Target detection in side-scan sonar images: expert fusion reduces false alarms

Nicola Neretti, Nathan Intrator and Quyen Huynh

Abstract— **We integrate several key components of a pattern recognition system for a mine-like targets detection problem. These include several image enhancements, post-processing and multi-expert fusion. The image enhancement includes wavelet de-noising and classical computer vision methods such as nonlinear and adaptive equalization and other filters. Our approach attempts to make the different experts as independent as possible so as to maximally exploit their fusion.**

Results are given on a shallow water mine-like targets detection problem using sonar data.

I. Introduction

Detection of mine-like targets from sonar data is a challenging problem due to the large variability in background clutter and the large variability in object appearance. Shallow water detection involves addressing the varying shape of the ocean surface and its vegetation. We address these issues by varying equalization methods, wavelet de-noising and image enhancement via difference of Gaussian filters.

We exploit the fact that ensemble of experts improve the overall performance of individual experts if the errors made by the individual experts are independent. We demonstrate how experts are made independent and how their fusion is feasible.

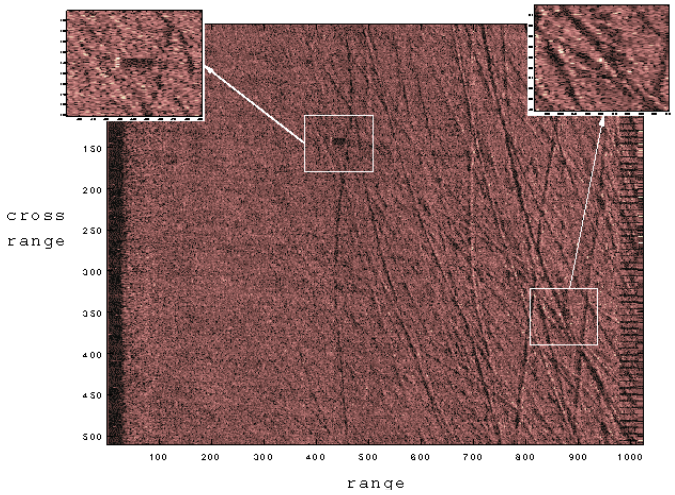

Fig. 1. Example of a side-scan sonar image containing two mine-like targets.

Nicola Neretti and Nathan Intrator are with the Institute for Brain and Neural Systems, Brown University, Providence, RI 02912, US. Quyen Huynh is with the Coastal Systems Station, Naval Surface Warfare Center, Panama City, FL 32407-7001. E-mail: HuynhQQ@ncsc.navy.mil.

II. Side-Scan Sonar

The sonar data is collected by a moving sonar fish, which emits an acoustic wave at regular intervals and records the reflected wave. An image of the sea floor is reconstructed from the acoustic waves. Objects observed at different distances from the sonar fish will present different intensities and shapes. The farther an object is from the sonar, the longer its shadow. The image is therefore divided into three regions (in the range direction) and filter parameters are optimized separately in each region.

We used two different databases to test our denoising techniques. The first one consists of a 60-image set from a side-scan sonar (Sonar-0). They are encoded as 8-bit gray scale images, 1024 range cells by 511 cross-range cells. The 60 images contain 33 targets; some contain more than one target while others contain no targets. Non-target objects which look as targets appear throughout the images. A typical mine-like target consists of a strong highlight on its left side and a long shadow down range on its right side. Unfortunately the presence of clutter can mask this structure. The second database we used consists of of an 80-image set from a side-scan sonar (Sonar-5) containing 34 targets. The images have 640 range cells by 1024 cross-range cells, and are encoded in 8-bit gray scale. This is a more difficult database since it exhibits drastic changes in background clutter. Both databases have been collected at the Naval Surface Warfare Center (NSWC) and processed by Dr. Gerry Dobeck.

Real sonar image data is preferred over simulated sonar data because sonar simulations are expensive and do not capture all the critical dynamics associated with actual sonar images.

III. The Steps of the Detection Process

We build several detectors, each based on a combination of various steps. In this section we describe each step in detail.

A. Image Normalization

A classical method for image normalization is histogram equalization. The image is transformed so that its histogram is as flat as possible in order to have roughly the same number of points in each intensity band. The top right image in Figure 2 shows the result: the background structure of the image is still present. We tried to apply a local version of histogram equalization, dividing the image into square cells and equalizing each cell separately. The result (Figure 2, bottom left) is an image that has discontinuities corresponding to the segmentation process;

