Information Theoretic Bounds of ATR Algorithm Performance for Sidescan Sonar Target Classification

Vincent L. Myers, Marc A. Pinto
NATO Undersea Research Centre
Viale San Bartolomeo, 400
La Spezia, Italy, 19038

myers@saclantc.nato.int / pinto@saclantc.nato.int

SUMMARY

Autonomous underwater vehicles (AUVs) are receiving considerable attention as platforms to carry real or synthetic aperture sidescan sonars capable of classifying and identifying mine-like targets on the seafloor. As research begins to focus on the independent decision-making capabilities and behaviours for these vehicles, some effort is being spent on developing automatic target recognition (ATR) algorithms that are able to operate with high reliability under a wide range of scenarios, particularly in areas of high clutter density, without human intervention. Because of the great diversity of pattern recognition methods and continuously improving sensor technology, there is an acute requirement for objective performance measures that are independent of any particular sensor, algorithm or target definitions.

This paper approaches the ATR problem from the point of view of information theory in an attempt to place bounds on the performance of target classification algorithms that are based on the acoustic shadow of proud targets. The information that can be used for classification found in sidescan sonar imagery is examined and common information theory relationships are used to derive some properties of the ATR problem.

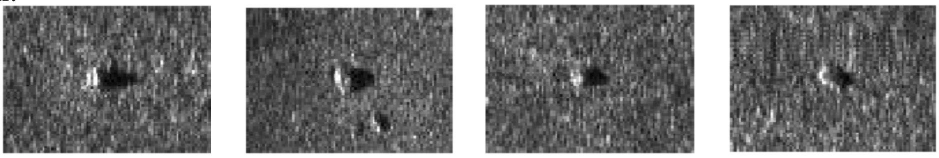
1 Introduction

Autonomous underwater vehicles (AUVs) equipped with high-resolution synthetic aperture sonars possess the potential to change the nature of mine countermeasures (MCM) operations. In order to realize this potential, highly reliable and accurate automatic target recognition (ATR) algorithms are needed. Mine classification has traditionally been an “operator in the loop” process, whereas in a concept of operations where one or many AUVs execute covert or semi-covert missions, some form of target discrimination is required in order to, as a minimum, reduce the amount of data to be stored, analyzed and possibly transmitted to the next stage of the mine hunting process resulting in reduced time and effort spent on prosecuting false contacts. Such algorithms are also necessary if we are to motivate the exploration of cooperative and dynamic AUV behaviours, as well as to augment their independent and covert capabilities.

Efforts have been made towards the development of satisfactory ATR systems for discriminating between mine-like objects and clutter. At times non-parametric machine learning algorithms are used, such as neural networks [1], support vector machines [2], decision trees [3] or some fusion thereof. Other times, model-based methods [4] or unsupervized algorithms are employed. In all cases, the objective is to separate objects

Paper presented at the RTO SET Symposium on “Capabilities of Acoustics in Air-Ground and Maritime Reconnaissance, Target Classification and Identification”, held in Lerici, Italy, 26-28 April 2004, and published in RTO-MP-SET-079.

Figure 1: Rocks and mines often have similar size and signatures, creating a difficult pattern recognition problem: Thumbnails (a) and (c) are acoustic images of rocks and (b) and (d) of mines using a commercial sidescan sonar.



(a) (b) (c) (d)

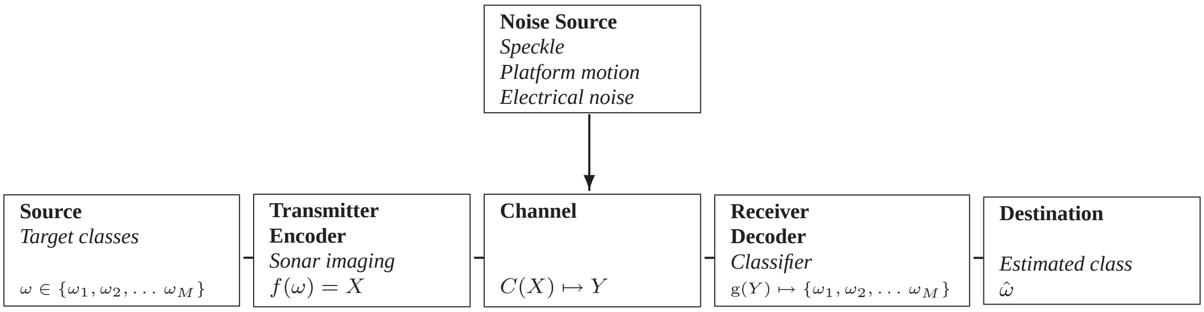
from background reverberation and to discriminate between targets (mines) and clutter of similar size and comparable acoustic signatures (Figure 1).

The capability to predict and bound sidescan sonar ATR algorithm performance is vital in providing objective benchmarks by which to measure the quality of the sensor, the sonar imagery, the image formation process as well as providing goals and evaluation criteria for pattern recognition algorithms. Since the detection and classification of mines is the ultimate goal of a minehunting system, bounds on performance could be used to assess design tradeoffs and carry out cost-benefit analyses during the sensor development phase rather than afterwards, when changes are costly and difficult. Finally, an upper bound on ATR algorithm performance for a given minehunting system could substantiate the need for increased research into improved pattern recognition techniques for sonar imagery, if the algorithms are performing significantly below the predicted upper bound, or conversely if effort should rather be spent on new and improved sensor development.

