# Wavelet Analysis of Side Scan Sonar Imagery for Classification

John Impagliazzo*, Walter Greene* and Quyen Huynh**

*Naval Undersea Warfare Center, Division Newport
Newport, Rhode Island, USA
**Naval Surface Warfare Center, Dahlgren Division, Coastal System Station
Panama City, Florida, USA

## Abstract

*Wavelet-based techniques are presented for compressing side scan sonar images communicated from UUV's through band limited communication channels. Testing with an automatic classifier on reconstructed images demonstrates exceptional performance in preserving features required to identify mine-like objects. Wavelet domain automatic target recognition techniques are studied to further improve coefficient selection in compression. These techniques serve to enhance image fidelity in the vicinity of minelike objects while maintaining a high compression ratio.*

## 1. Introduction

Discrete wavelet transform analysis techniques are applied to side scan sonar imagery to develop techniques which efficiently characterize images and features for communication over limited bandwidth channels [1]. A block diagram of the proposed image compression algorithm is shown in figure 1. It combines choosing an optimal wavelet basis to preserve mine-like features as well as a methodology and criteria for reducing the coefficient matrix to a sparse array. Wavelet-based detection and classification methods are incorporated to identify coefficients associated with the local structure of mine-like objects. The surviving coefficient criteria are then weighted to preserve these coefficients when generating the sparse wavelet matrix for compression.

The techniques described here involve two different methods of analyzing wavelet packet coefficients for the presence of mine-like objects. The first method involves the matching pursuit algorithm. A dictionary of functions consisting of the wavelet coefficients corresponding to various object types is created. The method determines which function or combination of functions in the dictionary are most like the object in question. The second method involves a statistical analysis of the coefficients. By calculating the probability density functions (pdf's) of the coefficients, a likelihood ratio test can be performed to indicate the presence of a mine-like object. The advantages of these methods are that they use the wavelet-based time-frequency atoms already computed during the compression process rather than heuristic features, and that they can be used for many different types of image data. All that is required is a set of known objects from the images to create the pdf's or the dictionary functions.

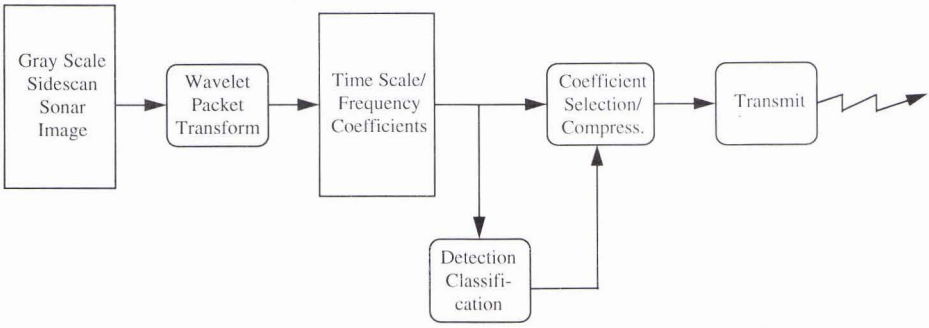


Figure 1. Block Diagram of Wavelet-Based Image Compression Algorithm

### 1.1 Overall Tactical Picture

Minehunting systems using side scan sonar will require 6 inch resolution in the direction of travel to avoid missing a 21 inch diameter mine-like object. A remotely-deployed minehunting system using an unmanned underwater vehicle (UUV) traveling at six knots could collect 511 sonar pings every 25 seconds. Also, if each sonar ping consists of 1024

samples at a 6 inch resolution, and port and starboard images were collected simultaneously, two 1024 by 511 pixel (picture element) images would be generated every 25 seconds. Without data compression, it could take as long as 45 minutes to transmit 25 seconds of data. To maintain real-time processing requirements, compression ratios of from 25:1 to 100:1 are required for an acoustic communications link or band limited radio frequency (RF) or satellite link.

When a side scan sonar pings, an underwater mine or similar object moored or sitting on the ocean bottom will prevent sound from the sonar system from reaching the sea floor for some distance beyond the object. This produces a characteristic highlight and adjacent shadow highly localized in the side scan sonar image. An automatic target recognition algorithm, therefore, might categorize the dimensions and relative intensity of the highlight and shadow to determine if the object should be classified as mine-like. This algorithm would, however, be susceptible to false alarms if the data compression techniques employed produced artifacts in the compressed image resembling the highlight-shadow characteristic. This is an issue with the conventional Joint Photographic Expert Group (JPEG) technique at high compression ratios.