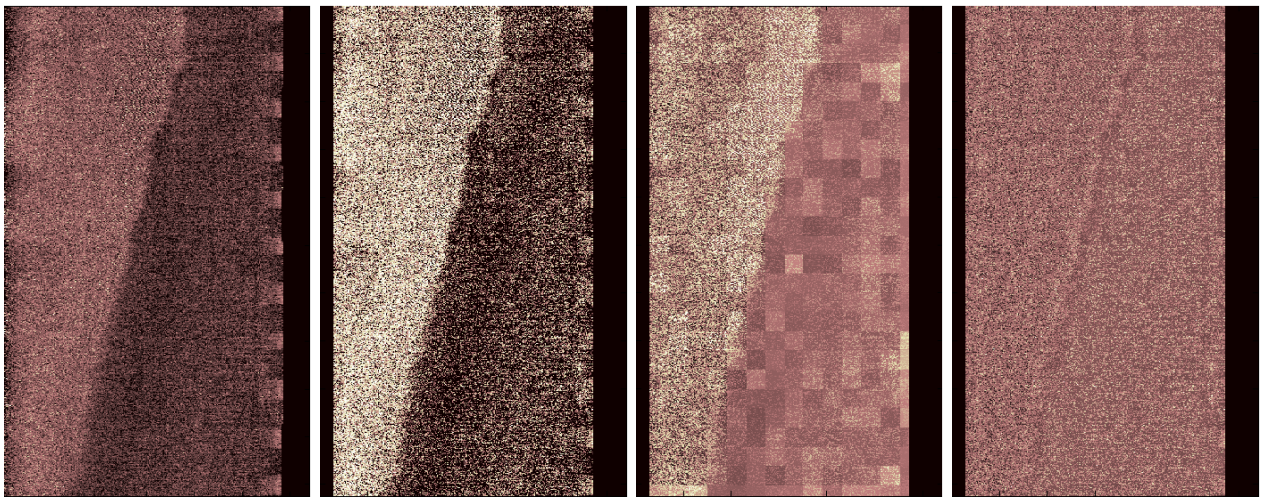

Fig. 2. Different examples of image normalization. Going from left to right: original image; histogram equalization; local histogram equalization; local standardization. Only the local standardization removes the background structure.

moreover the background structure has not been removed.

To address the problem of localized background structure removal, we developed a method based on a local standardization procedure. For each point (i, j) in the image, we compute the mean and standard deviation of the points in a neighborhood $I_d(i, j) = \{(i, j) : -d \leq i, j \leq d - 1\}$. Then, subtract from each point's value the corresponding mean and divide by the standard deviation:

$$x_{ij} \mapsto \frac{x_{ij} - \mu_{ij}}{\sigma_{ij}}, \tag{1}$$

where

$$\begin{aligned} \mu_{ij} &= E_{I_d(i,j)}[x] \\ &= \tfrac{1}{4d^2} \textstyle\sum_{l,m=-d}^{d-1} x(i-l, j-m), \\ \sigma_{ij} &= E_{I_d(i,j)} \left[(x - \mu_{ij})^2 \right] \\ &= \tfrac{1}{4d^2-1} \textstyle\sum_{l,m=-d}^{d-1} \left(x(i-l, j-m) - \mu_{ij} \right)^2 . \end{aligned} \tag{2}$$

The strength of this method is that, although the transform is based on local information, it does not introduce discontinuities in the image. Thus, we can equalize the image locally and get rid of the background structure at the scale we want. We just need to adjust the size d of the neighborhood I_d. The problem with the above formulation is that the algorithm is very slow. The most time consuming part is computing mean and standard deviation for each point in the image. We developed an alternative algorithm that reduces the number of computations giving comparable results. We divide the image using a grid of square cells, compute the mean μ_{ij} and standard deviation σ_{ij} of cell (i, j) and associate those values to the center of the cell (i, j). Then, we interpolate those values to estimate the mean and standard deviation of the other points in the image. The transform (1) can then be applied just like we did before. In this way, mean and standard deviation have to be computed for a much smaller number of points. Interpolation is a very fast algorithm, and preserves the continuity of transform across the image.

B. Denoising

In the wavelet based de-noising we used two different approaches. The first one is the direct application of Donoho's shrinkage [1]. It consists of choosing a certain level in the wavelet representation, which we suspect, contains noise that could affect the detection, and then shrinking its coefficients. We considered two types of mother wavelets: Coiflet-5 and Symmlet-8. It is also possible to shrink at different levels and even shrink with different mother wavelets based on a careful examination of the signal. Following Coifman and Majid [2], we first shrinked the coefficients at a certain level and then shrinked again at a different level the de-noised (reconstructed) image from the first level. Again, we used Coiflet-5 and Symmlet-8 mother wavelets. The scales for shrinkage were chooses so as to fit approximately the mine-like targets dimension. It turned out that a good choice would include levels between the first and the third, the first level corresponding to the finest scale.

Other methods that we have considered are based on more common filters. In particular we used a Gaussian filter with $\sigma = 2$, and a DOG filter (Difference Of Gaussians) with $\sigma_1 = 1$ and $\sigma_2 = 3$. Their parameters have been chosen so as not to smear the difference between the highlight of the mine and its shadow.