The need for performance prediction and bounding has been established for ATR in the field of real and synthetic aperture radar (for instance, see Dudgeon [5]). Horrowitz & Brendal [6] computed the probability of classification error using a Monte Carlo simulation to evaluate the Bayes error integral. Horne [7] used mutual information to estimate a number of possible target “states” that can be represented by a set of image pixels, combined with a number of “distinguished” states to ultimately compute an overall probability of error. Briles [8] relates mutual information to the Bayesian probability of false classification of a target identification system. Some work has also been done on bounding the performance of ATR for sonar. Pinto [9] defined a performance index based on a model for the standard deviation of the shadow boundary. Florin et al. [10] used a similar method to define a probability of false classification and Kessel [11] suggested that the probability of classification is connected with the probability of finding the correct boundary between shadow and background.

We develop a method for bounding ATR algorithm performance based partly upon the information theory of Briles and the circle-square discrimination problem of Florin et al. To this end, only the bounds of algorithms that employ shadow-based recognition methods using size, shape and texture features, are obtained. We first present ATR as a problem of communicating a message over a noisy channel. Sources of message corruption, such as noise distributions, the shadow to reverberation contrast and the effect of resolution examined. Using rate-distortion theory and the Chernoff information, bounds to the performance of any classifier for the defined classification task are found. The implications of these limits are discussed, with a particular emphasis placed on design considerations such as contrast and resolution.

Figure 2: Shannon's communication problem, revised for the ATR problem.



2 Pattern recognition as a communication problem

Figure 2 shows the five stage communication process originally described by Shannon in his seminal work on information theory [12]. A message ω randomly selected from a set of possible messages $\{\omega_1, \omega_2, \dots \omega_M\}$ is processed by an *encoder* function f, resulting in the signal X. This signal is transmitted to the receiver via the communication channel C resulting in the message Y based on the class conditional probability $p(y|x)$, that is, the probability that $X = x$ given that $Y = y$. On the receiver's side, a *decoder* function g is applied to Y, resulting in an estimate of the transmitted message $\hat{\omega}$ [†].

Posed in such a way, the communication process of Figure 2 can be interpreted in a manner that characterizes the sidescan sonar ATR problem. The source message is the true, unknown target class (e.g. mine and non-mine), each having an *a priori*, or prior probability of occurring, denoted as the probability mass function $P(\omega_i) = \Pr\{\omega = \omega_i\}$. The encoding function f is the process of sonar ensonification. At this stage in the process, the function $f(\omega) = X$ is deterministic on ω, and so we have the *a priori* probability $P(x_i) = P(\omega_i), \forall i$ and we will generally refer to the class or message with the random variable X. The channel is essentially an abstract entity that "transmits" the message from the transmitter to the receiver. Throughout this transmission the signal X is corrupted by various sources of noise, resulting in the received signal $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_N\}$, an N-dimensional random variable over $\mathcal{R}^N$ that is a general vectorized form of an $N = N_1 \times N_2$ sonar image. The virtual channel C follows the conditional distribution $p(\mathbf{y}|x) = \Pr\{\mathbf{Y} = \mathbf{y}|X = x\}$, the probability of encountering an observed sonar image $\mathbf{Y}$ given that the encoded message X was sent.

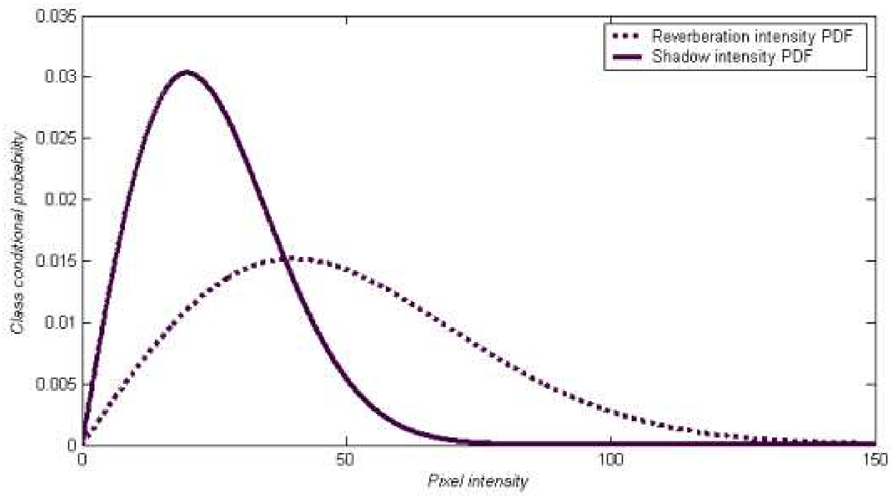
The received message, namely the sonar imagery, is interpreted by the decoding function g that makes a "guess" as to the unknown nature of the original transmitted message. The function g is the ATR process; we wish to bound the performance of g in terms of the probability of error. Finally, the estimated classification of the target $\hat{\omega}$ is passed onto the next stage of the process, for instance an AUV control system capable of adjusting its behaviour based on the nature of the contact.

2.1 Sources of noise

We now examine some of the sources of noise that corrupt the transmitted signal and cause classification errors to occur.

[†] Probability mass functions are denote using an uppercase P while probability density functions are denoted in lowercase p.

