Automated Detection/Classification of Sea Mines in Sonar Imagery

Gerald J. Dobeck, John C. Hyland and Le'Derick Smedley

Naval Surface Warfare Center, Dahlgren Division
Coastal Systems Station
Panama City, Florida 32407-7001

ABSTRACT

An advanced capability for automated detection and classification of sea mines in sonar imagery has been developed. The Advanced Mine Detection and Classification (AMDAC) algorithm consists of an improved detection density algorithm, a classification feature extractor that uses a stepwise feature selection strategy, a k-nearest neighbor attractor-based neural network (KNN) classifier, and an optimal discriminatory filter classifier (ODFC). The detection stage uses a nonlinear matched filter to identify mine-size regions in the sonar image that closely match a mine's signature. For each detected mine-like region, the feature extractor calculates a large set of candidate classification features. A stepwise feature selection process then determines the subset features that optimizes probability of detection and probability of classification (PdPc) for each of the classifiers while minimizing false alarms.

The AMDAC has been tested using 335 sonar images from three different sonar systems: a synthetic aperture sonar (SAS with 255 images), a side-scan sonar (SSS1 with 60 images) and another side-scan sonar (SSS2 with 60 images). The AMDAC's performance is: 90% PdPc with 0.42 false alarms per image for the SSS1; 92% PdPc with 0.64 false alarms per image for the SAS; and 91% PdPc with 0.12 false alarms per image for the SSS2. For these data bases the algorithm's performance is as good or slightly better than that of an expert sonar operator.

This work was funded by the Office of Naval Research, ONR 321TS , through the 6.2 Mine Countermeasures program element.

1. INTRODUCTION

This paper builds on our earlier work [1] and describes a new algorithm for the automated detection and classification (D/C) of sea mines in sonar imagery. Three sonar image data bases are used to evaluate the new D/C algorithm. The algorithm is referred to as the Advanced Mine Detection and Classification (AMDAC) algorithm. It is comprised of four stages: (1) Image Enhancement, (2) Detection, (3) Feature Extraction and Optimal Feature Selection, and (4) Classification (see Figure 1).

For the last seven years CSS has used three side-looking sonar data bases to evaluate mine detection and classification (D/C) algorithms developed in house, by industry and by academia [2], [3], [4], [5]. The data bases consist of a 255 image set from a synthetic aperture sonar (SAS), a 60 image set from a side-scan sonar (SSS1), and a 60 image set from another side-scan sonar (SSS2).

These data bases were selected because they demonstrate a variety of problematic issues and signal processing challenges. Mine threats in these data bases are bottom mines; they provide a significantly greater challenge to detect and classify than volume mines. Results for probability of detection and classification (PdPc) and false alarm rates referred to in this report are for single pass (one look) sonar operation. Single pass operation is of interest because of the implied higher search rate. Performance of an expert sonar operator was available for two of these databases, thus providing a benchmark for automated algorithm performance. Because these three data bases are somewhat small and do not represent all interesting scenarios, one must be cautious in generalizing any conclusions. However, the prohibitive cost of collecting and processing large data bases makes utilization of small data bases the typical practice when evaluating and comparing automated D/C algorithms. 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.

SPIE Vol. 3079 • 0277-786X/97/$10.00

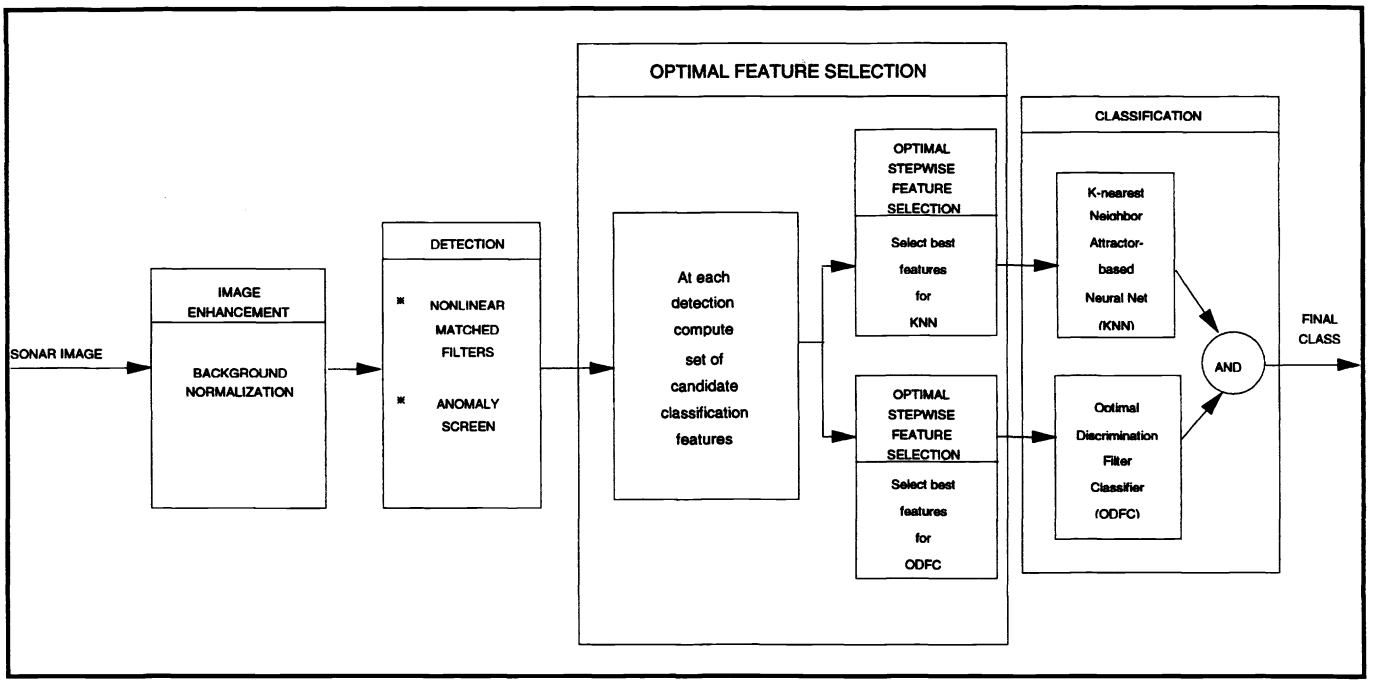

Figure 1. Overview of AMDAC Algorithm

The following sections, 2.1 through 2.6, describe the four stages of the AMDAC.

2. AMDAC ALGORITHM

2.1 Image Enhancement

The purpose of the image enhancement stage is to pre-condition the image so that the subsequent detection and classification stages are robust to variations in background level. This is accomplished by normalizing the background throughout the image to a constant level so that highlight and shadow levels are consistent and clearly stand out.

The principal reasons for variations of bottom brightness are (1) inadequate range-varying gains that are applied to the sonar return to compensate for the range-dependent reduction in signal strength cause by spherical spreading of acoustic energy, (2) the strong sonar returns from highly reflective bottom regions (e.g., gravel), and (3) the weak sonar returns from highly absorbent regions (e.g., mud). Thus, pre-conditioning reduces some of the background invariance and reinstates some degree of robustness to bottom variation.

For the three sonar data bases studied in this article the inadequate range-varying gain problem, as described above, was the dominate factor; so a simple range normalization was found to be adequate. This was done as follows.

$$I_r(i,j) = \text{raw 8-bit sonar image } (0 \le I_r(i,j) \le 255)$$
$$i = \text{range index} \quad (1 \le i \le n)$$
$$j = \text{cross-range index} \quad (1 \le j \le m)$$

$$I_n(i,j) = \text{normalized image } (0 \le I_n(i,j) \le 4)$$

$$= \text{MIN}\,(4, I_r(i,j)/b(i))$$

where

$$b(i) = \text{mean level of } I_r \text{ at range } i$$

$$= \frac{1}{m} \sum_{j=1}^{m} I_n(i,j)$$

Note the following facts about I_n:

(1) The mean background pixel intensity is essentially unity.
(2) The dynamic range for highlight pixel intensity is from 1 to 4.
(3) The dynamic range for shadow pixel intensity is from 1 to 0.

Experience of expert side-looking sonar operators has shown that the above scaling procedure is a good way to distribute the dynamic range between highlight and shadow.

In subsequent discussions we will make use of these three facts in setting parameter values. As an aside, it has been found that images from other environments may contain many large irregular regions, some of which are highly reflective while others are highly absorbent. In these cases the simple range normalization does not achieve the goal of making the background level uniformly near unity. Other computationally more intensive methods are needed that normalize the image based on local statistics within the image. As an example, low-pass median filtering has performed quite well for us.

2.2 Detection Stage

The purpose of the detection stage is to scan the entire image and identify candidate mine-like regions that will be more thoroughly analyzed by the subsequent classification stages. This is the most computationally intensive stage because a mine-size region surrounding each image pixel must be evaluated. Therefore, the goal is to keep the computations involved with each mine-size region small. The purpose of this computation is primarily not to determine mine-likeness but rather to screen out the non mine-like regions in the vast majority of the image and thus reduce the amount of data that must be processed by the computationally more intense classification stages. If the detection stage can reduce a typical image to about 10 detections (10 being very generous), the computational requirements of the classification stages are insignificant when compared to the detection screening.

The detection stage of the AMDAC is an improvement on the one reported in our earlier paper [1]. The detection stage operates on the normalized image. The algorithm divides the image into regions along range to account for the variability of the background and mine signature as a function of range. For this study the image was divided into three regions. Next, the image is convolved with a nonlinear matched filter, a different matched filter for each of the range regions. The procedure sets negative values from the match filter to zero. For each range region the resulting matched-filtered image is normalized by removing the region's mean and dividing by the region's standard deviation. Next, the procedure scans a target-size window over the normalized matched-filter image. A detection is declared by counting the total number of pixels within the window that exceed a specified "amplitude-detection threshold"; a pixel count above the specified "detection-count threshold" is declared a target.