To get some intuition about the effect of the de-noising methods, we analyzed their frequency response before and after de-noising. Figures 5 depict the Fourier transform of an original image (left picture, top row), and of the same image de-noised with different de-noising techniques. We note the presence of very high values in the low frequency domain in the original images. A possible interpretation is the presence of regular periodic structures (sand waves on the sea bottom, trails created by fish nets) and a correlation between pixels due to the slow movement of the sonar

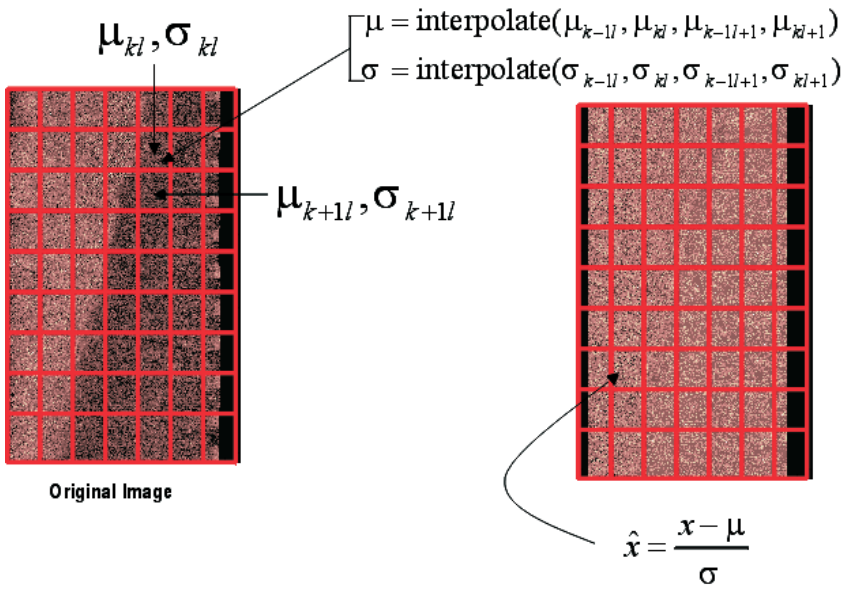

Fig. 3. Local standardization. For each point (i, j) in the image, we compute the mean and standard deviation of the points in a neighborhood. Then, subtract from each point's value the corresponding mean and divide by the standard deviation

detector. The wavelet de-noising had little effect on these low frequencies. The DOG filter (center picture, center row) has a stronger effect as it behaves more like a band-pass and thus, decreases both the high and low frequency response. This behavior is better seen in the histogram of the frequency response (Figure 6) where one can see the distortion to the histogram caused by the DOG and Gaussian (left picture, center row) filters vs. the distortion caused by the wavelet de-noising methods.

To gain better understanding of the effect of the different de-noising methods, we study the histogram of the matched filtered images. Figures 7 show these histograms for the original image (top), and for the same image de-noised with different techniques. The x-axis corresponds to the intensity of the pixel, the y-axis gives the log of the number of pixels having that intensity. The most important part in these histograms is the behavior at high matched filter responses (the far right part). The longer the tail, the higher the response of the matched filter, while the hight of this tail gives an indication to the possible number of false positives.

C. The Matched Filter

The matched filter is designed to detect a mine-like structure, a highlight with a shadow behind it. Relying on the existence of a shadow can dramatically reduce the false positive response of the detector. The challenge to a de-noising method is to preserve this sharp distinction between mine highlight and its shadow, while eliminating high frequency noise. This task is difficult, since a de-noising scheme is generally a low-pass filter that tends to smear edges. For a given false positive response, smearing these edges increases the false negative response, namely the undetected mines. The matched filter mask (Figure 8) contains four distinct

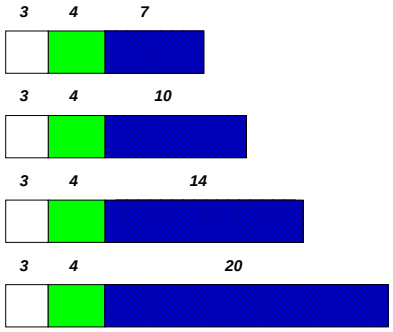

Fig. 8. Target signature map of the match-filter.

regions: pre-target, highlight, dead zone and shadow/post-target. It is defined as:

$$I_m(i,j) = \sum_{k=-N_1}^{N_2} \sum_{l=-M_1}^{M_2} g\left(h(k,l), I_n(i+k, j+l)\right) , \quad (3)$$

where (it is assumed that the input image to the matched filter is normalized so that the average background level is 1.)

$$g\left(h(k,l), I\right) = \begin{cases} h(k,l)(I-1) & \text{shadow, highlight,} \\ & \text{and dead zone regions} \\ h(k,l)|I-1| & \text{pre-target} \\ & \text{and post-target regions.} \end{cases} \quad (4)$$

In each of the four regions, the matched filter coefficients are constant and defined by,