Figure 3: Example of probability distributions for shadow and reverberation pixels. In this simplified case, there is a great deal of overlap between the two classes, and pixels can take on intensities of either class. This is also the case in practice, however it is unlikely that the distributions would be so clearly defined and would rarely remain constant throughout an entire survey, or even within a single ping



2.1.1 Contrast

Upon visual inspection of sidescan sonar imagery, one at once notices type of pixel to pixel "twinkling" called *speckle*. Speckle noise occurs when a coherent signal such as sonar, radar or laser, is used to illuminate a target. The in-phase signal, reflected by a number of sub-resolution sized scatterers, returns to the source having been altered by destructive and constructive interferences of phase. The result is an image where the background / seafloor region is without a constant signal. Speckle noise follows the Rayleigh distribution, defined as:

$$p(x) = \frac{x}{\beta^2} e^{-\frac{x^2}{2\beta^2}} \text{ for } x, \beta > 0. \tag{2.1}$$

The Rayleigh distribution has a single parameter β that is equal to its mode.

Speckle noise is not intrinsically responsible for loss of performance. Ambiguities arise when Rayleigh distributions for shadow and reverberation overlap, resulting in a situation where different regions can give rise to the same pixel intensities (Figure 3). What is significant is the separation of these two distributions, referred to as contrast. It is convenient to represent contrast as a single metric, the signal to noise ratio defined as $20\log_{10}(\frac{\beta_{\mathrm{S}}}{\beta_{\mathrm{R}}})$, where β_{S} is the mode of the distribution for the shadow areas and β_{R} is the mode of the background reverberation. Image contrast is a key factor for visual classification by human operators, and the contrast sensitivity of the human visual system is well-studied.

The principal source of poor contrast is the lack of resolution which is discussed below. However, even with a high resolution sonar, contrast can be affected by environmental factors such as multipath. Multipath occurs when second and higher-order returns contribute to the overall energy received in the shadow region, resulting in a general decrease in the distance of the shadow and reverberation signals. This effect is much more prevalent in shallow and very shallow water. Other physical processes such as diffraction also contribute to an imperfect (non-zero) shadow signal and system noise, although the latter is less prevalent in current sonar systems.

2.1.2 Resolution

The resolution of an image element is its physical dimensions. It is evident that if resolution is coarser than the size of the cues needed to perform recognition, then any satisfactory degree of recognition will not be achievable.

There is also an effect on recognition capacity attributed to the quantization that occurs when an image is partitioned into resolution cells that contributes to the contrast of the image. Pixels located on the boundary of the shadow and background regions lose contrast due to a resolution-sized low-pass filter over the shadow and background distributions. Pixels on the edge have a distribution that lies somewhere between β_{R} and β_{S} resulting in the loss of edge definition or crispness that follows the loss in contrast.

Resolution can be improved using various signal processing methods, such as synthetic aperture processing (SAS), or at the sensor or transducer level by increasing bandwidth. Resolution can also be separated into two components: Along-track resolution, the size of the pixel in the direction of movement of the sonar, and across-track resolution, the size perpendicular to the sonar track. These need not be the same, and in practice are not equal. For the sake of simplicity, however, pixels are assumed to be “squared” meaning that they are sampled equally in the along and across tracks.

2.2 The Bayes classifier

The ATR process is represented by the decision function g that predicts the class of a given detection based on the observed sonar imagery. Formally, a function $\mathrm{g}(\mathbf{Y}) : \mathbf{Y} \rightarrow \{x_1, x_2, \dots x_M\}$ that represents the predicted value of X given the series of measurements $\mathbf{Y}$. A misclassification error occurs when $\mathrm{g}(\mathbf{Y}) \neq X$, and the probability of error for a given classifier g is $L(\mathrm{g}) = \Pr\{\mathrm{g}(\mathbf{Y}) \neq X\}$. The best possible classifier is the one which minimizes the probability of error, defined as

$$\mathrm{g}^* = \min_{\mathrm{g}:\mathbf{Y}\rightarrow\{x_1,x_2,\dots x_M\}} \Pr\{\mathrm{g}(\mathbf{Y}) \neq X\}, \tag{2.2}$$

The function g^* is called the *Bayes classifier* and the minimum probability of error, $L^* = L(\mathrm{g}^*)$ is called the *Bayes error*.

If the class-conditional distribution $p(\mathbf{y}|x)$ is known, then g^* is also known; it chooses the class with the maximum *a posteriori*, or posterior probability:

$$\mathrm{g}^* = \mathrm{I} \underset{i=\{1,2,\dots M\}}{\arg\max}\ P(x_i)p(\mathbf{y}|x_i) \tag{2.3}$$

where I_A denotes the indicator of the set A. The Bayes error rate is therefore:

$$L^* = 1 - \int \cdots \int_{\mathcal{R}^N} \max_{i=\{1,2,\dots M\}} P(x_i)p(\mathbf{y}|x_i)dy. \tag{2.4}$$

The Bayes error is the best achievable error rate for any classification method, and although in some cases calculating L^* can be done analytically (Duda & Stork [13]),in general the typically large number of dimensions and the discontinuities inherent in the decision regions are enough to render the evaluation of the integral difficult or impossible.

3 An information theory for ATR

Posing target recognition as a communication problem allows us to use information theory to analyse the system[14]. The *entropy* $H(X)$ of a discrete random variable is defined as

$$H(X) = -\sum_{i=1}^{M} P(x_i) \log P(x_i). \tag{3.1}$$

If the log is base e entropy is measured in *nats*, and this is the convention followed in this paper. Entropy is a measure of uncertainty concerning a random variable. In the communication system defined above, entropy is used to measure the initial uncertainty that is held about the target classes, so that $P(x_i)$ is the prior probability of class i occurring. The entropy function $H(X)$ is maximum when all the $P(x_i)$'s are equal.