2.3 Nonlinear Matched Filter

The nonlinear matched filter is the work horse of the detection stage [6]. The matched-filter mask contains four distinct regions: pre-target, highlight, dead zone and shadow/post-target ; see Figure 2. It is defined as,

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

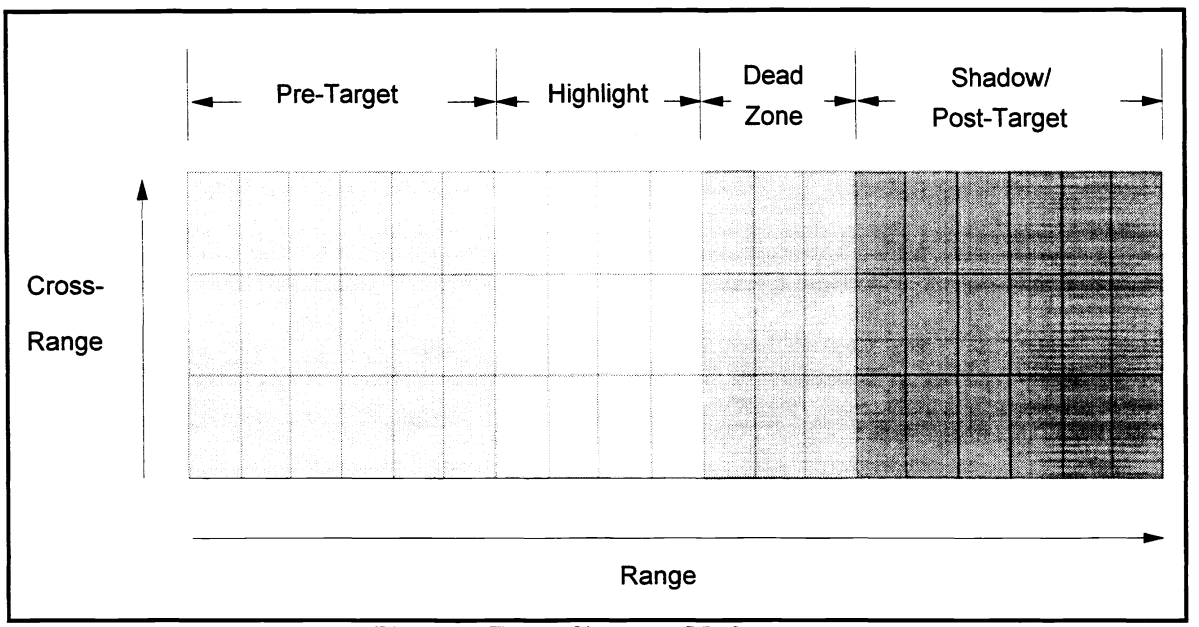

Figure 2. Target Signature Mask

where

$$g(h(k,l), I) = \begin{cases} h(k,l)\,(I-1) & \text{for } h(k,l) \text{ corresponding to shadow, highlight, and dead zone regions} \\ h(k,l)\,|I-1| & \text{for } h(k,l) \text{ corresponding to pre-target region and post-target regions} \end{cases}$$

Note that the nominal background level of unity is subtracted from I.

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

shadow region or post-target region: $h(k,l) = 1/(S_a(S_0 - 1))$
highlight region: $= 1/(H_a(H_0 - 1))$
dead zone region: $= 0$
pre-target and post-target regions: $= -1/(T_a|T_0 - 1|)$

where

S_a = area of shadow region in square pixels (see Tables 1, 2 and 3)

S_0 = reference shadow level
= 0.75 for this study

H_a = area of highlight region in square pixels (see Tables 1, 2 and 3)

H_0 = reference highlight level
= 1.5 for this study

T_a = area of pre-target region in square pixels (see Tables 1, 2 and 3)

T_0 = reference anomalous background level

The following two anomalous levels used in this study result in the same h(k,l) value.

$$T_0 = \begin{cases} 2.0, \text{ anomalous high backgound level} \\ 0.0, \text{ an anomalous low background level} \end{cases}$$

In this study the reference center h(0,0) corresponds to the center pixel of the highlight region. The following is a brief discussion of this matched filter.

The nonlinearity is associated only with the pre-target and post-target regions. The intent of this part of the matched filter was to prevent the detection of objects with highlights or shadows that were greater than mine size. The filter coefficient for this region is negative; thus, the output of the filter is decreased if the pre-target or post-target regions are dominated by low or high valued pixels. This region could have been selected to encircle the typical mine-size area. However, the large variations in the mine's cross-range size and shadow length, together with considerations of the extra computational burden, led us to use only the area in front of the highlight region. Since the mines in our SAS image data base have no shadows, a post-target region was used in place of the shadow region.

The selection of the coefficients, defined by the reference levels and the region's area, was done to make the contribution from each region of the matched filter equal for typical shadow, highlight, and anomalous background levels. Specifically, the coefficients are constructed such that the contribution for each region is unity when evaluated at the reference level of that region.

The optimal detection-amplitude and detection-count threshold pair is selected from the optimal receiver operator characteristics (ROC) table [6]. The ROC table contains optimal threshold pair that gives the least false alarm rate for a given Pd. A computationally fast method was developed to generate the ROC table that requires only a single pass through the data base.

The AMDAC uses a different filter mask for each mine type and each range region. For example, the SAS sonar has one mask for detecting mine type "A" in the first region, another mask for detecting mine type "A" in the second region, and another mask for detecting mine type "A" in the third region. To detect a different mine type, three different masks would be used. Reference [1] discusses target mask design in detail. Tables 1 through 3 show the masks' sizes for the three sonars and the three range regions.

Table 1. Target Mask Size for SSS1

Image Region	Mask Region			
	Pre-Target	Highlight	Dead Zone	Post-Target
First Region	12x3	5x3	4x3	10x3
Second Region	12x3	5x3	4x3	14x3
Third Region	12x3	5x3	4x3	20x3

Table 2. Target Mask Size for SAS				
Image Region	Mask Region			
	Pre-Target	Highlight	Dead Zone	Post-Target
First Region	12x3	9x3	0x3	12x3
Second Region	12x3	9x3	0x3	12x3
Third Region	12x3	9x3	0x3	12x3

Table 3. Target Mask Size for SSS2				
Range Region	Mask Region			
	Pre-Target	Highlight	Dead Zone	Shadow
First Region	10x3	3x3	3x3	5x3
Second Region	10x3	3x3	3x3	11x3
Third Region	10x3	3x3	3x3	11x3

2.4 Feature Extraction

In the feature extraction stage, a large set of classification features is computed from a target-size classification window centered at each detection location in both the normalized raw sonar image and the normalized matched-filtered image. The procedure calculates forty-five candidate features based on the size, shape and strength of the highlight and shadow, and histogram information of pixel intensity within the classification window. A complete list of features follows:

1. W_pix:	number of normalized image pixels in the window that exceed a threshold*
2. Mfw_pix:	number of matched-filter pixels in the window that exceed a threshold
3. M_wpix:	maximum normalized image pixel intensity in the window
4. Mmf_wpix:	maximum matched-filter pixel intensity in the window
5. T_str:	target strength computed from the normalized image image
6. Mft_str:	target strength computed from the matched-filter image
7. Mx_eig:	length of major axis of an ellipse fit to highlight region of the normalized image
8. Mn_eig:	width of minor axis of an ellipse fit to highlight region of the normalized image
9. Mfmx_eig:	length of major axis of an ellipse fit to bright region of the matched-filter image
10. Mfmn_eig:	width of minor axis of an ellipse fit to bright region of the matched-filter image
11. S_len:	shadow length
12. S_str:	shadow strength
13. Mmf_pclu:	maximum matched-filter intensity over the pixels in the detection cluster
14. C_pix:	number of pixels in the detection cluster
15. C_im:	number of detected clusters in the image
16. - 25. Hist_org(i):	number of normalized pixels above threshold(i) in the window, i from 1 to 10
26. - 35. Hist_mf(i):	number of matched-filter pixels above threshold(i) in the window
36. - 45. Hist_diff(i):	number of matched-filter pixels above threshold(i) in the classification window minus the number of matched-filter pixels above threshold(i) in the region that locally surrounds the classification window

* All thresholds are selected by the designer.

2.5 Optimal Feature Selection

When training classifiers with finite databases, Bellman discovered that the robustness of a classifier will collapse when the number of features becomes too large [7]. This is known as "the curse of dimensionality." Therefore, an important part of the overall classifier design is to select a small and robust set of classification features from the larger candidate set. Evaluating all possible combinations of 45 features is computationally not feasible (there are 2^{45} - 1 combinations, a number greater than 10^{12}). Therefore, a stepwise optimal selection process is used (select the best single feature, next select best of the remaining 44 features to add to the first, next select to best of the remaining 43 features to add to the previous two, etc.) [8]. For a 45 feature set this involves evaluating 1035 feature combinations. Computationally efficient algorithms have been developed for this selection process.

The two classifiers used in this algorithm were the k-nearest neighbor neural network (KNN) and the optimal discriminatory filter classifier (ODFC). Since two classifiers are used in this detection/classification algorithm, there are two selection algorithms; each one is tuned to the respective classifier. This is required because a set of features that is optimal for one classifier will, in general, be far from optimal for the other classifier. This usually results when each classifier uses a different mathematical/geometrical structure to partition feature space into class regions. A set of features amenable to one partitioning scheme may not separate well for a different partitioning scheme.