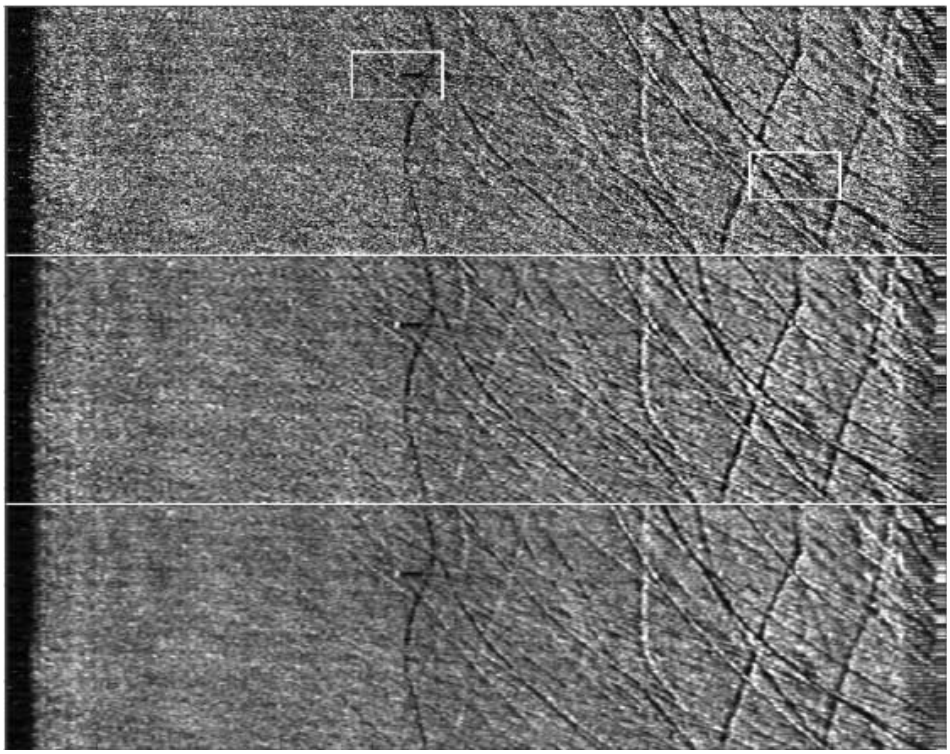

Fig. 4. Original image (top), wavelet de-noised image (center), and Gaussian filtered image (bottom). Mine-like objects in the original image have been enclosed in white squares.

$$h(k,l) = \begin{cases} 1/(S_a(S_o-1)) & \text{shadow region} \\ & \text{or post-target region} \\ 1/(H_a(H_o-1)) & \text{highlight region} \\ 0 & \text{dead zone region} \\ -1/(T_a|T_o-1|) & \text{pre-target} \\ & \text{and post-target regions.} \end{cases} \tag{5}$$

where,

$$\begin{array}{lcl} S_a & = & \text{area of shadow region in square pixels} \\ S_o & = & \text{reference shadow level} \\ H_a & = & \text{area of highlight region in square pixels} \\ H_o & = & \text{reference highlight level} \\ T_a & = & \text{area of pre-target region in square pixels} \\ T_o & = & \text{reference anomalous background} \end{array} \tag{6}$$

Full details of the filter construction are given in [3]. We then normalize the match-filter response by removing its range-dependent mean and dividing by the standard deviation.

D. Clustering and Grouping

After applying a threshold to the post-processed images, we group together first neighbor pixels over that threshold. The threshold varies between the different detectors, and is fixed according to the desired sensitivity of each detector. We then group together clusters that are within a certain distance from one another. We take as a distance threshold the average size of a mine-like target.

IV. Fusion

We identified several different problems connected to mine-like targets detection and classification, and think that it is not reasonable to try to address all of them with a single detection algorithm. The main problem is twofold: reduce the number of false alarms and detect the maximum number of targets. While the former would suggest a strict algorithm, the latter forces us to losen our requirements in order to account for the variability of the targets in the data. We decided to build several algorithms, each of them geared to address a specific problem, and to fuse their results together. The preferred approach for target detection is to analyze in detail the shape and intensity profile around each possible target; however this is computationally demanding. Figure 9 shows the cross-section distribution of intensities after applying the matched-filter. Previous approaches attempted to find a single optimal threshold for making a decision about the presence of a mine-like ob-