In the ATR problem there is a discrete number of classes but the pixels follow continuous distributions. The joint entropy $H(X, \mathbf{Y})$ for this case is

$$H(X, \mathbf{Y}) = -\sum_{i=1}^{M} \int \underset{\mathcal{R}^N}{\cdots} \int p(x_i, \mathbf{y}) \log p(x_i, \mathbf{y}), \tag{3.2}$$

and the conditional entropy $H(X|\mathbf{Y})$ is

$$H(X|\mathbf{Y}) = -\int \underset{\mathcal{R}^N}{\cdots} \int p(\mathbf{y}) H(X|\mathbf{Y} = \mathbf{y}), \tag{3.3}$$

where N is the number of pixels.

The conditional entropy is the remaining uncertainty about the nature of the variable X after having observed $\mathbf{y}$. This reduction in uncertainty is called *mutual information*. The mutual information $I(X; Y)$ of two random variables X and $\mathbf{Y}$ is defined as the distance between the joint probability $p(x, \mathbf{y})$ and the product of the marginal distributions $P(x)$ and $p(\mathbf{y})$:

$$I(X; Y) = \sum_{i=1}^{M} \int \underset{\mathcal{R}^N}{\cdots} \int p(x_i, \mathbf{y}) \log \frac{p(x_i, \mathbf{y})}{P(x_i) p(\mathbf{y})}. \tag{3.4}$$

or more simply:

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{3.5}$$

Mutual information is therefore a measure of the information carried by the channel C. An ATR system must make decisions based on the observation of Y. What is the best possible decision?

3.1 Rate distortion theory

Rate distortion theory [15], [8] is commonly used in data compression schemes in order to study the trade offs between the required nat Assume for the case of minehunting that the number of possible decisions is equal to the number of possible classes (i.e. M). The average distortion for a channel C with class conditional probabilities $p(\mathbf{y}|x)$ is

$$d_C = \sum_{i=1}^{M}\sum_{j=1}^{M} p(y_j|x_i)\rho(i,j)P(x_i). \tag{3.6}$$

For the Hamming distortion function, the average distortion is the probability of error. A channel is D-admissible if $d_C \leq D$ and the set of all D-admissible channels is

$$M_D = \{C : d_C \leq D\}\,. \tag{3.7}$$

The problem in rate distortion theory is to minimize the amount of mutual information required to reproduce the signal X with a distortion less than or equal some bound D^*. The minimization is done over all channels, and we are faced with the optimization problem:

$$R(D) = \min_C I(X;Y) \text{ given } d_C \leq D^*, \tag{3.8}$$

the result $R(D)$ being called the rate distortion bound. To put the target recognition problem within the context of rate distortion theory, we must define a slightly different distortion function

$$\rho_{\mathrm{g}}(X, \mathrm{g}(\mathbf{Y})) = \begin{cases} 1 & \text{if } \mathrm{g}(\mathbf{Y}) \neq X \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

This gives the probability of error of the classifier g, and we now have

$$R(L) = \min_C I(X;Y) \text{ given } d_C \leq L^*, \tag{3.10}$$

where L is the error rate and $R(L)$ is called the Bayes rate distortion bound. While the traditional rate distortion bound uses a simple average, the Bayes rate distortion bound requires the evaluation of a set of conditional probability distributions.

We state the following theorem without proof (see Kanaya & Nakagawa [16]):

Theorem 3.1 DECISIONMAKING THEOREM *Let $R(D)$ be the traditional rate distortion bound and $R(L)$ the Bayes rate distortion bound. The following equality holds:*

$$R(D) = R(L), \forall L > 0. \tag{3.11}$$

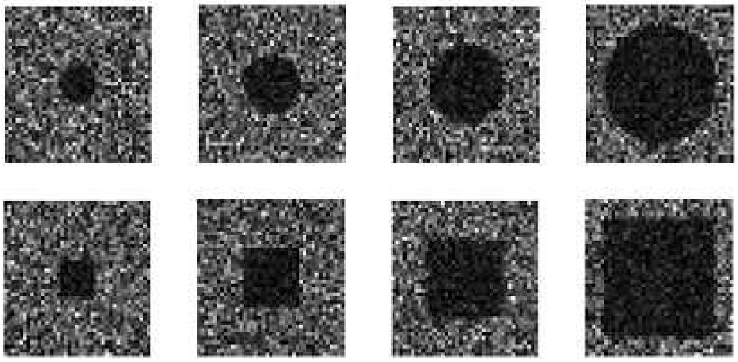
We are therefore able to circumvent calculating the more complex Bayes rate distortion bound, and use the much simpler and well understood conventional rate distortion bound. Berger [15] established that the rate distortion bound for equal class probabilities and the Hamming distortion function defined above is reduced to

$$R(L) = H(X) - H(L) - L\log(M-1), \tag{3.12}$$

where $H(X)$ is the entropy of the class prior probabilities and $H(L) = -L\log(L) - (1-L)\log(1-L)$ is the entropy of the error rate.

We are therefore in a position whereby establishing the mutual information of the ATR communication problem we can compute the Bayes error rate.