The Navy Imaging Database at the Naval Surface Warfare Center, Dahlgren Division (NSWCDD), Coastal System Station (CSS), was used in the research described here. The images in this database were produced by a sidescan sonar towed by a helicopter. The database consists of 60 images, 30 of which have been designated as training images and 30 as testing images. Fifteen of the 30 training images contain one mine signature each, 16 of the 30 testing images contain one mine, and one testing image contains two mines, for a total of 33 mine signatures. The mines in this database are cylindrical bottom mines that typically have both a highlight and shadow signature. For this sonar, a typical mine signature has around 36 pixels in the highlight region and about 120 pixels over the shadow zone, but this varies greatly. The data for each image consist of a matrix of 1024 by 511 8-bit unsigned integers. For processing purposes, the last column is duplicated to give 512 columns.

### 1.2 Side Scan Sonar Images

Figure 2, image si000206 (*s*onar *i*mage number 206), is an example from the Navy Imaging Database, which is referred to here as sonar0. Near range is at the top of this figure, with far range at the bottom. Cross-range (the direction of travel) is horizontal across the image. The near range appears to be smooth while the far range is rough; abnormalities appear as striations in the last 5% of range. The apparent smoothness of the near range is due to the higher angle of sound incidence. In reality, the roughness and tracks are distributed uniformly over the image. The dark tracks, in many cases, are caused by fishermen dragging shrimp nets. The orientation of the tracks is also evenly distributed, but more horizontal tracks show up because of their acoustic shadow. The axes shown are in pixels. Resolution in the direction of travel, approximately 15 cm (6 inches), is a function of the speed of the tow vehicle and the round-trip time of the ping to the farthest range. Resolution in range, also approximately 15 cm, is a function of the number of beams and the maximum range. The minimum resolution in range is also determined by the size of objects which need to be detected. Two mines are found at coordinates (427, 370) and (864, 159) and are shown in the blow ups in figure 2. The first mine, at (427, 370), has a modest horizontal highlight and a pronounced shadow. The second mine, at (864, 159), is difficult to see. It has a small strong highlight and a small shadow, which is somewhat disguised because it is located on the edge of a track.

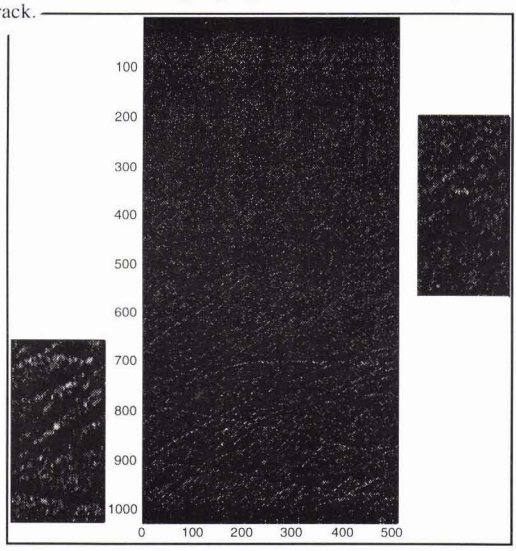


Figure 2. Original Image si000206 with Blowups of Mines

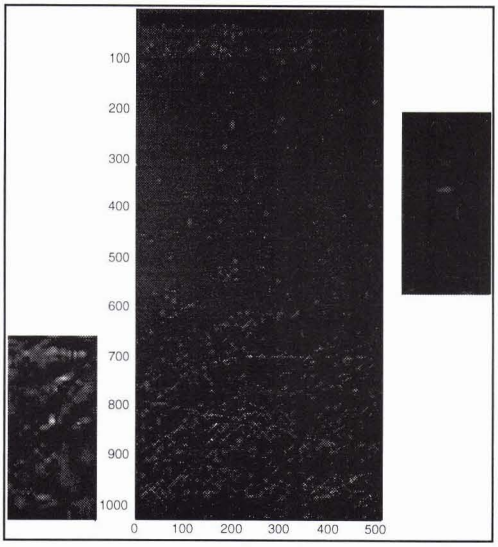


Figure 3. si000206 after 100:1 Compression with Blowups of Mines

## 2. Orthogonal Wavelets

### 2.1 Wavelet Transform Analysis

Efficient characterization of images is possible with wavelet time-frequency analysis techniques because of the local support property. Features well-localized in space are well represented by the set of coefficients overlapping the features' location. Other advantages of the discrete fast wavelet transform is that their complexity is of O(n) and they are implemented through finite impulse response filters. All of the discrete fast wavelet transforms provide perfect reconstruction. The primary differences among various types are the length of the filter, the required precision of the filter coefficients, and the relationship between forward and inverse filters. Orthogonal wavelet transform filters implement the same filtering at the forward and inverse transforms.