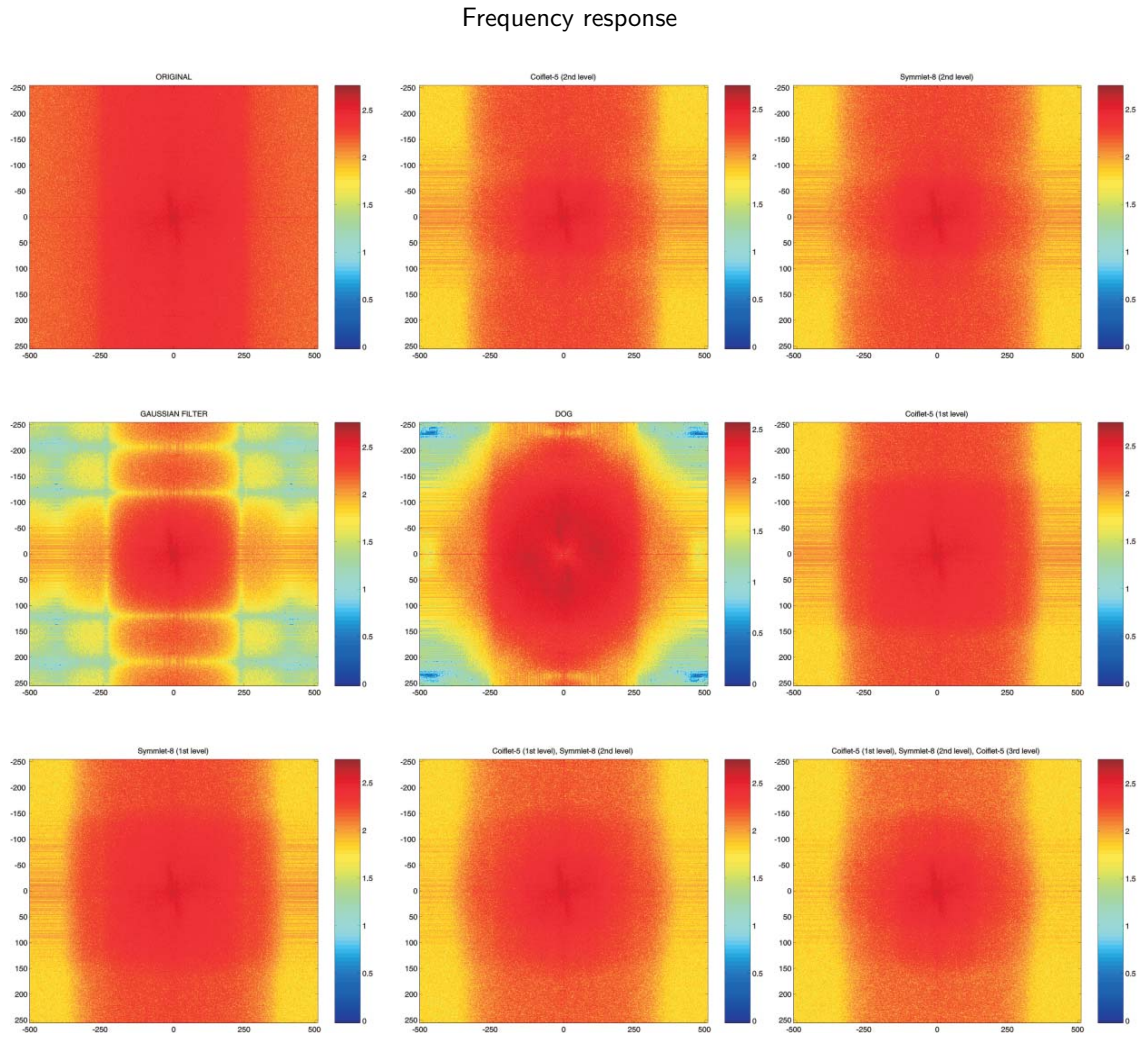

Fig. 5. Fourier transform of the original image, and of the same image de-noised with different de-noising techniques.

ject (left). We approach the problem by attempting to approximate the distribution more accurately using multiple combinations of amplitude and count thresholds, as well as other parameters (right). Figure 10 shows the results of each step in every expert. By trying different experts using different image pre-processing and expert parameters, we construct experts that have independent errors and then by fusing them together, we are able to reduce dramatically the number of false alarms.

V. Results

It appears that wavelet denoising can increase the number of correct detections, keeping the number of false alarms per image reasonably low. The improvement is around 6% which corresponds to the detection of two mine-like targets formerly missed by the detection algorithm. The Gaussian filter could not improve the performance of the detection algorithm. On the contrary, it increased the number of false alarms per image. In table I we do not report the results for the DOG filter. The reason is that the performance of the detection program on the DOG filtered images was too poor, the number of false alarms per image being too large.

Sonar-5 is, so far, the most difficult data-set due to its varied background. This data-set best demonstrates the importance of shrinking only at the level where the important information is, so as no to create artifacts from background clutter in other image scales.

It is quite difficult to infer the quality of the various denoising methods from the denoised images (Figure 4). It is however, evident that wavelet based denoising tends to pop up the highlights of the mine-like targets (Figure 4 center).

The frequency response of all the denoising methods we tested, apart from the DOG filter, is qualitatively the same. All of them act on the image reducing the values of high frequency coefficients. Thus, the difference in their performance is not directly linked to their frequency response but to their ability to retain higher order structure.

The matched filtered histograms show that there is indeed a difference in the way denoising is performed. As can be seen in Figures 7, images denoised using a wavelet based technique present a shorter tail. This means that