Briefly, two feature selection algorithms were developed for the KNN and ODFC, respectively. By optimal is met the following. Bellman's curse of dimensionality implies that, for a finite training set, the detection and classification performance (as judged by the validation/test set) does not continually improve as more features are added (as dimensionality is increased) to the classification process. With a finite data set it is possible only to determine a subset of features that will give robust performance. Optimal selection in this context means finding the subset of features from a much larger set that gives the best performance for both the finite training and validation (test) sets for a specific classifier (in our case, either KNN or ODFC). Specifically, the best subset is the subset of features which when fed into the classifier minimizes a risk metric. The risk metric that is used is defined as the maximum (between the training and validation data sets) of the weighted sum of the number of missed mines and number of false calls. As stated above, for a typical candidate feature set of size 45, there would be 2^{45} - 1 combinations of subsets to evaluate in order to determine a global optimum. Because this is too large of a number to evaluate, a Forward Stepwise Optimal Selection Process (FSOSP) is used instead of a Globally Optimum Selection process. The FSOSP selects as the first feature the one that gives the best classifier performance of any single feature. The second feature selected is the one that performs best when used with the first. As it name implies, the Forward Optimal Stepwise Selection Process continues at each stage by adding a new feature to the subset determined in the previous stage. This stepwise process is obviously not globally optimal, but has worked extremely well in this optimization application. Our research suggests that it is important that the classifier itself be used in the selection process. Other selection techniques (e.g., principal component analysis or multivariate normal models) are often used because they are computationally tractable. But they are usually not based on the same mathematical, statistical, or geometrical structure as the classifier and, therefore, do not select the best features for that classifier. Keeping in mind Bellman's curse of dimensionality, our research suggests that classifiers that train fast, for which the best subset of features can be optimally selected, will perform significantly better than more sophisticated classifiers for which the best subsets cannot be found because training is too computationally expensive.

The selection process has been further improved by adding a Backward Stepwise Optimal Selection Process (BSOSP) analogous to model order reduction used in statistics [9]. Because the candidate features are not orthogonal, adding a new feature to the current optimal feature subset can in fact decrease the performance metric; this is consistent with Bellman's curse of dimensionality. The augmented procedure is to first use the FSOSP to determine the initial feature set as previously described. Then, the BSOSP determines which features to remove. Application of the FSOSP and BSOSP can be repeated until performance improvement stops. As an example, using the SSS2 data base, the number of false alarms per image decreased from 0.28 (using only FSOSP) to 0.12 (using FSOSP followed by BSOSP) for the same PdPc of 91%.

2.6 Classification

The KNN and ODFC classifiers were chosen because they process the data using significantly different mathematical, statistical, and geometrical paradigms; "ANDING" the two classifiers produces highly effective results. In general, many classifiers could be "ANDED" but for the sonar data presented here, two have worked very well. Also the process of classifier selection relied strongly on the fact that the optimal feature selection process employed would be "tuned" to each classifier. With this in mind the two classifiers were chosen because: (a) they train very fast - this fact makes the algorithm used in the stepwise optimal selection process feasible, (b) they each divide feature space into different geometrical models - the KNN into hyperspheres and the ODFC into hyperplanes -- this tends to make the classifiers complementary,(c) the classifiers use different "noise" tolerant paradigms. The KNN uses sample Bayesian conditional probabilities as determined from the training data. The ODFC uses noise rejection and signal enhancement schemes based on concepts of matched-filter design.

The KNN is a probabilistic-based neural network that employs radial-basis neurons [10]. The neural network has two layers: the attractor layer and the confidence layer. The attractor layer is constructed during training in the following way. Feature space is partitioned by a set of attractors. A feature vector, f, is said to belong to attractor i if $N(f, f_i) < R_i$ where $N(f, f_i)$ is some distance measure between vectors f and f_i (for this study an L_2 norm was used), f_i is the center of attractor i and R_i is the radius of attractor i.

For a given $N(.,.)$, R_i, and f_i the probability that attractor i contains a given class can be estimated from sample probabilities computed from the training set. Each attractor is represented as a radial-basis neuron that has the feature vector as an input and multiple outputs (one output for each class). The i-th neuron *fires* when the input feature vector belongs to attractor i; when it fires, the level of each output, associated with a given class, equals the estimated probability that the attractor contains that class.

During training, the number of attractors, their center locations (f_i), radii (R_i), and degree of overlap are determined in an optimal fashion to encompass the entire training set. The radii are selected using the well-known k-nearest neighbor scheme to achieve accurate probability estimates and adequate class boundary resolution for good class discrimination. The more training samples in an attractor, the more accurate the sample probability estimates will be. In a relative sense, an attractor with a large radius will encompass a large number of training samples. However, the size of the radii is balanced against the resolution into which feature space is partitioned; the smaller the radii, the better the discriminatory resolution at complex class boundaries. Finally, attractor overlap is used to regulate the *smoothness* of probability estimates across class boundaries.

Training of this neural network also takes into account that the training data is usually statistically unfairly sampled; i.e., the number of training samples in a class (or subclass) does not reflect the true probability of occurrence of that class but instead reflects the artificial conditions over which the data was collected. For example, one might have more of one class than another, or more data collected under one condition than another. In reality, fairly sampled data is seldom available because the cost to collect an adequate data base is unacceptably high or simply because one does not have adequate access to another country's targets or environment in which to collect representative data. Under such uncertainty, it was determined that the best statistical property to reinstate into the training set is that of equally likely classes (or subclasses). This is accomplished by a specially devised counting method that is used to determine the sample probabilities. Thus, the training of this neural network is not biased to favor those classes that were most plentiful in the training and test sets and to ignore those classes that were least abundant. Other neural network training algorithms that optimize the number of correct classifications are typically biased by such data.

The output of the attractor layer is then fed to the confidence layer which determines the confidence that the input feature vector belongs to each class. First all the attractors that contain the input feature vector (or are sufficiently near it) are identified. The confidence that the feature vector belongs to a given class is then determined by interpolating the individual probability estimates that the feature vector belongs to a class over this set of attractors. In the interpolation, the contribution of each attractor's probability estimate is weighed according to how close the feature vector is to the attractor's center.

This neural network has the desirable characteristic that it can determine if a feature vector cannot be reliably classified. Specifically, if there are no attractors sufficiently near, then the feature vector is considered statistically different than the training vectors and is said to belong to an unknown class. Another way to view this is to say that during training, the neural network never experienced anything that was representative of this new vector (i.e., the new vector is not within the neural network's experience base). Under this circumstance, there is no basis to select among the known classes and all outputs of the neural network are set to zero. Other noteworthy characteristics of this network are that the training algorithm is non-iterative, fast and the network sizes are reasonable thereby permitting very quick execution so specialized hardware is not needed.

The ODFC is a classifier with its basis in linear discrimination theory [10]. Two banks of linear filters are determined from the training set: one bank is sensitive to mine characteristics and the other is sensitive to clutter characteristics. The following is a brief description on how the discrimination process works.

Let f be a feature vector. Let $F_m(f,i)$ be the energy output for input f of the i-th filter in the mine-filter bank, and $F_c(f,j)$ be the energy output for input f of the j-th filter in the clutter-filter bank. F_m and F_c are linear filters; the outputs of these filters are linear combinations of the input features plus a DC offset and a warping term made up of a linear combination of nonlinear functions of the features. The filter's coefficients, DC offset, and warping coefficients are determined such that, on the average, the energy output of the mine-filter bank is greater than the energy output of the clutter-filter bank when the input feature vector corresponds to a mine. And vice versa for a feature vector that corresponds to clutter. That is, on the average,

$$\max_i F_m(f(\text{mine}),i) > \max_j F_c(f(\text{mine}),j)$$

$$\max_i F_m(f(\text{clutter}),i) < \max_j F_c(f(\text{clutter}),j)$$

The solution involves the solving of a generalized eigenvalue problem. In the ODCF linear combinations of features permit boundaries between classes to be described by sets of hyperplanes. By adding a linear combination of nonlinear functions of the features one changes the hyperplane surface to one that is curved (warped). These nonlinear terms permit the partitioning surfaces in feature space to more efficiently fit irregular class boundaries.

Classification is determined by "ANDING" the output of both classifiers. Both the KNN and ODFC have two outputs corresponding to the confidence that the input feature vector is associated with a mine or clutter. Let $C_{knn}(\text{mine})$ and $C_{knn}(\text{clutter})$ designate these two confident levels for the KNN; and $C_{odfc}(\text{mine})$ and $C_{odfc}(\text{clutter})$ for the ODFC. If

$$T_{knn}\, C_{knn}(\text{mine}) > C_{knn}(\text{clutter})$$
$$\text{AND}$$
$$T_{odfc}\, C_{odfc}(\text{mine}) > C_{odfc}(\text{clutter})$$

a mine classification is declared. The classification gain thresholds, T_{knn} and T_{odfc}, are adjusted to select the desired balance between PdPc and false alarms.