Evaluation of compression performance is based on the visual fidelity of the image and on comparisons of automatic classifier performance on the original image with performance on a duplicate image reconstructed from the substantially compressed image file. Figure 3 shows image si000206 after 100:1 compression. The automatic classifier used to evaluate the images was developed by Dr. Gerald Dobeck and his colleagues at the Naval Surface Warfare Center, Dahlgren Division, Coastal Systems Station, Panama City, Florida, USA [2].

The impressive success of wavelets is due mainly to the discovery of multiresolution analysis by Mallat [3]. Multiresolution analysis constitutes a useful functional analysis tool in wavelet theory and leads to the development of the very fast pyramid scheme to compute the wavelet coefficients. Although the fleet side scan sonar image is processed in two dimensions with the orthogonal wavelet algorithm, a brief description of the theory is presented here for one dimension. In practice, the one dimensional algorithm is applied twice, first to each row of the input matrix and then to each column of the row processed matrix.

In the continuous wavelet transform, for a given function $x(t)$, the coefficients are defined as follows:

$$C_x(a,b) = \langle x, \psi_{ab} \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi_{ab}\left(\frac{t-b}{a}\right)dt, \quad b \in R, \quad a > 0. \tag{1}$$

The parameters $a$ and $b$ have the effects of dilation and translation respectively. To discretize the transform in the time/frequency plane, we let $a = 1/2^j$ and $b = k/2^j$, where $j, k \in Z$. The coefficients thus become

$$C_x\left(\frac{1}{2^j}, \frac{k}{2^j}\right) = 2^{j/2} \int_{-\infty}^{+\infty} x(t)\psi_{jk}\left(2^j t - k\right)dt. \tag{2}$$

Finally, discretizing in time gives

$$C_x\left(\frac{1}{2^j}, \frac{k}{2^j}\right) = 2^{j/2} \sum_{n=1}^{N} x[n]\psi_{jk}\left(2^j n - k\right), \tag{3}$$

where $N$ is the length of the input vector $x[n]$. The function $\psi \in L^2(R)$ is called an orthogonal wavelet if the family $\{\psi_{jk}\}$ is an orthonormal basis of $L^2(R)$: that is, $\langle \psi_{jk}, \psi_{lm} \rangle = \delta_{jl}\delta_{km}$, where $\delta$ is the Kronecker delta function. Note that the wavelet function is in $L^2(R)$ so that it has finite support in time. This differs from the trigonometric functions in Fourier analysis and gives the wavelet its ability to produce time information as well as frequency.

The discrete orthogonal wavelet algorithm is actually implemented as a series of convolution and decimation operations with discrete-time wavelet filter banks, such as those developed by Daubechies [4]. We adopt the compactly-supported wavelet, Daubechies 6. The length of the wavelet was chosen fairly arbitrarily. However, some of our classification work has suggested that longer wavelet filters (e.g., Daubechies 20) tend to miss some mines, while shorter wavelet filters (e.g., Daubechies 2) have many false alarms.

The discrete wavelet transform is implemented by a series of convolution and decimation operations with a pair of filters. Let $x = \{x[k]\}_{k=0}^{K-1}$ be the discrete version of input signal $x(t)$ of length $K = 2^n$. This can be either a row of the image or a column of the coefficients after the rows have been processed. In the fast discrete wavelet transform, the signal $x$ is first decomposed into low and high frequency bands by the convolution-decimation (subsampling by two) operations of $x$ with the pair of a low-pass filter $G = \{g_k\}_{k=0}^{L-1}$ and a high-pass filter $H = \{h_k\}_{k=0}^{L-1}$, where $L$ is the length of the filter. In orthogonal wavelets, the length of the two filters is the same. The filters $G$ and $H$ satisfy the orthogonality conditions:

$$GH^* = HG^* = 0, \quad \text{and} \quad G^*G + H^*H = I. \tag{4}$$

$G$ and $H$ are Quadrature Mirror Filters (QMFs), which allow perfect reconstruction. The decomposition process continues iteratively on the resulting low frequency bands and each time the high frequency bands are left intact. The iteration stops when there is one low frequency coefficient and one high frequency coefficient. As a result, the frequency axis is partitioned *smoothly* and *dyadically* in an octave-band fashion, as shown in figure 4. Figure 4 shows the phase plane produced by the wavelet transform. The wavelet transform converts one-dimensional data into two-dimensional data. The horizontal, or $t$, axis can be labeled by time or position, depending on the nature of the data, and increases to