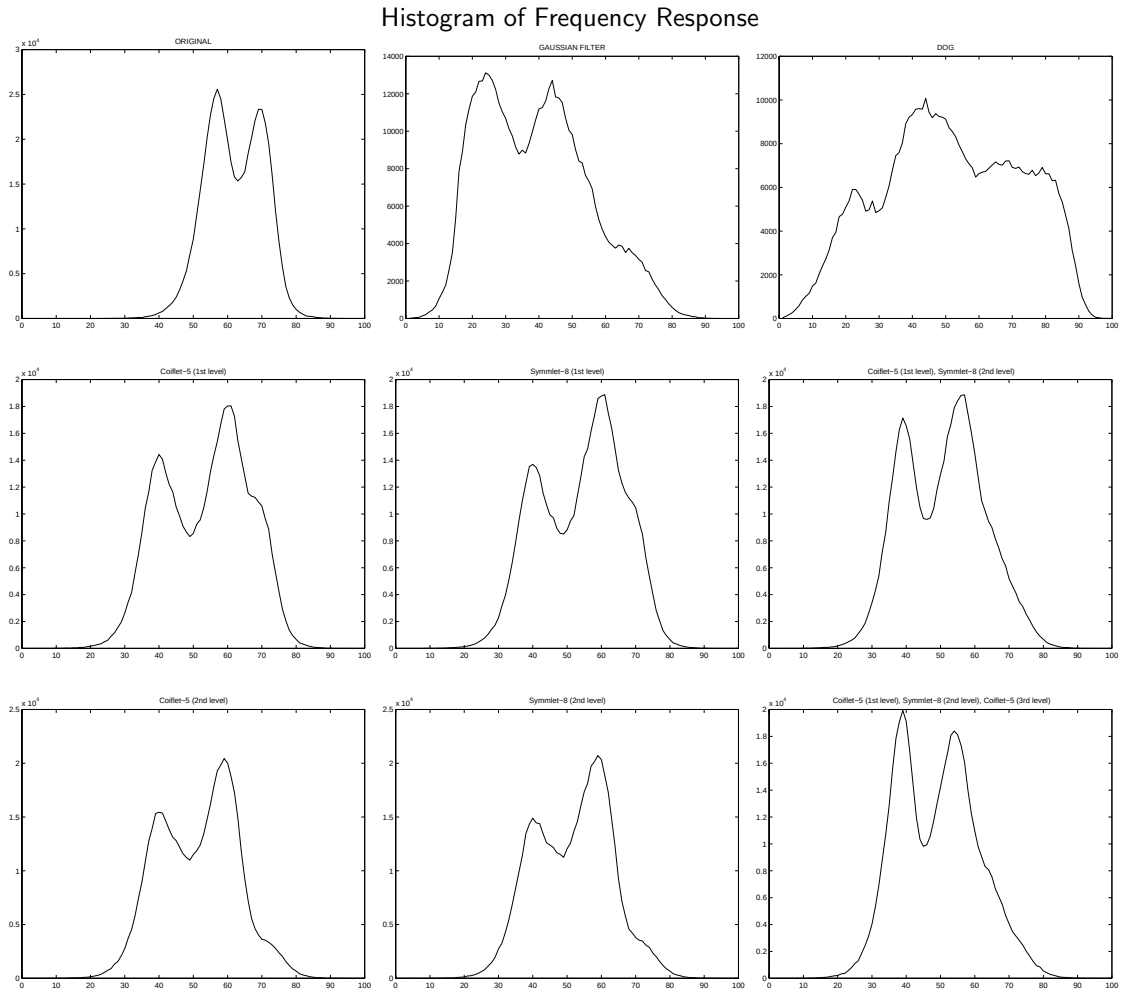

Fig. 6. Histogram of the log of the frequency response of different de-noising methods. It is evident that the DOG and Gaussian filters distort the histogram while the other de-noising methods retain a relatively similar response histogram.

Performance summary: FA/Image

Algorithm	Sonar-0	Sonar-5
Expert #1	10.3	13.8
Expert #2	10.1	13.2
Expert #3	10.7	11.9
Expert #4	10.8	14.8
Expert #5	10.5	14.4
Expert #6	11.1	14.8
Fusion	5.0	7.8

TABLE I

Performance of the different experts. FA/Image are given for the two datasets and correspond to 100% positively detected targets. The bottom row correspond to the fusion of the six experts.

the number of high value pixels in the matched filtered image is lower. Since detections are concentrated in this region, this results in a lower number of false alarms per image. On the other hand, both the Gaussian filtered and the DOG filtered images present a tail comparable to that of the original one.

A further interpretation of the different performance is that, using a convolution filter to denoise the image, we modify the shape of the mine-like targets as well. Thus, the matched filter in the detection program is no longer optimal. On the contrary, wavelet denoising projects the image over an orthonormal basis and shrinks only the coefficients corresponding to wavelets whose support is of the same order of the mine-like targets. This does not affect