3. RESULTS

3.1 SSS2 Example 1

In Figures 3 through 10, Image 0 and Image 51 from the 60 image SSS2 data base are used to illustrate the D/C algorithm. Figure 3 shows the normalized image and the results of the detection stage. The gray scale indicator on the extreme right of the image indicates pixel intensity. White indicates a pixel intensity of 4.0 and represents the maximum pixel intensity value displayed. Black indicates a pixel intensity of 0.0 and represents the minimum pixel intensity value displayed. To give a size perspective, each sonar image has a two-meter by two-

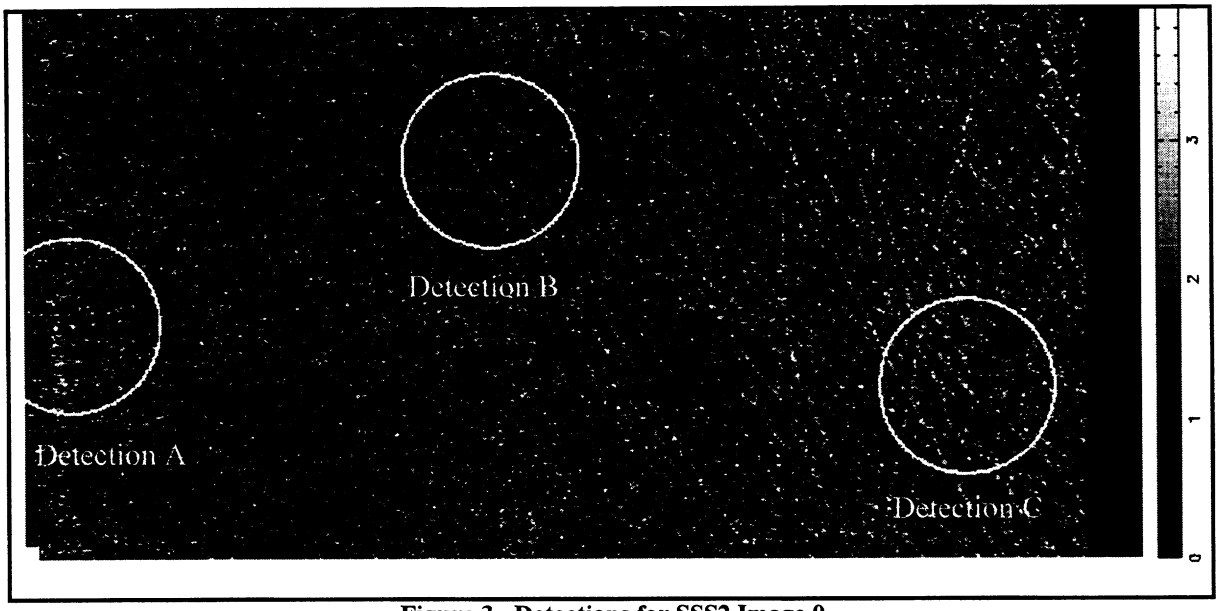

Figure 3. Detections for SSS2 Image 0

meter square highlight in the lower left-hand corner of the image. In all images, range from the sonar increases from left to right implying target shadows are to the right of the highlights. The detected mine-like targets typically resemble the target-signature mask--a highlight region followed by a shadow and preceded by a uniform background. With a strong highlight and pronounced shadow, Detection B is easily visible. Although Detection C has a noticeable highlight followed by a shadow, the shadow's proximity to a trough made by a fisherman's drag net obscures the shadow region. False Detection A is triggered by a subtle highlight and shadow.

Figures 4 and 5 show that both the KNN and the ODFC classify Detections B and C as mine-like; they both reject Detection A. "ANDING" the two classifiers results in Final Classifications B and C as illustrated in Figure 6. The squares in Figure 6 indicate actual mines; there one sees that both mines are detected and classified and one false target is detected but not classified as mine-like by either classifier.

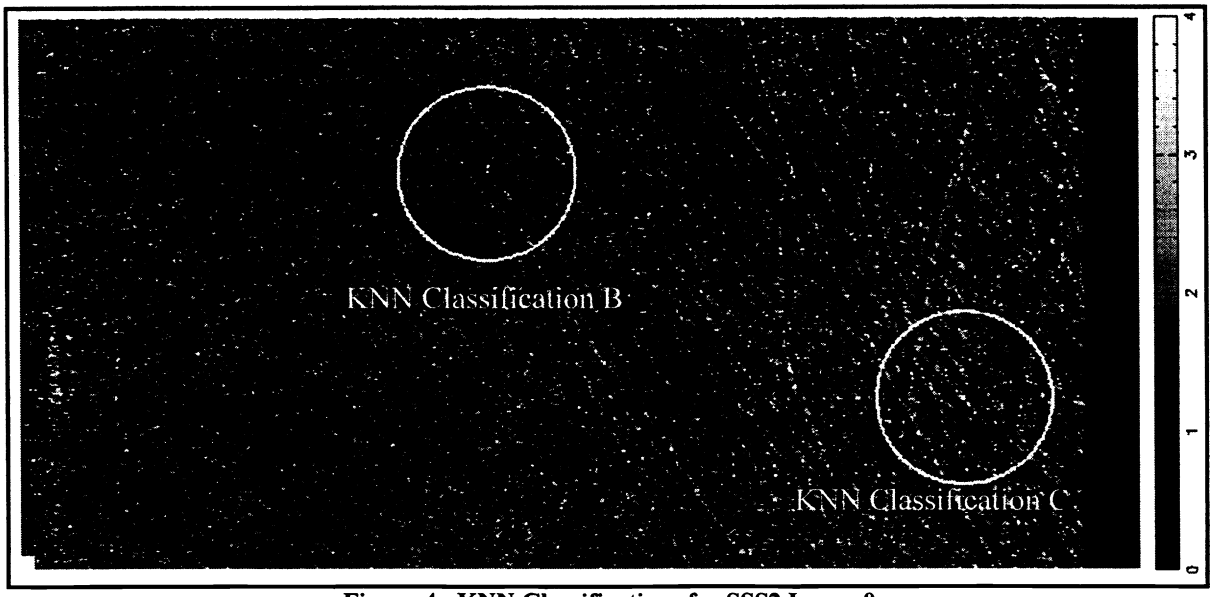

Figure 4. KNN Classifications for SSS2 Image 0

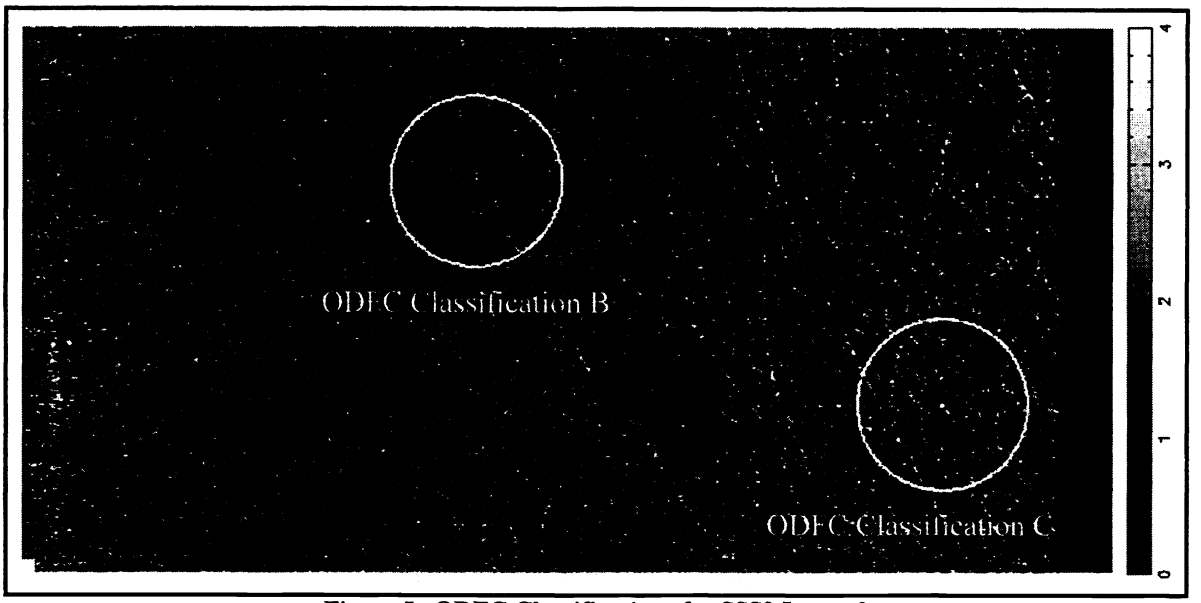

Figure 5. ODFC Classifications for SSS2 Image 0

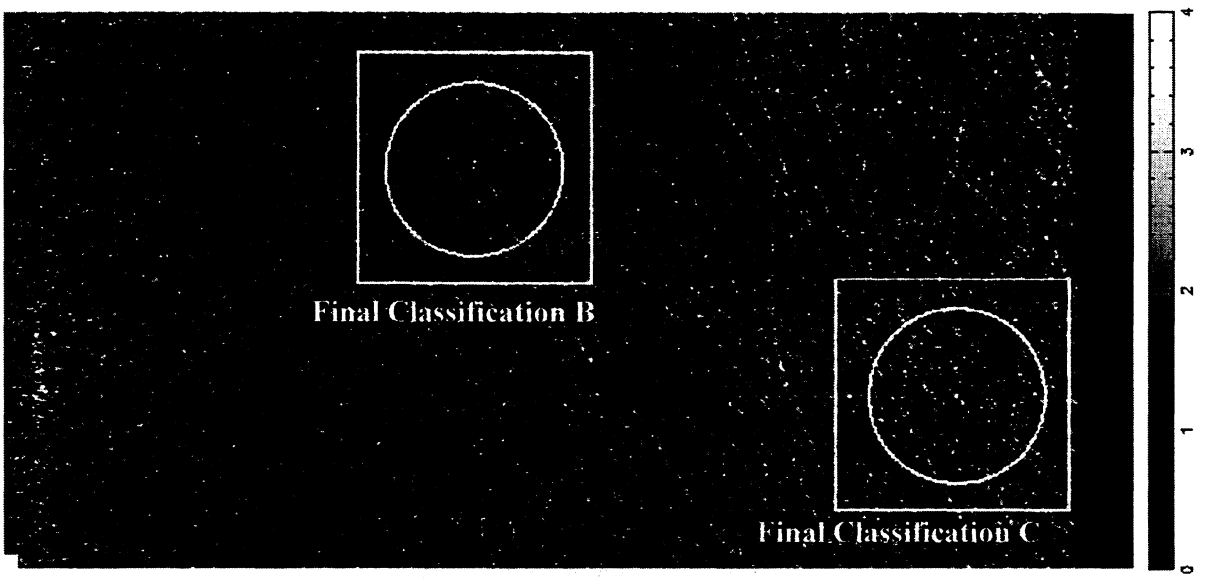

Figure 6. KNN and ODFC Classifications for SSS2 Image 0

3.2 SSS2 Example 2

Figure 7 shows another SSS2 example. This particular image illustrates some of the difficulties associated with normalizing the background in the sonar image. Note that the background level of the image is not uniform. And there is a large band of bright pixels that extends from the upper left to the lower right of the image caused by a highly reflective bottom type. Despite the lack of uniformity in the background intensity, the AMDAC detects four candidate mines.