Figure 4: The circle-square test for a number of resolutions and fixed contrast



4 Principal Results

In the general target recognition problem we have a situation in which there are many non-target detections with signatures that are similar to target signatures. Although we are able to define the statistics of the noise present in sonar imagery, there is nevertheless the problem of properly defining what is meant by "target" and "non-target". In reality, defining these classes properly is not only virtually impossible but is likely to be in fact impossible because, while accurately modeling shadow shape characteristics for known mine targets is relatively easy, precisely knowing the distribution of non-target contacts that can possibly be found on the seabed requires an enormous amount of data and can vary as the geological features (for rocks) or environmental history (for man-made objects) of an area changes, sometimes during the course of a single mission. Data sets gathered to this end are often collected stratified with equal samples of each class, whereas the actual class prior probabilities of occurring are likely skewed.

4.1 The circle-square test

Florin et al [10] suggested that an object recognition task could be reduced to differentiating between a circle and a square (Figure 4). In practice, target and non-target signatures are not so easily represented, nor are they as dissimilar and often shadow signatures for some rocks are significantly mine-like so as to be effectively identical. However, the simple task of circle-square discrimination is still a useful concept. For objects of the same size, it presents an undemanding classification task and consequently, if an ATR system is unable to achieve satisfactory performance on this problem, then acceptable performance on more complex shadow definitions is unattainable. Such a model also permits us to isolate the sonar characteristics such as contrast and resolution and their contributions to the performance of the ATR system.

We begin by establishing a property of the target recognition problem that will facilitate the calculation of the information theoretic measures.

Property 4.1 *For the communication system defined above, if y_j follows the same distribution for all $x_i, i \in \{1, 2, \ldots M\}$, then y_j does not convey any information that can be used for classification.*

PROOF: Since

$$I(X;Y) = H(X) - H(X|Y), \tag{4.1}$$

if $p(\mathbf{y})$ does not change for all x_i then $H(X|Y) = H(X)$, resulting in $I(X;Y) = 0$. □

Property 4.1 allows us to reduce the complexity of mutual information calculation by limiting the number of dimensions to the ones that contribute to the overall result, i.e. pixels with different probability distribution functions for different classes (see Figure 5). This result supports the intuitive belief that regions of the sonar image not related to the target can be disregarded when performing object recognition, however it also presents evidence that the shadow texture (as opposed to purely the shape) or artifacts (such as ringing) are effective cues for target recognition.

The class-conditional entropy calculation for the target recognition problem is therefore

$$H(X|Y) = -\underset{\mathcal{R}^{N^\Delta}}{\int \cdots \int} p(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N^\Delta}) H(X|Y = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N^\Delta}\}) \tag{4.2}$$

where N^Δ is the number of "difference" pixels in the shadow definition. For the Rayleigh distributions that characterize the ATR sonar problem, this calculation for class $X = x$ is

$$H(X = x|\mathbf{Y}) = -\underset{\mathcal{R}^{N^\Delta}}{\int \cdots \int} P(x) \log \left(\frac{p(\mathbf{y}|x)P(x)}{p(\mathbf{y}|x)P(x) + p(\mathbf{y}|\bar{x})P(\bar{x})} \right), \tag{4.3}$$

where $p(\mathbf{y}|x)$ follows a set of independent Rayleigh distributions:

$$p(\mathbf{y}|x) = A \times y_1 y_2 \ldots y_{N^\Delta} e^{-(\alpha_1 y_1^2 + \alpha_2 y_2^2 + \cdots + \alpha_{N^\Delta} y_{N^\Delta}^2)}, \tag{4.4}$$

$$\alpha_i = \frac{1}{2\beta_i^2}, \tag{4.5}$$

$$A = \prod_{i=1}^{N^\Delta} \frac{1}{\beta_i^2}, \tag{4.6}$$

and β_i is the mode parameter for the i^{th} pixel of class x, and $\bar{x}$ is the probability of the complement class of x, i.e. for the two-class problem, $P(\bar{x}_1) = P(x_2)$ and vice-versa. The total class-conditional entropy is $H(X = x|\mathbf{Y}) + H(X = \bar{x}|\mathbf{Y})$. Unfortunately, an analytical solution to the general, multi-dimensional class-conditional entropy cannot be found. Although Equation (4.4) is faster to evaluate than the Bayesian error integral (2.2), a Monte Carlo integration [17] was necessary to compute the value of $H(X|Y)$. Below, we examine a different bound where an analytical solution can be found.

Figure 6 shows the predicted Bayes error rate using the rate distortion bound of the circle-square test for a number of sonar design parameters. Three contrast ratios were considered, corresponding roughly to low (4 dB), medium (7 dB) and high (10 dB). Rather than repeat the calculations for endless combinations of object sizes and resolutions, we have coupled the two measures in a single quantity that is the number of along-track pixels on the target. As the pixels are assumed to be square, it is also the number of across-track pixels as well.

Figure 5: Differing sizes of circular and square patterns generated using the sonar imaging process described in Section 2. The top two rows shows the Rayleigh mode parameter for each pixel after the low pass filtering attributable to the division into resolution cells. A perfect image would have but two colours, background and shadow. Here we clearly see the blurring effect of the low pass filtering. The bottom row is simply the difference between the two first rows. Non-zero pixels in this image give some information that can help an algorithm make the correct classification. The higher the value, the greater the contrast (and the greater the information content).