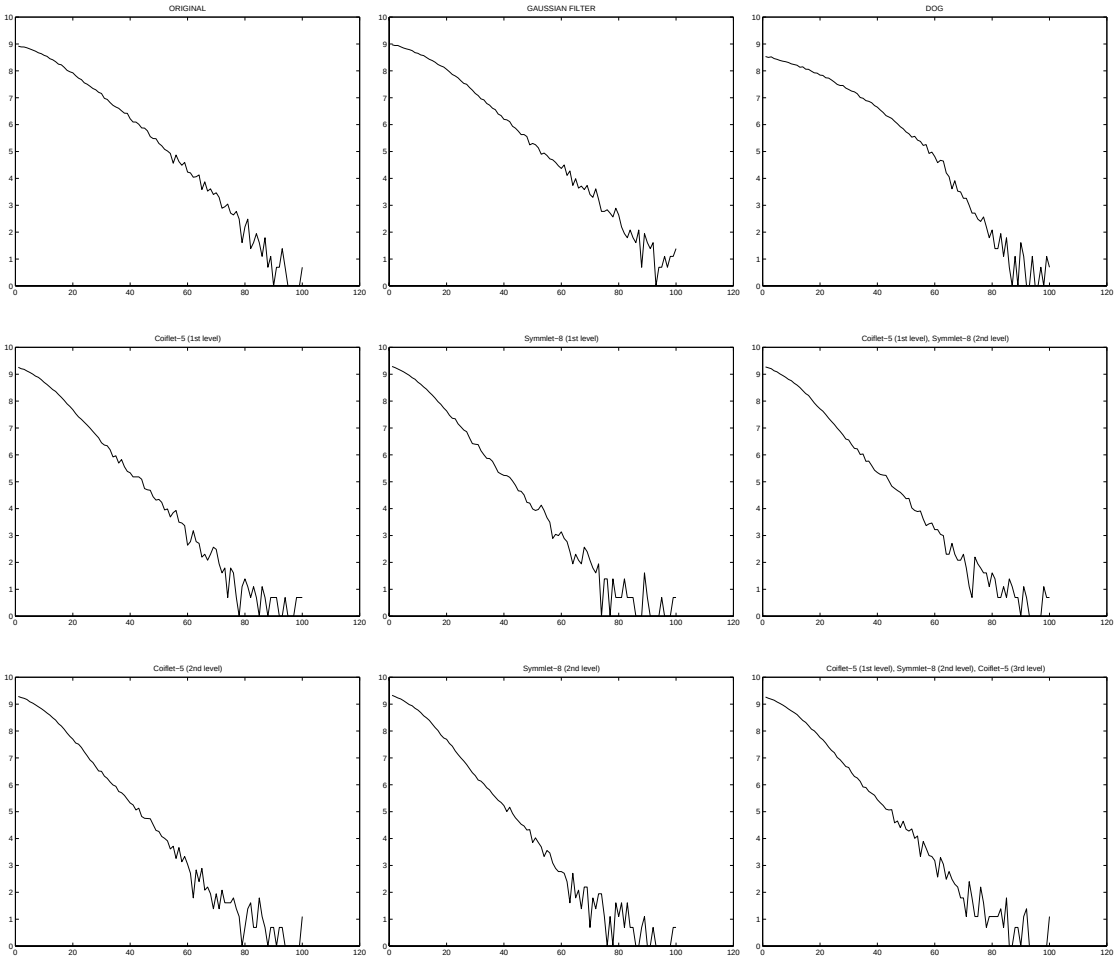

Fig. 7. Histogram of the matched filtered image for the original image (top left), and for the same image de-noised with different techniques. All graphics are in semilogarithmic scale.

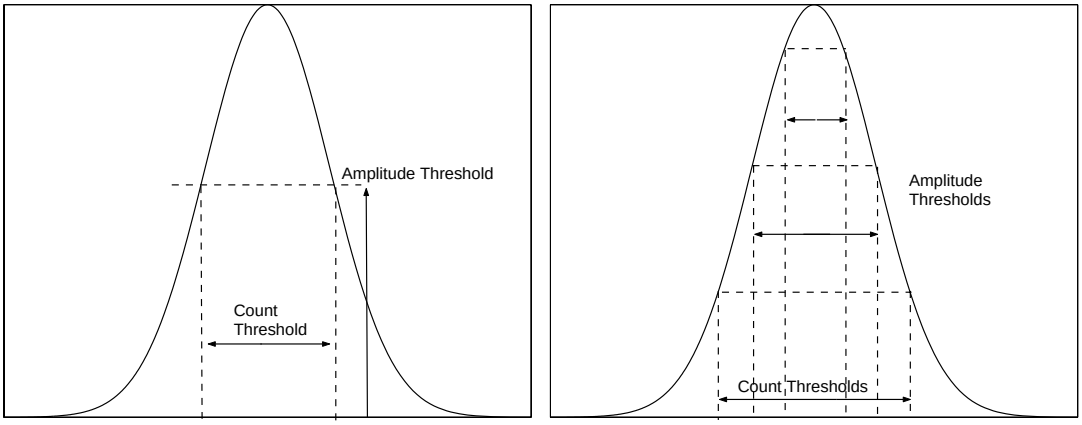

Fig. 9. This figure shows the cross-section distribution of intensities after applying the match-filter. A conventional approach finds a single optimal threshold value and a single optimal pixel count (left). Our approach is to approximate the distribution more accurately using multiple combinations of amplitude and count thresholds (right).

the shape of the targets.

In our analysis we noticed that the performance does not depend on the type of mother wavelet used. Coiflet-5 and Symmlet-8 gave comparable results when the shrinking was applied to the same level. The performance is mostly affected by the choice of the level. That makes sense, since it is equivalent to choosing the scale at which the noise is present. The best results were obtained when the second level was shrinked. So far, we have not seen an effect to the use of different wavelets at different levels.

We have used fusion in the most strict sense, where all experts have to agree about the existence of a mine a in a