the right. The vertical, or $f$, axis is usually labeled frequency, or scale, and increases upward. Different spatial resolutions are given to different frequency bands. Low frequency with low spatial resolution is at the bottom, while, toward the top of the figure, the frequency resolution is decreased and the spatial resolution is increased. The entire phase plane is covered by disjoint cells of equal area which are called Heisenberg cells [5]. The uncertainty principle can be interpreted as a rectangular cell located around a position in the phase plane, $(t,f)$, that represents an uncertainty region associated with $(t,f)$. The total number of cells is equal to the dimension of the input vector.

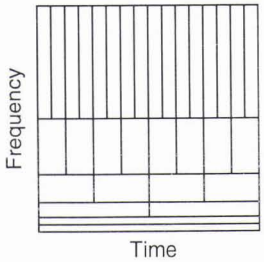


Figure 4. The Time/Frequency Phase Plane for the Standard Wavelet Bases

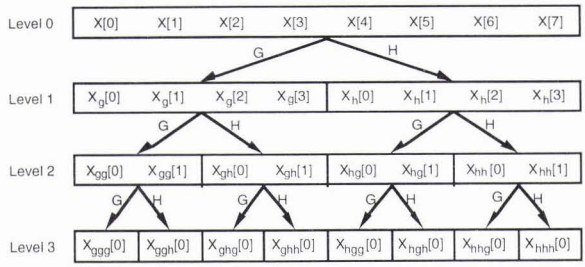


Figure 5. The Complete Wavelet Packet Decomposition Tree

## 2.2 Wavelet Packet

For the wavelet packet transform [6], the high frequency band, which is left intact during each iteration of the wavelet transform, is also decomposed into finer frequency bands. Figure 5 depicts the entire wavelet packet decomposition tree. Level 0 represents the original signal $x$. The level 1 decomposition generates $Hx$, labeled $x_h$, and $Gx$, labeled $x_g$. $Gx$ represents the low frequency band and $Hx$ represents the high frequency band. Applying the low-pass filter $G$ and the high-pass filter $H$ to both low and high frequency bands, we obtain four frequency bands $H^2x$, $GHx$, $HGx$ and $G^2x$ ordered in decreasing frequency. This is the level 2 decomposition. The decomposition process continues to the maximum depth of $J$, where $J = \log_2(n)$, and finer frequency resolutions are obtained toward lower levels. Since each decomposition level generates equal boxes corresponding to a uniform partition of the frequency axis, the time axis is also windowed uniformly. Hence, the extent of the support of each basis function is essentially constant for a decomposition level. Note that each level forms an orthonormal basis on which to project the side scan sonar image.

The full wavelet packet transform produces a wavelet packet tree structure containing many more coefficients than are needed to reconstruct the image. A basis vector of coefficients should be selected. The basis vector of transform coefficients can be constructed by selecting all the coefficients in a level, referred to as a level basis, or by selecting coefficients from different levels to obtain a "best" basis for the input data vector $x$. Here "best" can be in whatever measure desired. Certain rules apply which constrain the selection of coefficient sets to those having parent-child relationships within the wavelet packet decomposition tree [6]. For the fleet side scan sonar application, a level basis was selected which best characterized the spatial and spectral characteristics of the targets. The transform vector contains the same number of coefficients as the dimension of the input vectors from the original image, so this does not lead to compression. However, the forward transform increases the amount of energy contained in some individual coefficients and decreases others. Increasing the magnitude of individual coefficients increases the data requirements of the transform, but because the energy is conserved in the transform, most of the remaining coefficients are very small, approaching zero. Since deleting small coefficients does not significantly affect the total energy in the image, it will not cause significant distortion in the reconstructed image. The surviving coefficients now comprise the transform vector and can be encoded for compression using a zero run-length encoding technique.