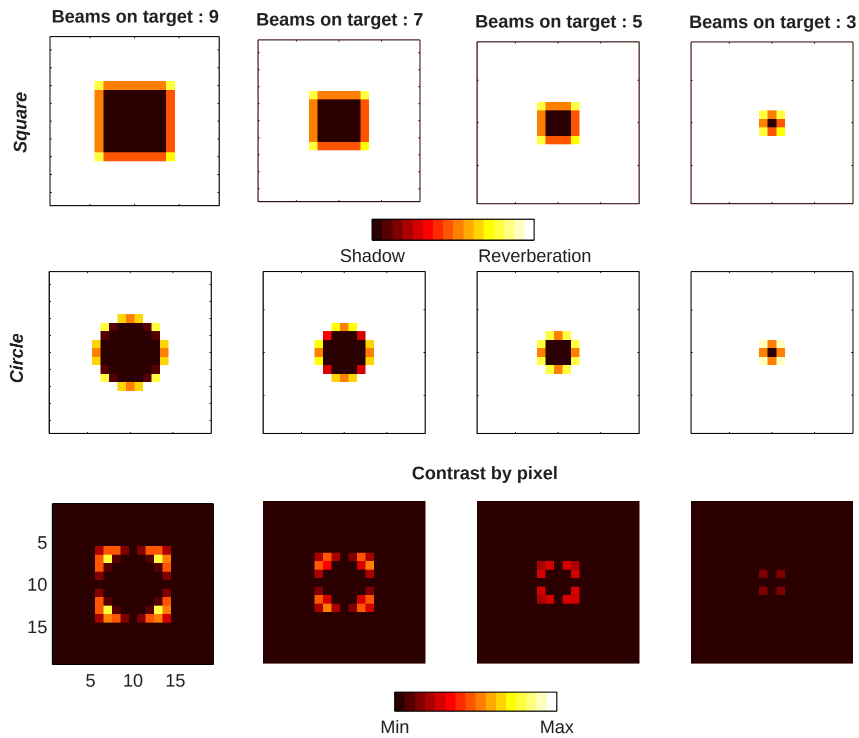
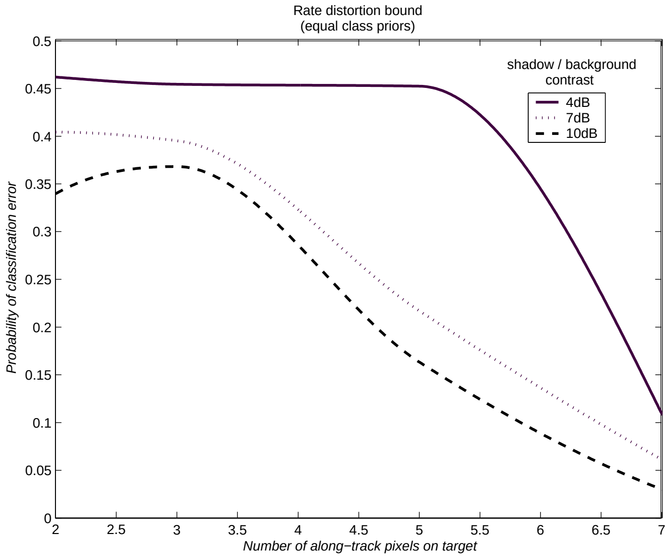


Figure 6: Rate distortion bounds for ATR performance using the number of along track pixels as a function of the probability of error for low, medium and high contrast sonar imagery.



4.2 The Chernoff information

So far we have considered the use of mutual information to estimate the Bayesian error rate by applying rate distortion theory. Although we have circumvented a direct evaluation of the Bayes error integral, we have shifted the problem to calculating another complex although simpler, multidimensional integral. There is also an upper limit on the probability of error called the *Chernoff bound*, defined as:

$$L^* \leq P(x_1)^\lambda P(x_2)^{1-\lambda} e^{C(p(\mathbf{y}|x_1),p(\mathbf{y}|x_2))}, \tag{4.7}$$

where

$$C(p(\mathbf{y}|x_1), p(\mathbf{y}|x_2)) \triangleq -\min_{0\leq\lambda\leq1} \log \left(\int_{\mathcal{R}^{N^\Delta}} \cdots \int p(\mathbf{y}|x_1)^\lambda p(\mathbf{y}|x_2)^{1-\lambda} \right), \tag{4.8}$$

is called the *Chernoff information* (see Thomas and Cover [14]). For a series of independent Rayleigh distributions, an analytical solution can be found ([18]), giving:

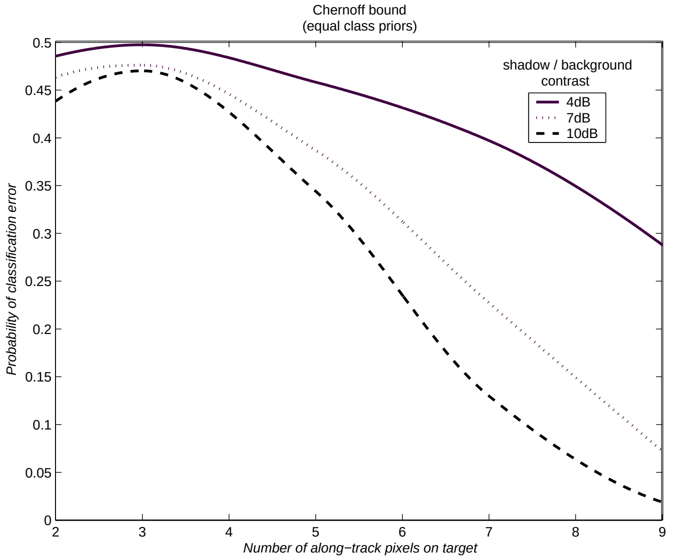
$$C(p(\mathbf{y}|x_1), p(\mathbf{y}|x_2)) \triangleq -\min_{0\leq\lambda\leq1} \log \left(A_1^\lambda A_2^{1-\lambda} \prod_{i=1}^{N^\Delta} \frac{1}{2\frac{\lambda}{2\beta_{1(i)}^2} + \frac{1-\lambda}{2\beta_{2(i)}^2}} \right), \tag{4.9}$$

where $\beta_{j(i)}$ is the i^{th} Rayleigh mode parameter for class $x_j, j \in \{1, 2\}$ and

$$A_j = \prod_{i=1}^{N^\Delta} \left(\frac{1}{\beta_{j(i)}^2} \right), j \in \{1, 2\}. \tag{4.10}$$

