Canonical Coordinates for Detection and Classification of Underwater Objects From Sonar Imagery

James D. Tucker [a], Mahmood R. Azimi-Sadjadi [a], and Gerry J. Dobeck [b]
[a] Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, Colorado 80523, Email: dtucker@engr.colostate.edu
[b] Naval Surface Warfare Center Panama City Panama City, FL, USA 32407-7001

***Abstract*—Detection and classification of underwater objects in sonar imagery are challenging problems. In this paper, a new coherent-based method for detecting and classifying potential targets in high-resolution sonar imagery is developed using canonical correlation analysis (CCA). Canonical coordinate decomposition allows us to quantify the changes between the returns from the bottom and any target activity in sonar images and at the same time extract useful features for subsequent classification without the need to perform separate detection and feature extraction. Moreover, in situations where any visual analysis or verification by human operators is required, the detected/classified objects can be reconstructed from the coherent features. In this paper, underwater target detection and classification using the canonical correlations extracted from regions of interest within the sonar image is considered. Test results of the proposed method on underwater side-scan sonar images provided by the Naval Surface Warfare Center (NSWC) in Panama City, FL is presented. This database contains synthesized targets in real background varying in degree of difficulty and bottom clutter. Results illustrating the effectiveness of the CCA-based detection and classification method are presented for various densities of background clutter and bottom difficulty.**

I. Introduction

Detection and classification of underwater objects in sonar imagery is a complicated problem due to various factors such as variations in operating and environmental conditions, presence of spatially varying clutter, variations in target shapes, compositions and orientation. Moreover, bottom features such as coral reefs, sand formations, and the vegetation may totally obscure a target object.

Various methods have been explored for target detection and classification in sonar imagery. In [1], [2] Dobeck utilizes a nonlinear matched filter to identify mine-size regions in the sonar image that match the target signature. For each detected region, several features are extracted based on the size, shape, and strength of the target signature. A stepwise feature selection process is then used to determine the subset of features that optimizes the probability of detection and classification. A k-nearest neighbor and an optimal discrimination filter classifier are used to classify each feature vector and the decisions of the two classifiers are fused for the final decision. In [3] Ciany segments the sonar images into "sub-frames" on which each frame is adaptively thresholded to identify the target structure. Geometric features are then extracted from contiguous target structure regions of interest within the sub-frame. Classification of each region as target or non target is done through a multi-level weighted scoring-based classification system. In [4] Adridges presented an adaptive clutter filter detector which exploits the difference in correlation characteristics between clutter and targets. After detection, features are extracted and then orthogonalized and then the classification is performed on the orthogonal feature set through an optimal Bayesian classifier. Chandran in [5] presented the use of a matched filter designed to capture the target structure. Higher order spectra are then extracted as the feature set to classify objects using a k-nearest neighbor classifier, a minimum distance classifier, and a threshold classifier based on the minimum and maximum values of a feature obtained over all classes, and the final decision are fused.

One detection and feature extraction method that has not been explored for sonar imagery is Canonical Coordinate Analysis or (CCA). This method has shown great promise in underwater target classification problems using sonar backscatter [6]. The canonical coordinate decomposition method determines linear dependence [7] or coherence between two data channels. This method not only determines the amount of dependence or independence between two data channels (e.g. two sonar pings with certain separation) but also extracts, via the canonical coordinates, a subset of the most coherent features for classification purposes. Canonical coordinate decomposition allows us to quantify the changes between the returns from the bottom and when target activities are present and at the same time extract useful features for target classification without the need to perform separate detection and anomaly feature extraction.

In this paper a new coherent-based detection and classification method for high-resolution sonar imagery is developed using CCA as an optimal Neyman-Pearson detection scheme and a feature extraction process. In both cases, the canonical correlations are formed from region of interests (ROI) within the sonar image. From these canonical correlations, coherence (or incoherence) can be measured and used to determine if a target is present in the processed ROI and then the extracted

features can then be used to classify the detected ROI's. The data set used in this study was provided by the NSWC in Panama City, FL. The data-set consists of high-resolution side-looking sonar imagery that contains either no targets, one target, or multiple targets.

This paper is organized as follows: Section II reviews the CCA method and its application as a feature extraction (estimation framework) or as a detection tool for implementing the Neyman-Pearson detector. Section III describes the preprocessing and feature extraction process done on the NSWC Scrub data set to extract the canonical correlations. In Section IV, the detection and classification results of using CCA to detect and classify of underwater targets in sonar imagery are presented. Finally, conclusions and observations are made in Section V.