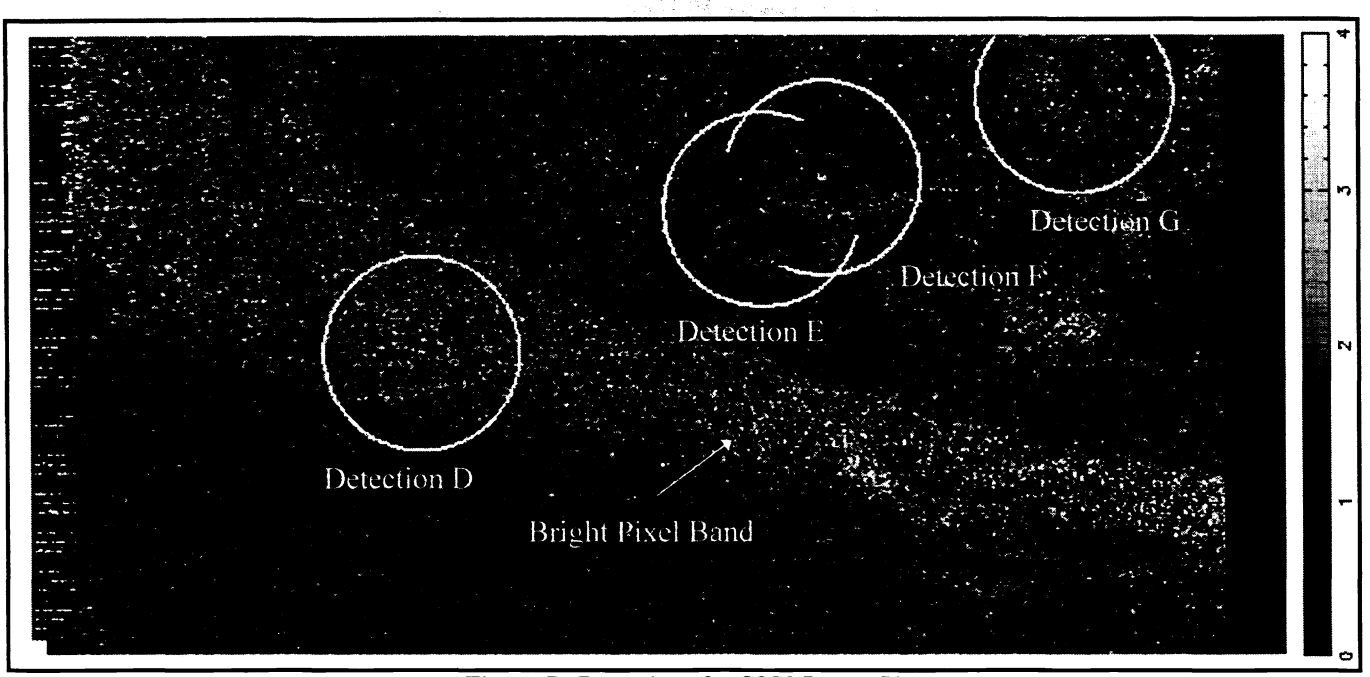

Figure 7. Detections for SSS2 Image 51

In Figure 8, the KNN classifies only Detections D and E as mines and rejects Detections C and F. Figure 9 shows that the ODFC classifies Detections E and F as mines and rejects Detections C and D. Results from "ANDING" the two classifiers are shown in Figure 10. As the ground truth square indicates, the final mine classification is correct and all false alarms have been eliminated. This example clearly illustrates the power of combining two different classification approaches; separately, each approach would have had one correct classification and one false alarm. "ANDING" the two classifiers eliminated the false alarms.