Finding λ for which the Chernoff information reaches a minimum allows us to bound $L*$ from above using Equation (4.7). The Chernoff bounds on the error probability for equal class priors for different contrast and resolution are shown in Figure 7.

Figure 7: Chernoff bounds for ATR algorithms using the number of along track pixels as a function of the probability of error for low, medium and high contrast sonar imagery.



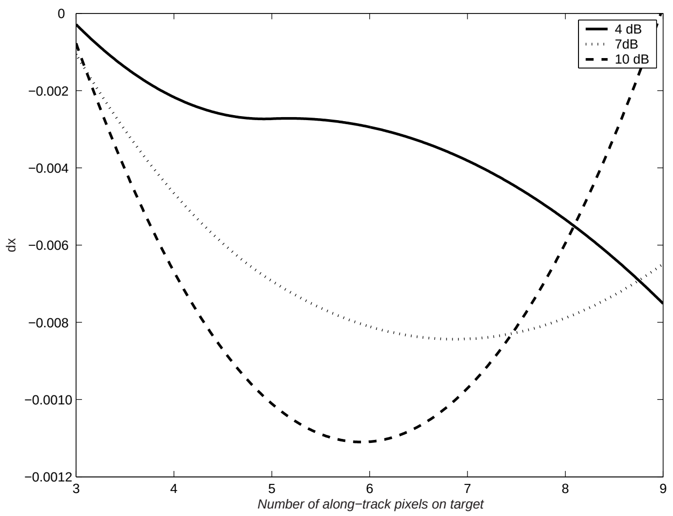
5 Observations

The performance bounds derived in the previous section give way to commentary. Our first remark is that the overall error rate does not decrease linearly as we increase the number of pixels on the target, which is itself directly correlated to the resolution of the sonar imaging system. This effect is perhaps due to the non-linear mixing of Rayleigh distributions for shadow/reverberation pixels when dividing into resolution cells. Further examination shows that gains in performance are more significant when combined with an increase in contrast. Certainly in all cases, better contrast (or signal to noise ratio) results in better accuracy regardless of resolution, however the gain is not simply a constant offset. There are often trade-offs to be made when determining resolution and contrast requirements of a sonar system and the performance bounds derived above for automatic target recognition algorithms imply that it is possible to optimize these parameters.

In the contrast cases shown, the “high” contrast is of primary interest since most modern minehunting sonars generally fall into this category. However, the low and medium contrast cases are not shown solely for academic purposes. For instance, in very shallow water operations depending on the environment or on areas of seabed with particular acoustic properties, shadows can be filled in by multipath or lost in the background even if the sonar itself is otherwise a high contrast sensor. Using the rate-distortion bounded error probability, predicted performance for a high contrast (10dB) sonar achieves a probability of correct classification of over 95% at around 7 pixels on the target.

Which increases in along track pixels most contribute to a reduction in error probability? Figure 8 shows the derivative of the Chernoff bounded probability for the three contrast situations used. In the high contrast case, increasing from 5 to 6 along track pixels on the target gives the largest single step increase in classification performance. For a medium contrast situation, the gains are greatest when increasing the number of pixels from 6 to 7 and for low contrast imagery gains become significant from 7 to over 9 pixels. Note that little to no gain can be expected over 9 along track pixels on a target in the case of a high contrast sonar.

Figure 8: The derivative of the Chernoff bounded error probability.



What of the use of post-processing techniques such as image enhancement algorithms designed to improve contrast after the imagine process, or of resampling methods intended to artificially increase resolution? One might assume that by applying such methods to a low resolution or poor contrast image that we could boost ATR algorithm performance with little expense. However, information theory prevents us from taking such shortcuts as is stated in what is known as the *information processing theorem*, which states:

Theorem 5.1 *If* $X \rightarrow Y \rightarrow Z$ *then* $I(X;Y) \geq I(X;Z)$.

Therefore no processing, random or deterministic, can increase the amount of information that a variable conveys about another. This theorem is vital in lending confidence to our derived performance bounds and avoids examination of countless implementations of image processing or sampling techniques that could possibly change the performance bounds. In practice, it is likely that sonar imagery would be processed in order to calculate features to present to an ATR system implementation. The intention here is rather to bound the performance of that system.

5.1 Human operators, mutual information and risk

The capacity for a computer algorithm to recognize targets is not the same as the ability of a trained human operator to classify man-made and natural objects using sidescan sonar imagery. The majority of navies employ human operators as the main decision maker during operations and the evaluation of the human aptitude for visually interpreting sidescan sonar imagery and the relation of this performance to the information theoretic measures established here could possibly be established. The human decision making process differs from optimal decision making (in the Bayesian sense) in that, training and confidence aside, decisions are made based on individually perceived levels of information and assumptions such as those made regarding the prior class probabilities or the associated risk, the bias of the operator or even the expectation of the mission outcome. Psychological aspects enter into play whenever humans are put “in the loop” and it may be possible to quantify these using a similar analysis as the one that has been proposed here for ATR algorithms.