II. Review of Canonical Coordinate Decomposition

CCA is a method that determines linear dependence (or coherence) between two data channels by mapping the data to their *canonical coordinates* where linear dependence is easily measured by the corresponding *canonical correlations*. The language and terminology used in this section is taken mostly from [7], [8].

Consider two data channels $\mathbf{x} \in \mathbb{R}^{m\times 1}$ and $\mathbf{y} \in \mathbb{R}^{n\times 1}$ where $m \leq n$, when it is assumed that $\mathbf{x}$ and $\mathbf{y}$ are zero mean random vectors that yield a composite covariance matrix

$$E\left[\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \begin{pmatrix} \mathbf{x}^H & \mathbf{y}^H \end{pmatrix} \right] = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix}, \quad (1)$$

where, R_{xx}, R_{yy}, and $R_{xy} = R_{yx}^H$, are the covariance matrices of data channels $\mathbf{x}$, $\mathbf{y}$, and between $\mathbf{x}$ and $\mathbf{y}$ respectively. The singular value decomposition (SVD) of the coherence matrix C may then be written as [7]

$$\begin{gathered} C = R_{xx}^{-1/2} R_{xy} R_{yy}^{-H/2} = FKG^H \quad \text{and} \quad F^H C G = K, \\ F^H F = I, \;\; G^H G = I, \qquad K = \text{diag}[k_1, k_2, \ldots, k_m]; \end{gathered} \quad (2)$$

where $R_{xx}^{-1/2} R_{xx} R_{xx}^{-H/2} = I$, and $R_{xx}^{1/2} R_{xx}^{H/2} = R_{xx}$. Note that for this SVD, only the nonzero singular values of C and their corresponding nonzero vectors are considered.

The *canonical coordinates* of $\mathbf{x}$ and $\mathbf{y}$ can now be defined as [7]

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} F^H & 0 \\ 0 & G^H \end{bmatrix} \begin{bmatrix} R_{xx}^{-1/2} & 0 \\ 0 & R_{yy}^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \quad (3)$$

where elements of the $\mathbf{u}$ and $\mathbf{v}$ vectors are the canonical coordinates of $\mathbf{x}$ and $\mathbf{y}$, respectively.

Hence, $\mathbf{x}$ and $\mathbf{y}$ are mapped to their respective canonical coordinates using

$$\mathbf{u} = W^H \mathbf{x} \quad \text{and} \quad \mathbf{v} = D^H \mathbf{y}. \quad (4)$$

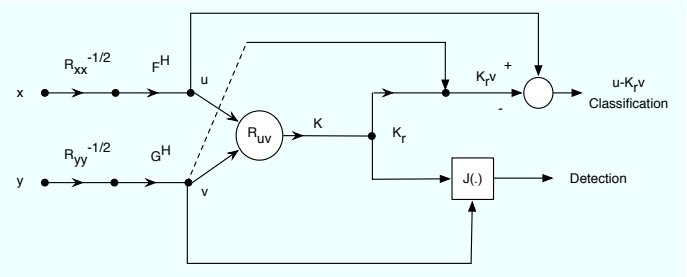

Fig. 1. Canonical Coordinate Decomposition Process.

via the canonical decomposition matrices $W^H = F^H R_{xx}^{-1/2}$ and $D^H = G^H R_{yy}^{-1/2}$ (see Figure 1). The canonical coordinates $\mathbf{u}$ and $\mathbf{v}$ share the composite covariance matrix

$$\begin{aligned} E\left[\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \begin{pmatrix} \mathbf{u}^H & \mathbf{v}^H \end{pmatrix} \right] &= \begin{bmatrix} R_{uu} & R_{uv} \\ R_{vu} & R_{vv} \end{bmatrix} \\ &= \begin{bmatrix} W^H R_{xx} W & W^H R_{xy} D \\ D^H R_{yx} W & D^H R_{yy} D \end{bmatrix} = \begin{bmatrix} I & K \\ K & I \end{bmatrix}, \end{aligned} \quad (5)$$

where the diagonal cross-covariance matrix

$$K = R_{uv} = F^H C G = W^H R_{xy} D \quad (6)$$

is the *canonical correlation* matrix of canonical correlations k_i; $i = 1, 2, \ldots, m$, with $1 \geq k_1 \geq k_2 \geq \cdots \geq k_m > 0$ [7]. Note that we have $E[u_i v_j] = k_i \delta(i-j)$, where $\delta(\cdot)$ represents the Kronecker delta function.