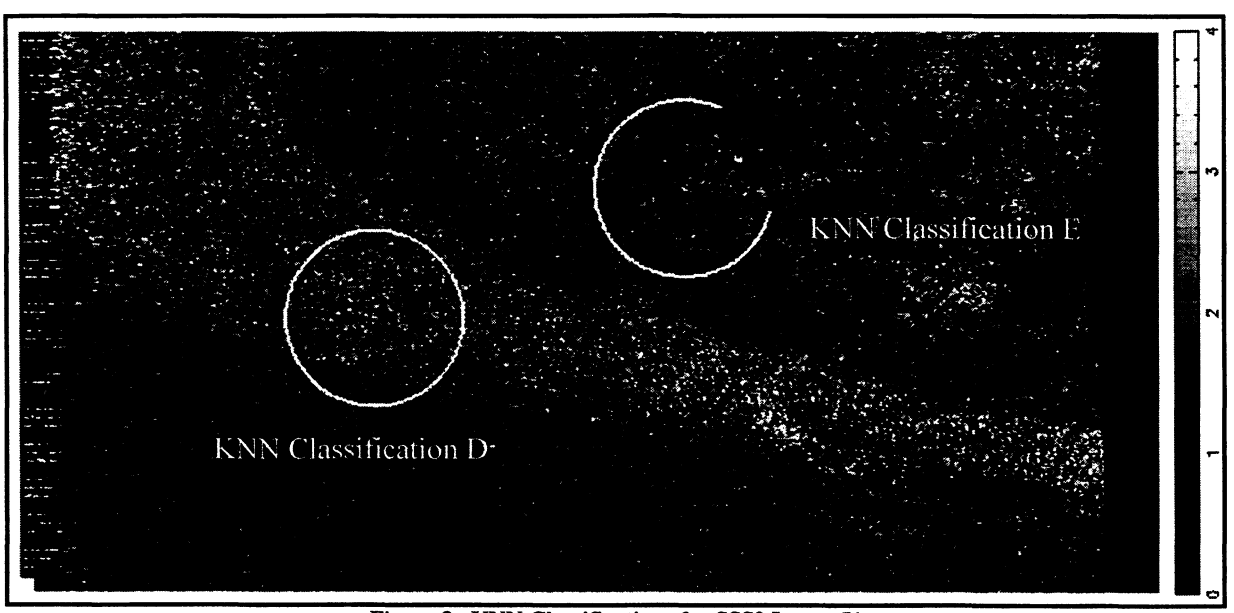

Figure 8. KNN Classifications for SSS2 Image 51

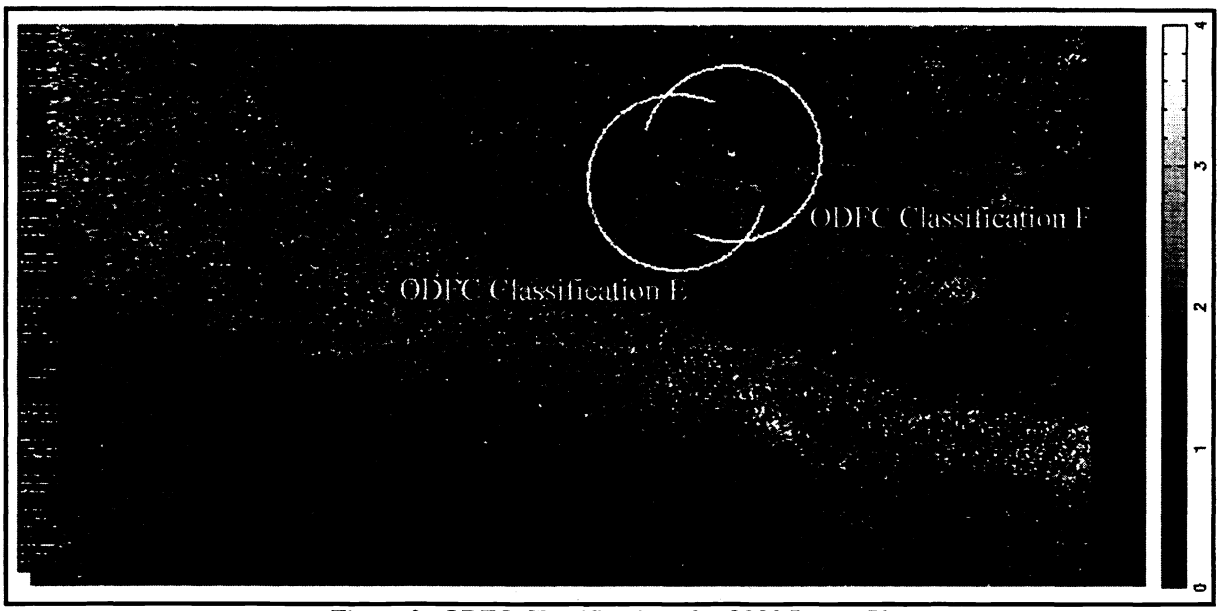

Figure 9. ODFC Classifications for SSS2 Image 51

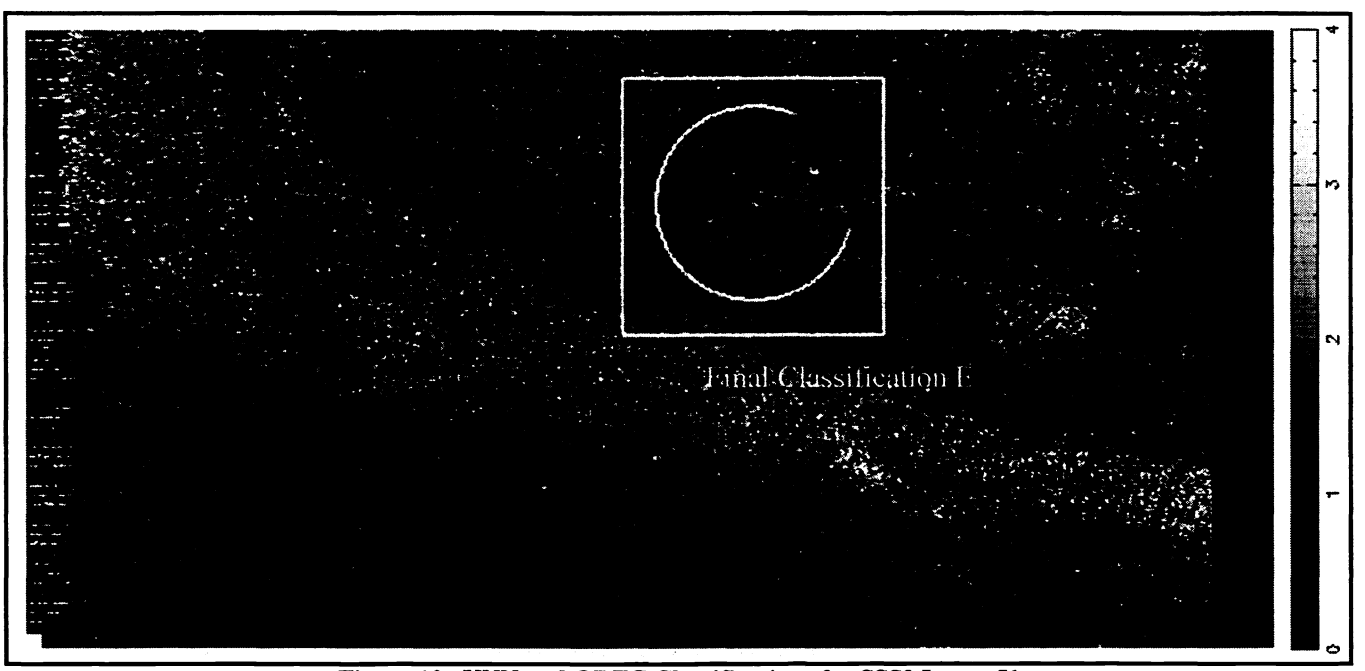

Figure 10. KNN and ODFC Classifications for SSS2 Image 51

3.3 SSS2 False Alarms, Missed Detections and Overall Performance

Overall, the AMDAC detected and classified 91% of the mines in the SSS2 data base with an average false alarm rate of 0.12 false alarms per image. Figure 11 shows the seven false alarms detected and three missed mines in the entire SSS2 data base. The false targets are quite mine-like. The top and bottom missed mines in Figure 11 have weak highlights and very poor shadows. The middle missed mine is embedded in very bright surface reverberation which made the mine signature blend in with the background.

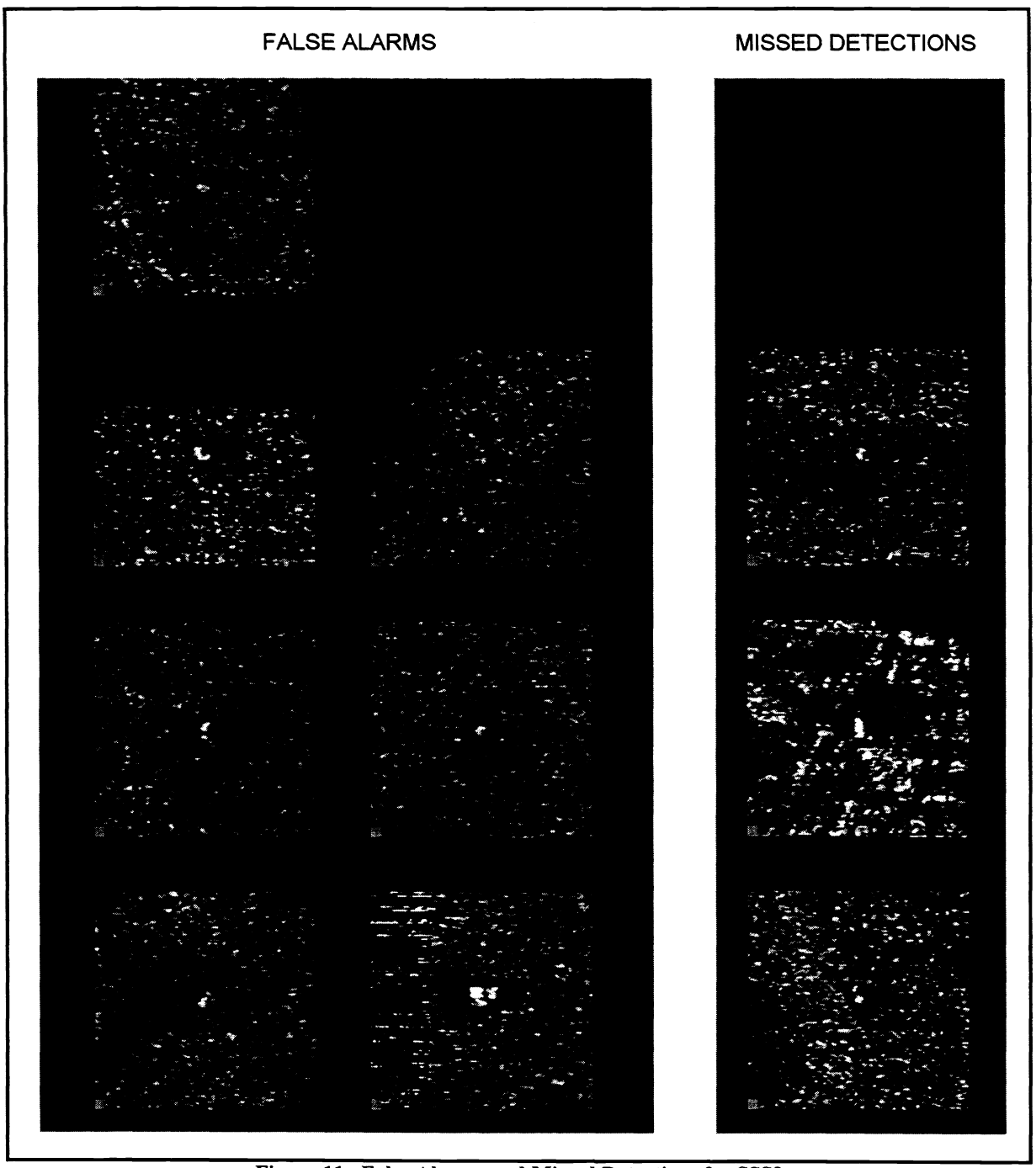

Figure 11. False Alarms and Missed Detections for SSS2

3.4 SAS Example

In Figures 12 through 15, Image 116 of the 215 image SAS image data base is used to further illustrate the D/C algorithm. The sonar image contains data from both the starboard and port sides of the sonar. Returns from the water column in the middle of the sonar image were blanked out prior to processing. Sonar operators ignore this region because it consists of returns from the water volume prior to the first bottom return. Note that there are no shadows on the targets; because the data was collected during shallow water operations, surface reverberation came into play and filled in the shadows. The lack of a target shadow makes the detection and classification process quite difficult. When a target exhibits only a highlight, the detection thresholds are typically set to accept a higher false detection rate. The classification phase is then relied upon to reject the extra false targets that are detected.

Figure 12 shows that the AMDAC has detected eight mine-like regions in the SAS image. All eight candidate regions do exhibit the desired mine-size highlight surrounded by a uniform background. Figure 13 shows that the KNN rejects two of the detections and Figure 14 shows the ODFC rejects five of the detections. "ANDING" the classifiers results in the classification of the two mines and one false target as shown in Figure 15.

3.5 Results Summary

Table 4 summarizes the AMDAC's performance for the all three sonars, the SSS1, the SAS and the SSS2. Table 4 also lists expert sonar operator's performance for the SSS1 and SAS data bases. Operator performance for the SSS2 is unavailable. Because of the similarities between the SSS1 and the SSS2, the operator performance for the SSS2 should be comparable to that of the SSS1.

For the SSS1, the AMDAC detects and classifies 90% of the mines with 0.42 false alarms per image while the expert operator detected 80% with 0.72 false alarms per image. For the SAS, the AMDAC performs similarly to the expert sonar operator at 68% PdPc. However, the AMDAC can also be adjusted to 92% PdPc with 0.64 false alarms per image. And for the SSS2, the AMDAC shows an impressive 91% PdPc with only 0.12 false alarms per image.

Table 4. AMDAC Performance Summary

Sonar	Operator		AMDAC	
	PdPc (%)	Fa/Image	PdPc (%)	Fa/Image
SSS1	80	0.72	90	0.42
SAS	68	0.23	68 92	0.21 0.64
SSS2	N/A	N/A	91	0.12