As with human operators, the performance of an ATR system should not be reduced to a single metric, namely the probability of error. In military detection and classification tasks, such as minehunting, the consequences that result from classifying a non-target as a target, a *false alarm*, are typically additional time and resources spent on prosecution. However, missing a target signature will almost always be ensued by ramifications that can range from delays to seriously jeopardizing mission objectives. On the other hand, the target class is generally a *rare event*, occurring with far less frequency than the non target class. This search for equilibrium between costs and prior probabilities adds another layer of complexity to the sidescan sonar ATR problem. Costs, by their very nature, are subjective and usually impossible to quantify and accurate estimates of prior class probabilities are just as unattainable. In fact, both of these quantities are subject to change as the nature of the mission changes, and it is essential that ATR systems designed for military applications remain flexible.

6 Conclusions

We have presented a method for bounding the accuracy of automatic target recognition algorithms applied to sidescan sonar imagery using information theory. A method for calculating the information that is conveyed about an unknown target when ensonified using a sidescan sonar was proposed on the basis of a simple visual shape recognition task and this measure of information was then related to the Bayesian probability of error using rate-distortion theory and the Chernoff information. The method was used to predict an upper bound for the performance of automatic target recognition algorithms under varying environmental and noise conditions.

This study has resided for the most part in the theoretical domain and we have not attempted to carry out a comparison of commercial sonar systems, their suitability to be used as minehunting sonars or the performance of ATR systems in use. Certainly new sidescan sonars should be designed in such a way as to maximize the probability of success of ATR algorithms in a wide range of environments. This implies increasing contrast and resolution as modern mines become smaller and stealthier and for operations in shallow and very shallow water in particular.

Where lies the problem of successful automatic target recognition, with the sensor or the algorithms? No general study has been carried out to answer this question in practical terms. However, sensors currently in development are fast approaching the perfect classification horizon for current targets of interest and consequently ATR researchers can focus upon achieving the superior discriminatory capabilities that these sensors allow.

As sonar resolution increases, the traditional minehunting stages of detection and classification become blurred and questions about using high-frequency sidescan sonar for object recognition or identification are being asked. This study has established objective and unbiased performance prediction methods that will hopefully allow system designers to answer such questions.

Acknowledgements

Klein 5500 sonar imagery was collected at the RMS-TD trial in November 2002 by DRDC-Atlantic, Halifax, Canada.

References

[1] R. Balasubramanian and M. Stevenson, “Pattern recognition for underwater mine detection,” in *Proceedings CAD/CAC 2001*, (Halifax, Nova Scotia, Canada), November 2001.

[2] G. Dobeck and J. Cobb, “Fusion of multiple quadratic penalty support vector machines (QPSVM) for automated sea mine detection and classification,” in *Proceedings of the SPIE*, vol. 4742, April 2002.

[3] V. Myers, “Decision trees for computer-aided detection and classification of mines in sidescan sonar imagery,” Technical Memorandum TM 2002-144, Defence Research and Development Canada, Halifax, Nova Scotia, Canada, February 2003.

[4] S. Reed, Y. Pettilot, and J. Bell, “An automatic approach to the detection and extraction of mine features in side scan sonar imagery,” *IEEE Journal of Oceanic Engineering*, vol. 28, pp. 90–105, January 2003.

[5] D. Dudgeon, “ATR performance modelling and estimation,” Tech. Rep. 1051, MIT, Lexington, MA, December 1998.

[6] L. Horrowitz and G. Brendel, “Fundamental SAR ATR performance predictions for design tradeoffs,” in *Proceedings of the SPIE*, vol. 3070, pp. 267–284, April 1997.

[7] A. Horne, “An information theory for the prediction of SAR target classification performance,” in *Proceedings of the SPIE*, vol. 4382, pp. 404–415, 2001.

[8] S. Briles, “Information-theoretic performance bounding of bayesian identifiers,” in *Proceedings of the SPIE*, vol. 1960, pp. 255–266, 1993.

[9] M. Pinto, “Performance index for shadow classification in minehunting sonar,” in *UDT 97 Conference Proceedings*, (Hamburg, Germany), pp. 159–163, June 1997.

[10] F. Florin, F. Van Zeebroeck, I. Quidu, and N. Le Bouffant, “Classification performances of minehunting sonar: Theory, practical, results and operational applications,” in *UDT 2003 Conference Proceedings*, (Sweden), June 2003.

[11] R. Kessel, “Estimating the limitations that image resolution and contrast place on target recognition,” in *Proceedings of the SPIE*, vol. 4726, April 2002.

[12] C. Shannon, “A mathematical theory of communication,” *Bell Systems Technical Journal*, vol. 27, pp. 378–423, July 1948.

[13] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. New York, NY: John Wiley and Sons, 2001.

[14] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY: Wiley-Interscience, 1991.

[15] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[16] F. Kanaya and K. Nakagawa, "On the practical implication of mutual information for statistical decisionmaking," *IEEE Transactions on Information Theory*, vol. 37, no. 4, pp. 1151–1156, 1991.

[17] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi, *GNU Scientific Library – Reference Manual*, January 2002.

[18] V. Myers and M. Pinto, "Information theoretic bounds for the performance of atr algorithm performance for sidescan sonar target classification," Tech. Rep. SR-391, NATO Undersea Research Centre, La Spezia, Italy, 2004.