One of the most important properties of canonical correlations is that they are invariant under uncoupled nonsingular transformations of $\mathbf{x}$ and $\mathbf{y}$ [7]. Canonical correlations can also be used to determine useful properties regarding $\mathbf{x}$ and $\mathbf{y}$, such as linear dependence (coherence) and mutual information [7].

A. CCA As An Detection Tool [8]

The detection problem is stated as a hypothesis testing problem of distinguishing hypothesis of $H_0 : \mathbf{y} : CN_n[0, R_{nn}]$, i.e. noise alone, versus hypothesis $H_1 : \mathbf{y} : CN_n[0, R_{yy} = R_{xx} + R_{nn}]$, i.e. signal plus noise, where $CN_n[0, R_{xx}]$ denotes the n-variate proper complex normal distribution with mean vector zero and covariance matrix R_{xx}. In [8] the Neyman-Pearson detector for testing H_0 and H_1 as defined above is written as the log-likelihood ratio in terms of canonical coordinates and correlations as,

$$\begin{aligned} l(\mathbf{y}) = &(R_{yy}^{-1/2}\mathbf{y})^H (I - C^H)(R_{yy}^{-1/2}\mathbf{y})) \\ &+ (G^H R_{yy}^{-1/2}\mathbf{y})^H ([I - K^2]^{-1} - I)(G^H R_{YY}^{-1/2}\mathbf{y}). \end{aligned} \quad (7)$$

This is the standard Gauss-Gauss log-likelihood ratio, but in the canonical coordinates $\mathbf{v} = G^H R_{yy}^{-1/2}\mathbf{y}$ and $l(\mathbf{y})$ is the weighted sum of the magnitude-squared of the canonical coordinates weighted by canonical correlation-dependent weights. If we define $G = [g_1, \cdots, g_n]$, then $l(\mathbf{y})$ can be written as follows,

$$l(\mathbf{y}) = \sum_{i=1}^{n} |g_i^H R_{yy}^{-1/2}\mathbf{y}|^2 \left(\frac{k_i^2}{1 - k_i^2} \right) \quad (8)$$

The J-divergence [8] between H_1 and H_0 is

$$\begin{aligned} J &= E_{H_1} l(\mathbf{y}) - E_{H_0} l(\mathbf{y}) = tr(CC^H + (CC^H)^{-1} - 2I) \\ &= tr(KK^H + (KK^H)^{-1} - 2I) \end{aligned} \tag{9}$$

where E_{H_0} and E_{H_1} are the expected values of $l(\mathbf{y})$ under H_0 and H_1, respectively. We then can rewrite the J-divergence in terms of the canonical correlations as,

$$J = \sum_{i=1}^{m^2} (k_i - \frac{1}{k_i})^2 \tag{10}$$

The function $(k_i - \frac{1}{k_i})^2$ is non-increasing in the interval (0, 1]. *Consequently the rank-r detector that maximizes the divergence is the detector that uses the dominant coordinates corresponding to the dominant canonical correlations k_i.* The divergence between the two hypothesis considering the r dominant canonical correlations is,

$$J_r = \sum_{i=1}^{r^2} (k_i - \frac{1}{k_i})^2 \tag{11}$$

Thus, for building low-rank detectors, the dominant canonical coordinates need to be retained in order to find the coherence between the two data channels $\mathbf{x}$ and $\mathbf{y}$. Using the coherence one can find the information necessary to detect presence of a target in the environment.

B. CCA As An Estimation or Feature Extraction Tool [8]

The estimation problem is stated as estimating channel $\mathbf{x}$ (signal) from channel $\mathbf{y}$ (observation). The minimum mean square error estimator of $\mathbf{x}$ from $\mathbf{y}$ can be written as [8],

$$\hat{\mathbf{x}} = R_{xx}^{-1/2} F K G^H R_{yy}^{-1/2} \mathbf{y}, \tag{12}$$

with minimum error covariance

$$Q_{xx} = R_{xx}^{-1/2} F (I - K^2) F^H R_{xx}^{H/2}. \tag{13}$$

The volume of the error concentration ellipse divided by the volume of the prior concentration ellipse is

$$V = \frac{\det(Q_{xx})}{\det(R_{xx})} = \det(I - K^2). \tag{14}$$

The processing gain (PG) and the information rate (R) given [8] in terms of V as $PG = V^{-1}$ and $R = -(1/2)\log V$, respectively. It is clear that the *dominant* canonical coordinates are the ones that minimize V, maximize PG, and maximize R. Thus the optimal rank-$r \leq n$ estimator of $\hat{\mathbf{x}}$, of $\mathbf{x}$, from $\mathbf{y}$, is the estimator that retains only the dominant canonical coordinates and maximizes the coherence between $\mathbf{x}$ and $\mathbf{y}$. In canonical coordinate domain this rank-r estimator can be defined as,