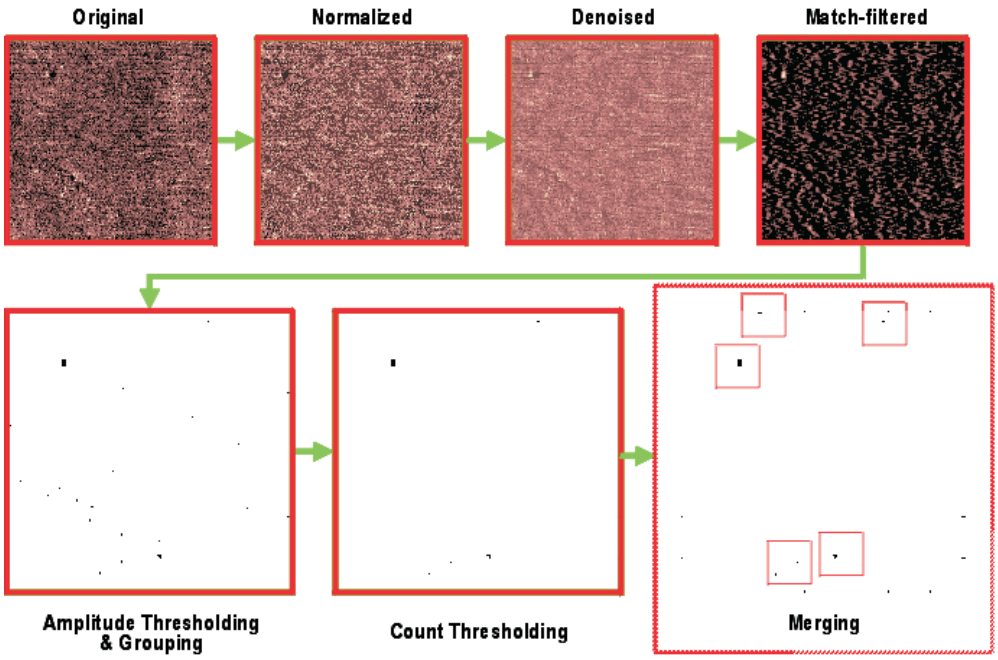

Fig. 10. Steps in each detector.

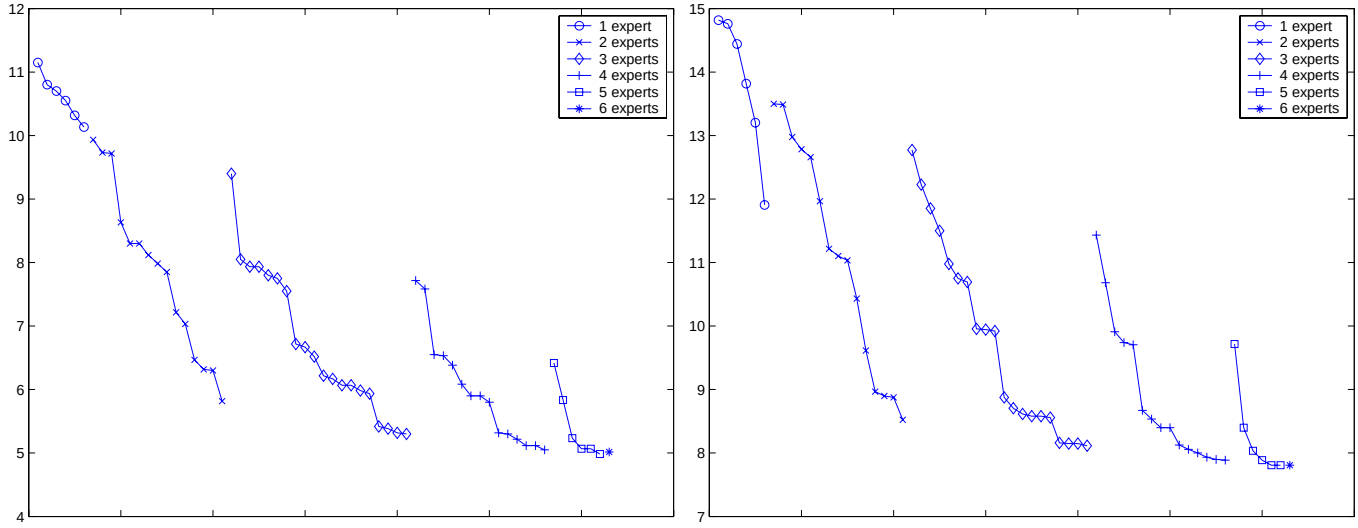

Fig. 11. Performance of different combinations of expert on sonar-0 (left) and sonar-5 (right).

certain location, for the ensemble to vote for a mine there. While we have experimented with various fusion schemes for this problem, we have found this simple one to be most effective in controlling the number of false alarms while achieving the maximal possible sensitivity. When experts are fused in this way, it can be easily seen that the number of false alarms goes down drastically and keeps on going down as long as we add experts (Figure 11). This is of course due to the fact that we train (or modify the parameters) of each of the experts to achieve 100% detection. Since the errors that different experts are relatively independent, due to the different data representations and the different preprocessing of the image normalization as well as image denoising, the fusion is effective in reducing the number of false alarms. Table I shows the number of false alarms per image for the two datasets corresponding to 100% positively detected targets.

This approach demonstrates the usefulness of multiple experts in addressing different background in sonar images as well as mine orientation and contrast. We have not been able to achieve close by results with any other method which does not employ fusion.

VI. Acknowledgements

This work is supported by the Office of Naval Research (code 311).

References

[1] David L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[2] R. Coifman and F. Majid, "Adapted waveform analysis and denoising," in *International Conference Wavelets and Applications*, Toulouse, France, 1992.

[3] G. J. Dobeck and J.C. Hyland, "Sea mine detection and classification using side-looking sonars," *Proceedings of the SPIE Annual International Symposium on Aerospace/Defense Sensing, Simulation and Control*, vol. 2496, 1995.