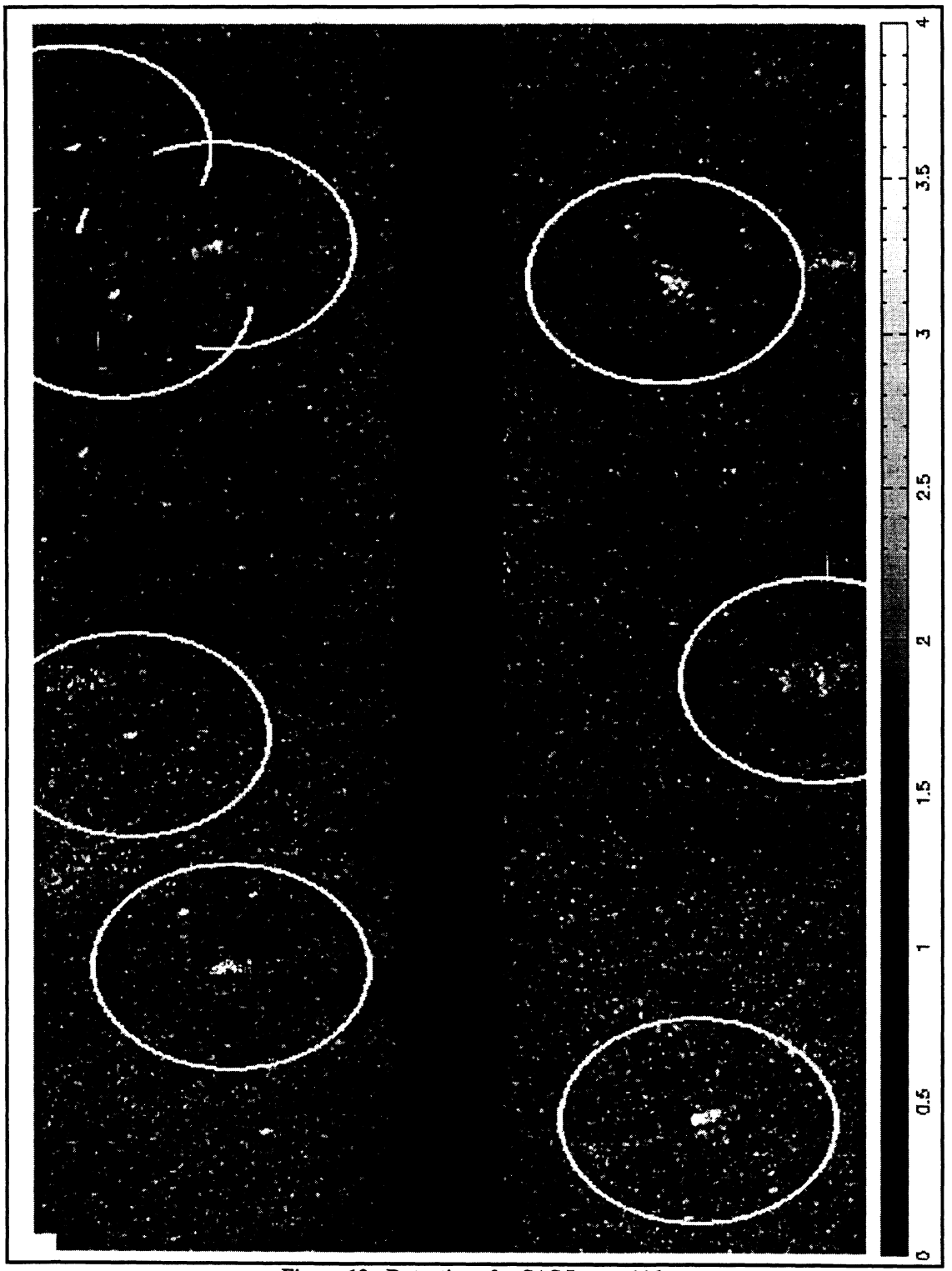

Figure 12. Detections for SAS Image 116

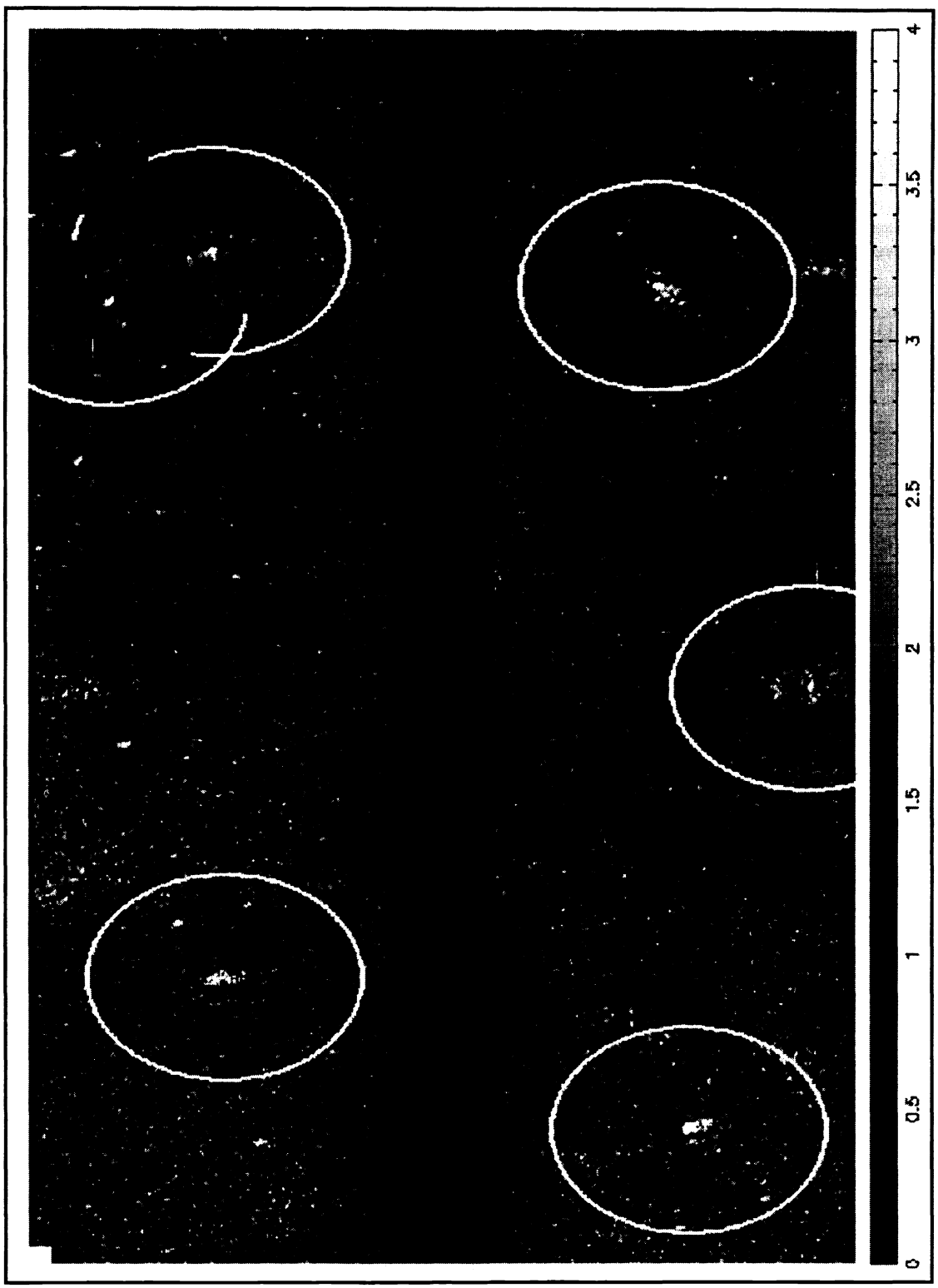

Figure 13. KNN Classifications for SAS Image 116

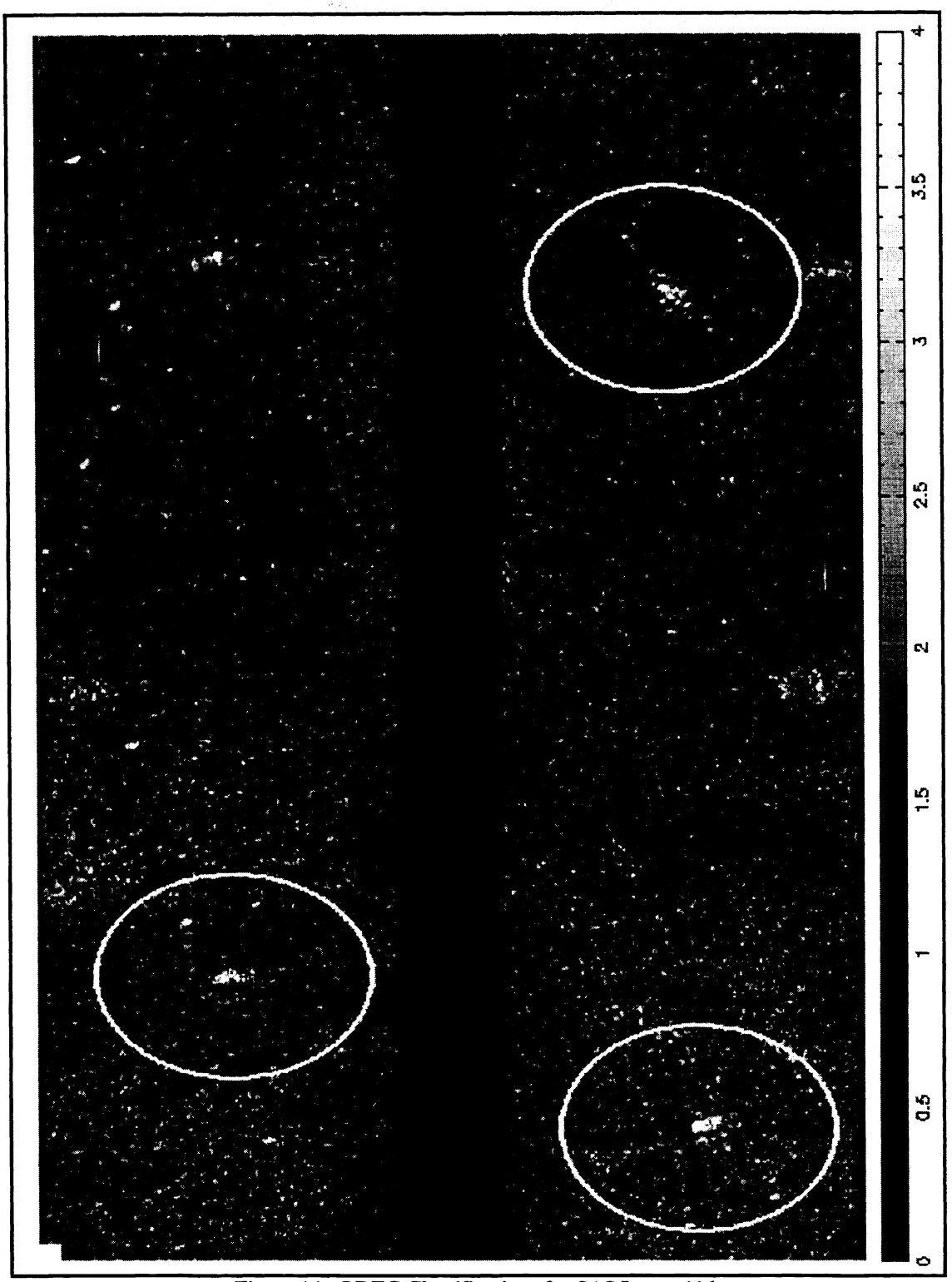

Figure 14. ODFC Classifications for SAS Image 116

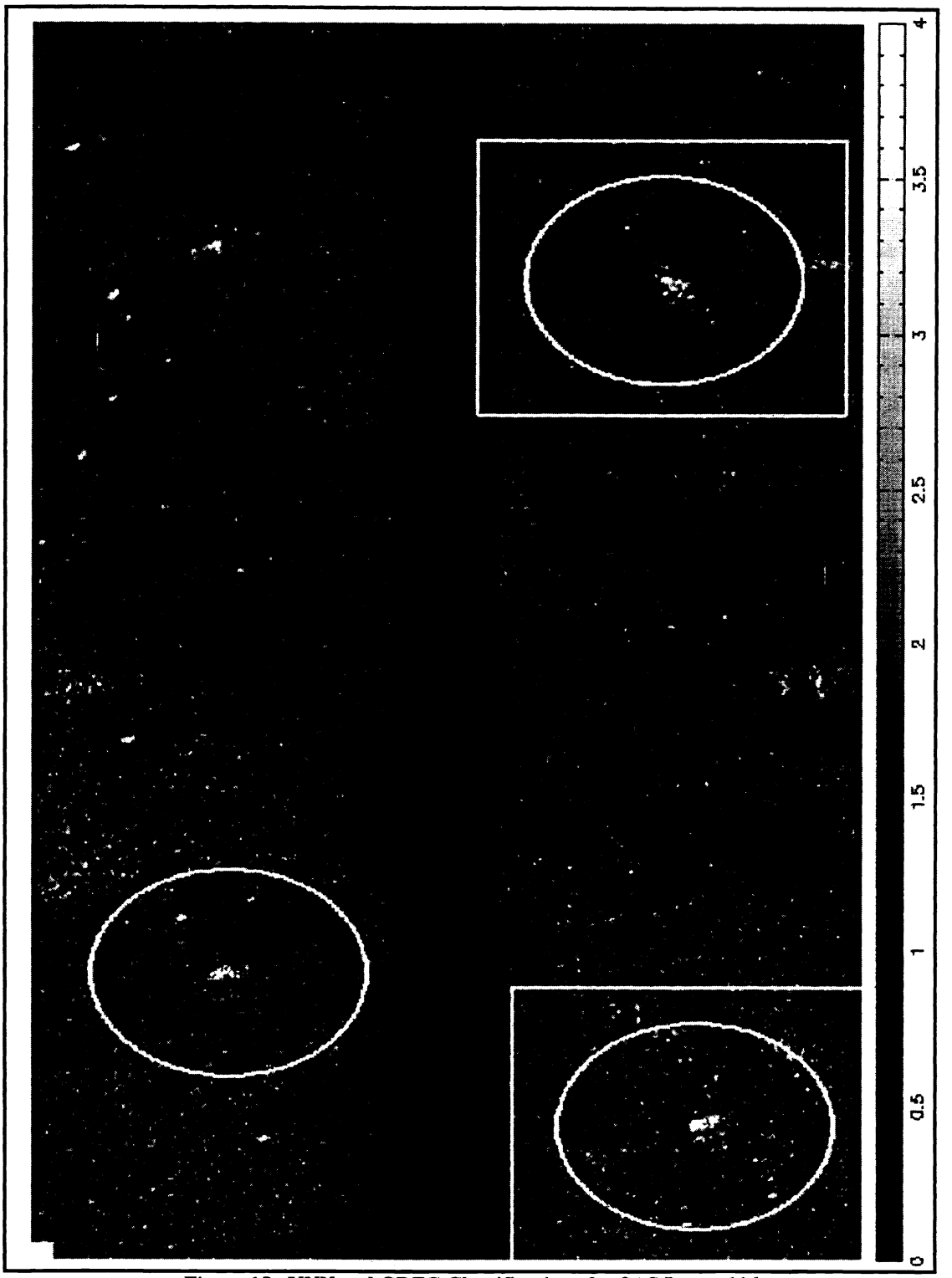

Figure 15. KNN and ODFC Classifications for SAS Image 116

4. CONCLUSIONS AND FUTURE RESEARCH

Five fundamental principles have come out of this research on automated detection and classification.

(1) The detection stage is by far the most computationally intensive stage. Because the entire image must be scanned, the basic design focus is on keeping the number of computations per pixel small. The main goal is not to detect mine-like objects, but rather to eliminate the majority of the image that is non-mine-like. The resulting detections will be processed in depth by the classification stages which require many more computations per pixel in order to extract robust and discriminatory mine-like features. If the detector produces 10 mine-like detections per image on the average (a very generous number), we have found that the computational requirements of our classification stages are negligible compared to those of the detector. It then follows that this classifier's main goal is to eliminate the false targets while preserving mine detections.

(2) The classification stage uses two classifiers: the KNN and ODFC. The stepwise optimal selection procedure, described herein, dramatically improves the performance of each classifier by permitting a computationally efficient means to select the "best" subset of classification features from a much larger set of candidate features. As such it mitigates problems associated with Bellman's curse of dimensionality.

(3) Significant improvement in classifier performance results if the feature selection process is highly "tuned" to the respective classifier. Classifiers use statistical, mathematical, and geometrical constructs to partition feature space into class regions. In this research we found that a subset of features optimal for one classifier's partitioning scheme will be far from optimal for another classifier which uses a different partitioning approach. Therefore, we us the same partitioning approach when determining the best feature subset for a particular classifier.

(4) Because training procedure for both the KNN and ODFC is very fast, it was possible to develop a stepwise optimal selection procedure capable of efficiently evaluating thousands of feature combinations. Without such an automated procedure it was impossible for us to determine the "best" features. Being able to use the "best" features dramatically improved the performance of our classifiers. This strongly suggests the following. Rapid-trainable classifiers, whose "best" features can be determined by stepwise optimal selection, may significantly outperform more complex classifiers that have computationally intense training procedures and must rely on selection methods that are either heuristic or not well "tuned" to the classifier.