## 2.3 Orthogonal Transform Algorithm

The algorithm for the orthogonal wavelet case is as follows: the one dimensional wavelet packet algorithm is applied twice, first to each row of the input matrix and then to each column of the row processed matrix. Only the level 3 basis coefficients for the lower half of the frequency spectrum are retained. This provides some filtering of high frequency noise. Next, the coefficients are sorted by magnitude for the spectrum of interest. Processing time is substantially reduced by only processing the pyramid algorithm to level 3 and by limiting the spectrum of interest to the lower half of the image bandwidth. The number of coefficients required to give the desired compression ratio is calculated; only this number of the largest coefficients, along with their locations in the coefficient matrix, are retained in the compressed image file. Zero run length encoding is currently used to encode coefficient locations. For reconstruction, the surviving coefficients are placed into their proper locations in the coefficient matrix while the rest of the coefficients are set to zero. The inverse wavelet packet transform is performed on each column, replacing the data that was there, and then on each row to produce the reconstructed image. This is the reconstructed image which is processed by the classification algorithm.

### 2.4 Overall Performance

The fleet side scan sonar images were compressed 25:1, 50:1, and 100:1 using the orthogonal wavelet technique. The original image si000206 and si000206 after 100:1 compression are shown in figures 2 and 3. The regions containing each of the two mines are shown in enlargements. At a compression ratio of 25:1, the most noticeable difference from the original image is the reduction of high frequency texture information from the background. This effect increases at 50:1 and, as can be seen at 100:1, the texture information has been mostly removed from the near range region. A significant weakening of the highlight or shadow contrast is not observed in either of the mines as the compression ratio is increased to 100:1.

Comparison of the automatic classifier on the original and wavelet compressed data sets demonstrated that only minimal degradation was realized in the compression process. On the original data set, the classifier demonstrated a probability of detection and classification (PdPc) of 91% and 0.28 false alarms per image (FA/image). For the data set compressed by the orthogonal transform algorithm, PdPc = 83% and FA/image = 1.1.

## 3. Matching Pursuit

### 3.1 Description

Mallat and Zhang [7] introduced a matching pursuit algorithm which allows a signal function to be decomposed into a linear expansion of functions belonging to a redundant dictionary of waveforms. Here, these waveforms are time-frequency atoms computed from sample mine and non-mine images. The assumption is that the time-frequency atoms consist of a pattern of wavelet coefficients related to the local structure of the target. The structure can be difficult to detect from individual coefficients because the forward transform diffuses the information across the basis. The advantage of the wavelet domain is that the signal waveform and dictionary waveform can be compressed using wavelet image compression techniques, preserving information about the local target structure without making assumptions about the nature of the target. This compression, in turn, minimizes the computational requirements on the matching pursuit algorithm.

Mallat and Zhang define a family, $D = (g_\gamma)_{\gamma \in \Gamma}$, of vectors in $H$, $H = L^2(R)$, such that $\|g_\gamma\| = 1$. Letting $f \in H$, a linear expansion of $f$ is computed over a set of vectors selected from $D$ to best match the local target structure. This is done by successive approximations of $f$ with orthogonal projections on elements of $D$. Let $g_{\gamma_0} \in D$. The vector $f$ can be decomposed into

$$f = \langle f, g_{\gamma_0} \rangle g_{\gamma_0} + \mathrm{R}f, \tag{5}$$

where R$f$ is the residual vector after approximating $f$ in the direction of $g_{\gamma_0}$. The vector $g_{\gamma_0}$ is orthogonal to R$f$, hence

$$\|f\|^2 = |\langle f, g_{\gamma_0} \rangle| + \|\mathrm{R}f\|^2. \tag{6}$$

To minimize $\|\mathrm{R}f\|$, $g_{\gamma_0} \in D$, is selected such that $|\langle f, g_{\gamma_0} \rangle|$ is maximized.

Considering the iterative approach, let $\mathrm{R}^0 f = f$. To compute the residue $\mathrm{R}^n f$ at the n$^{\text{th}}$ iteration, for $n \geq 0$, a vector $g_{\gamma_n} \in D$ is chosen which best matches the residue $\mathrm{R}^n f$. The residue $\mathrm{R}^n f$ is decomposed into

$$\mathrm{R}^n f = \langle \mathrm{R}^n f, g_{\gamma_n} \rangle g_{\gamma_n} + \mathrm{R}^{n+1} f, \tag{7}$$

which defines the residue at the order $n+1$. Since $\mathrm{R}^{n+1} f$ is orthogonal to $g_{\gamma_n}$,

$$\|\mathrm{R}^n f\|^2 = |\langle \mathrm{R}^n f, g_{\gamma_n} \rangle|^2 + \|\mathrm{R}^{n+1} f\|^2. \tag{8}$$

Extending this decomposition up to order $m$, equation (7) yields

$$f = \sum_{n=0}^{m-1} \langle \mathrm{R}^n f, g_{\gamma_n} \rangle g_{\gamma_n} + \mathrm{R}^m f, \tag{9}$$

and equation (8) yields the energy conservation equation

$$\|f\|^2 = \sum_{n=0}^{m-1} |\langle \mathrm{R}^n f, g_{\gamma_n} \rangle|^2 + \|\mathrm{R}^m f\|^2. \tag{10}$$

The original vector $f$ is decomposed into a sum of dictionary elements that are chosen to best match its residues. Although the decomposition is nonlinear, it maintains an energy composition as if it were a linear orthogonal decomposition.

