeric confidence rather than just a simple classification. Further, it uses all the information contained in the distance vector, and it facilitates adjustment of false alarm rates. The method is applied to several different types to demonstrate its generality.

14. SUBJECT TERMS

pattern recognition
nearest neighbor
space surveillance
range profile
classification
neural network
radar signature

15. NUMBER OF PAGES
50

16. PRICE CODE

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