(5) There is a significant improvement in overall classification performance that results from "ANDING" classifiers which use complementary statistical, mathematical, and geometrical constructs to describe class boundaries in feature space. "ANDING" the KNN and ODFC proved very adept at reducing false alarms while maintaining a high probability of mine detection and classification. This concept can be illustrated using the paradigm of "getting a second doctor's opinion". Two doctors look at data from a patient to arrive at their diagnoses. Even though they look at the same data, each doctor emphasizes or de-emphasizes different aspects based on their training and experience. If they conclude the same diagnosis, there is little doubt of its validity. This also suggests that "ANDING" two or more simple classifiers may lead to improved performance over a single more complex classifier. This leads the authors to believe that extremely high performance in automated mine D/C can be achieved by "ANDING" (fusing) the diverse D/C algorithms currently being developed through ONR under their Mine Countermeasures research programs.

Overall, the new detection and classification algorithm developed has been shown to perform significantly better than the detection and classification algorithm previously reported in SPIE 95 [1]. The new algorithm used in this study performs as well or slightly better than an expert sonar operator on the three sonar image data bases.

The algorithm's robustness will be further evaluated on new advanced synthetic aperture and side-looking sonars that are currently undergoing their initial sea testing. We will investigate the benefits of using three classifiers rather than just two. Using more than three classifiers produce some combinatorial design issues. Also, algorithms that adapt detection and classification thresholds will be developed in order to increase robustness to varying environments and bottom types.

5. ACKNOWLEDGEMENTS

This work was funded by the Office of Naval Research, ONR 321TS as part of the 6.2 Mine Countermeasures program element The technical agent for this work is the NSWC Coastal Systems Station, Dahlgren Division. The point of contact is Dr. Gerald J. Dobeck, Code 10T2, Panama City, Florida 32407-7001, Phone: 904-234-4222.

6. REFERENCES

[1] 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: Detection Technologies for Mines and Mine-like Targets, 17-21 April 1995, Orlando, Florida.

[2] M. G. Bello, "Hierarchical Multilayer Perceptron Network-Based Fusion Algorithms for Detection/Classification of Mines Using Multiple Acoustic Images and Magnetic Data," Proceedings of SPIE '96, Vol. 2765, 9-12 April, 1996, Orlando, Florida.

[3] T. Aridgides et al,, "Adaptive-Filter/Feature Orthogonalization Processing String for Optimal LLRT Mine Classification in Side-Scan Sonar Imagery," Proceedings of SPIE '96, Vol. 2765, 9-12 April, 1996, Orlando, Florida.

[4] D. Kil and F. Shin, "The Impact of Lossy Image Compression on Automated Target Recognition Performance," Proceedings of IEEE Oceans '96, September, 1996, Fort Lauderdale, Florida .

[5] S. Nelson and S. Tuovila, "Automated Recognition of Acoustic Image Clutter," Proceedings of SPIE '96, Vol. 2765, 9-12 April, 1996, Orlando, Florida.

[6] H. Van Trees, Detection, Estimation and Modulation Theory, Part I, John Wiley and Sons, 1968.

[7] R. Duda and P. Hart, Pattern Recognition and Scene Analysis, John Wiley and Sons, 1973.

[8] K. Enslein, A. Ralston and H. Wolf, Statistical Methods for Digital Computers, Volume II, John Wiley and Sons, 1977.

[9] L. Ott, An Introduction to Statisitcal Methods and Data Analysis, Third Edition, PWS-KENT Publishing Company, Boston, Massachusetts, 1988.

[10] B. D. Rilpley, Pattern Recognition and Neural Networks, Cambridge University Press, 1996.