In the matching pursuit algorithm for target classification, the inner product of the signal function with each of the dictionary waveforms is computed. The waveform which best matches the signal function is selected for the iteration and a residue is computed from the signal function. The residue is formed by subtracting the selected waveform, scaled by

the correlation coefficient, from the signal function to produce a new signal function for the next iteration. After the last iteration the signal function is represented as a linear expansion of the scaled dictionary waveforms. Target-like objects are discriminated from non-target signal functions by comparing the energy in the dictionary's target waveforms to that of the non-target waveforms. The class associated with the greater energy is assigned to the signal waveform.

### 3.2 Neural Network

The classification of the target-like signal functions can be further refined by a back-propagation neural network. In order to implement the matching pursuit/neural network classifier, it was necessary to divide the training set of data into two subsets, A and B. Half of the training set, subset A, was used as target waveforms for the matching pursuit dictionary. Non-target waveforms were also in the dictionary and were selected from areas away from the target from the same target file. The remaining half of the training set, subset B, was processed using the matching pursuit algorithm having subset A in the target dictionary. These results were then scored to form two lists for training the neural network, a list of the functions for correctly classified targets and a list for the false alarms. The list of functions for correctly classified targets was augmented by an additional set generated from targets in training subset B with offset centers. Offsetting the target center in a 3 by 3 pattern increased the number of target waveforms by a factor of nine. Each displaced center was located four pixels horizontally and vertically from its neighbor. The target and false alarm lists were used to train a neural network to discriminate targets from false alarms for the limited set of target-like signal functions classified by the matching pursuit algorithm.

### 3.3 Libraries

The sonar0 training set consisted of 30 image files, 15 containing bottom targets. In order to decide how to group the targets into subsets A and B, each target in the training set was processed by a matching pursuit algorithm having a dictionary containing the other 14 targets. This test was conducted to identify target images having unique features and those have features redundant with other target images in the dictionary. Two of the targets were found to have unique characteristics based on the results of the matching pursuit processing. A minimum set of 5 targets, including the unique targets, was selected for the matching pursuit dictionary. This approach maximized the number of target samples available to train the neural network stage of the classifier. The dictionary therefore consisted of the 5 target samples and 5 non-target samples randomly selected from the same image file. To scan an image file, a 32 by 32 pixel subimage was selected. Therefore each of the images in the dictionary also corresponded to a 32 by 32 subimage. The number of entries in the dictionary was increased nine fold by duplicating each of the target entries in a 3 by 3 offset pattern, where each displaced center was located four pixels horizontally and vertically from its neighbor. This improved the efficiency of scanning an image file by stepping the subimage processed as the signal function in 8 pixel increments. The sonar0 side scan sonar images contain 512 by 1024 pixels. Scanning a 32 by 32 subimage through a complete side scan sonar image required 8,196 increments.

### 3.4 Results

The signal functions and dictionary waveforms for the matching pursuit algorithm were compressed wavelet transformed subimages. The Daubechies 6 wavelet packet was used with filtering at levels 1 and 2. This reduced the 32 by 32 coefficient matrix to a 16 by 16 and 8 by 8 which were reformatted to a column vector. This allowed the projection of the signal function onto the dictionary waveforms to be computed as an inner product with the dictionary vectors.

Because of the limited number of target samples in the training set, only modest results were realized after preliminary testing of matching pursuit on the sonar0 test set. With filtering at level 1, PdPc = 89% and at level 2, PdPc = 78%. The neural network further reduced the FA/image by 50%, but dividing the training set into subset A and B degraded the performance of both classifiers. To implement both a matching pursuit dictionary and a neural net training set, unique targets existing in the training set therefore could only be assigned to one or the other subset. The subset that did not have information on the unique feature did not detect targets which would have otherwise been found.

## 4. Wavelet Coefficient Baysian Classifier

### 4.1 Generating Probability Density Functions for Coefficients

There are many more coefficients in the wavelet tree structure, shown in figure 5, than are needed for reconstruction of the image or for classification. The question is which coefficients should be used for classification. The matching pursuit algorithm answers this by using a subset of the coefficients in a basis based on compression. The selection of the basis was fairly arbitrary. The statistical classifier described here is an attempt to determine the ability of individual coefficients in the wavelet packet transform to discriminate between two classes of objects. There are a total of $n\log(n)$ coefficients plus the $n$ original data values themselves to choose from. In general, the number of inputs to a classifier will be less than or equal to the number of data points in the original signal. In order to determine which coefficients are best for classification, the probability density function (pdf) for each coefficient can be estimated. By generating the conditional pdf's for each coefficient, one can develop the optimal, or Baysian, classifier. The problem here is to determine the presence or absence of a mine-like object. Thus two conditional pdf's correspond to each coefficient. Transforming the image when there is a mine present, one obtains the values of each coefficient for the mine-present case. By transforming enough images with mines, one can use the distribution of the magnitude of each coefficient as an estimate of its pdf when a mine is present. Transforming an image with no mine produces the values for each coefficient

for the no-mine case. With images where no mine is present, the distribution of the coefficient magnitudes is an estimate of the pdf for the no-mine case.

There are two problems which come up immediately with this procedure. First, the number of images available with and without mines is very limited. To get a good estimate of the pdf's, a large number of examples is needed in each case. Second, not every coefficient is affected by the presence of a mine in the image. Because of the resolution in position of the wavelet transform, the extent of the coefficient may not be on the mine in the image. In fact, if the entire image is transformed, there are very few coefficients which are actually affected by the presence of a mine. The others should all be considered examples of non-mine coefficients.

The first problem is handled by not relying on an accurate pdf for each coefficient. Since there are a large number of coefficients, one can compensate for the inaccuracies of any one pdf by using many of them. There are two methods of handling the second problem. First, one can take the transform of only a small window of the image, approximately the size of a mine. Then, most, if not all, of the coefficients are affected by a mine in the window. The window can be slid through the image and a classification made for each location of the window. This method works well, but is very costly in the number of calculations needed. A second method involves only a single transform of the image. In this method, one transform is done on the entire image producing a large set of coefficients. Instead of scanning through the image in position, the coefficients from a location in the image are determined and only those are fed to the classifier. In effect, the image is scanned by selecting the proper set of coefficients in the right order.

### 4.2 Discrimination Threshold

In the binary classification case there are the following two hypothesis:

1) $H_0$ - no mine present, and

2) $H_1$ - at least one mine present.

The Bayesian test can be written as

$$H_0 \text{ if}: \frac{p_{A|H_1}(R|H_1)}{p_{A|H_0}(R|H_0)} < \eta, \quad H_1 \text{ if}: \frac{p_{A|H_1}(R|H_1)}{p_{A|H_0}(R|H_0)} > \eta \tag{11}$$

where $p_{A|H_0}(R|H_0)$ and $p_{A|H_1}(R|H_1)$ are the conditional probability densities and $\eta$ is the threshold [8]. The ratio on the right is the likelihood ratio and is a random variable. The threshold is a number based on the a priori probabilities and the cost of each course of action. If the likelihood ratio is greater than $\eta$, the $H_1$ hypothesis is chosen and if the likelihood ratio is less than $\eta$, the $H_0$ hypothesis is chosen. Because it is difficult to assign costs and a priori probabilities in this problem, the Neyman-Pearson test is used. In this test, the conditional probabilities, $P_F$ , probability of false alarm, and $P_D$ , probability of detection, are used. A value of $P_F$ is specified and this determines the value of $\eta$.

### 4.3 Selection of Coefficients for Classifier

Given a desired probability of false alarm, the pdf's of the coefficients and the threshold determine a probability of detection. This could be done for individual coefficients, although that was not done here. At a particular level in the wavelet packet transform, and at a specific frequency (a frequency bin), there is a set of coefficients which depend only on position. Each coefficient corresponds to the same waveform which only differs by a shift in position. Since the probability of the mine being at any particular location is the same, it doesn't make sense to pick one coefficient out of the set as being better than the others for classification. Instead, all of the coefficients in the set which only differ by a shift are grouped into one pdf. This produces more examples for each pdf and so helps to improve the estimate. The pdf's now provide information about which frequency bins are best for classification.

### 4.4 Classification

Because the pdf's for the mine and non-mine cases overlap and are not perfect, there would be considerable missclassification using any one pdf. To improve the classification, many pdf's are used. To classify an area of the image, the magnitude of all of the coefficients which overlap that area are checked via the Bayesian test. If a majority indicate the presence of a mine, the area is classified as mine-like. To reduce the calculations required, only coefficients with the best pdf's, those with the highest $P_D$, can be used for the voting. The overall classification procedure is as follows. The entire image is transformed producing a matrix of coefficients the same size as the image. The bases used for this transform should contain the coefficient bins which were determined to be best for classification. Each coefficient in the transform is compared to the threshold as shown in equation (11) for its particular frequency bin, and a classification is made. All the coefficients which detected a mine are compared to see if they fall on the same location in the image. If enough mine classifications are at the same location, that location is classified as mine-like.

### 4.5 Results

Preliminary results produced using the statistical classifier have been encouraging. The method described using the transform of only a small window of the image, approximately the size of a mine, was run on each mine in the training set to generate the pdf's for the coefficients. Because of the small size of the window, the pdf for every coefficient at every level was estimated. The best pdf's were then used to classify the images in the test set. The receiver operating

characteristic (ROC) curve for the results is still being generated. Work has started on the transform coefficients from the whole image. Observations of the mine and non-mine cases show that pdf's for each case appear to be very similar. The best pdf's for classification have yet to be determined.

## 5. Conclusions

It has been shown that wavelet packet based transforms are an effective compression scheme. A high level of automatic classifier performance on reconstructed side scan sonar images, compressed by as much as 100:1, demonstrates that these schemes effectively preserve the classification features of underwater mines. Promising results have also been presented for detection and classification techniques in the wavelet domain as a guide to the selection of coefficients in the compression process. The matching pursuit algorithm, which attempts to find a best match in a library of coefficient sets, and the statistical analysis, which attempts an optimal discrimination based on individual coefficients, both have shown good results in detecting mine-like areas. Correctly identifying coefficients related to the local structure of mine-like objects provides the ability to enhance the fidelity of the associated areas in the reconstructed image, while maintaining the required compression ratio.

Further analysis will be conducted using a larger side scan sonar data set to resolve problems realized when there are only a small number of mine signatures available, some of which are unique. Redundant dictionaries of time-frequency atoms over multiple bases will also be explored. Even with improvements made, operator intervention will continue to be needed to reconcile false alarms from detected mines. The voluminous amounts of data collected by a 6 knot UUV will still generate a number of false alarms per hour. The enhanced fidelity provided by the methods examined in this paper furnishes operators with a higher quality image with which to reconcile targets and false alarms.

## Acknowledgments

The authors would like to thank ONR321TS, Dr. Randy Jacobson, for sponsoring this work, Dr. Gerald Dobeck and his colleagues at the Naval Surface Warfare Center, Dahlgren Division, Coastal Systems Station, for their automatic classifier, and Dr. Börn Jawerth and his colleagues at the University of South Carolina and Summus Ltd. for their work with us using the biorthogonal wavelet transform.

## References

[1] J. Impagliazzo, W. Greene, G. Dobeck, Q. Huynh, L. Smedley, B. Jawerth, and L. Mygatt, "Wavelet Based Data Compression for Communication of Side Scan Sonar Images", JUA, July 1997.

[2] G.J. Dobeck and J.C. Hyland, "Automated Detection and Classification of Sea Mines in Sonar Imagery," in *Proceedings of the SPIE 11-th Annual International Symposium on Aerospace/Defense Sensing, Simulation and Control,* 21-24 April 1997, Orlando, Florida, Vol. 3079: Detection and Remediation Technologies for Mines and Minelike Targets.

[3] S. Mallat, "Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(R)$", Trans. Amer. Math. Soc. **315**, 69-88, 1989.

[4] I. Daubechies, *Ten Lectures on Wavelets*, volume 61 of CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM (1992).

[5] R.R. Coifman and M.V. Wickerhouser, "Entropy-Based Algorithms for Best Basis Selection," IEEE Trans. Inf. Theory, **38** (2), March 1992.

[6] R.R. Coifman, Y. Meyer, and M.V. Wickerhauser, "Wavelet Analysis and Signal Processing," in *Wavelets and their Applications*, eds. Ruskai et al. (Jones and Bartlett, Boston 1992), pp. 153-178.

[7] S G Mallat and Z Zhang, "Matching Pursuit With Time-Frequency Dictionaries," IEEE Trans. on Sig. Proc., Vol 41, no. 12, pp. 3397-3415, December 1993.

[8] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York, John Whiley & Sons, Part I Detection, Estimation, and Linear Modulation Theory, ch. 2, pp. 23-36, 1968.