$$\hat{\mathbf{x}} = R_{xx}^{-1/2} F K_r G^H R_{yy}^{-1/2} \mathbf{y}, \tag{15}$$

and it can be shown

$$\hat{\mathbf{u}} = K_r \mathbf{v}, \tag{16}$$

where K_r holds the top r dominant canonical correlations. In this framework, classification can then be made by using either the dominant canonical correlations as a feature vector or using $\mathbf{u} - K_r \mathbf{v}$ which shows how well $\mathbf{x}$ can be estimated from $\mathbf{y}$ where the target is more coherent than the environment in which it is found.

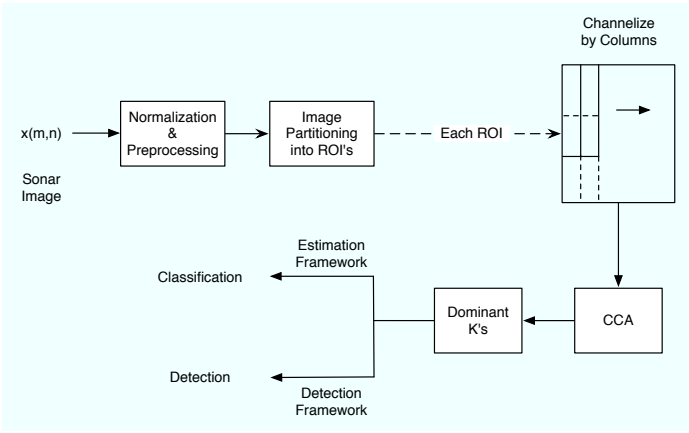

Fig. 2. Block Diagram of the Side Scan Sonar CCA-based Detection Method.

III. CCA Analysis of High Resolution Sonar Imagery

In order to prepare the data for CCA, the high resolution images are first normalized using a serpentine forward-backward filter [9]. The purpose of this normalization is to help distinguish the target's highlight and shadow signature from the bottom and artifacts present in the image. The purpose of the normalization is to reduce the variability of the local mean throughout the image in order to use it as a reference level so that the highlight and shadow of the target can be more easily identified. The serpentine forward-backward filter (SFBF) normalizer first uses a forward second order digital filter and attempts to select a path in which the filter output best follows the original image. Basically the path is generated by taking the next point on the path to a point where its neighbors intensity are near the filtered output value. The purpose of the forward and backward filters is to estimate the local mean on one side of a pixel and the backward filter is used to estimate the local mean on the opposite side of the pixel. The local mean estimate that is nearest to the original image value is selected to normalize that pixel. Processing the image in the forward-backward direction helps preserve target edges, highlights, and shadows. Any background non-ideals, i.e. target size artifacts, are normalized out to the underlying mean of the image. For a more detailed explanation of the normalization algorithm the reader is referred to [9].

After the normalization process, the first 120 pixels are ignored which corresponds to the sonar altitude as it traveled through the water column, which is $1/10^{th}$ of the maxium range. Next, the image is partitioned into overlapping ROI's of size $M \times N$. For this data set the ROI size of 12×34 was experimentally determined to be optimal considering the average size and shape of the targets in the data set. The overlap along the horizontal and vertical directions was 50%

in order to ensure that a target would be covered by more than one ROI in case of splitting. Each ROI is then channelized in a column-wise fashion for CCA. The $\mathbf{x}$ and $\mathbf{y}$ channels consist of the first 8 pixels in one column $\mathbf{x}$ and the first 8 pixels in the adjacent column $\mathbf{y}$. This process is continued moving in the horizontal direction across the ROI. On the next pass through the ROI, the channels are given a 50% overlap in the vertical direction to ensure complete coverage of the target in the ROI (see Figure 2). The idea behind this channelization is to look for common coherent attributes that can be used to relate one channel to the other according to the framework discussed in Section II-A. Clearly, for background ROI's high level of coherence among consecutive columns (channels) does not exist. Note that in this paper we use the dominant canonical correlations which hold the coherent information between the two channels. One may also use the subdominant canonical correlations in the detection framework to detect incoherence (or change) between the two channels. From the dominant canonical correlations, namely k_1 and k_2, a scalar detection measure of $k_1 \times k_2$ was formed for each processed ROI. Based on this measure (instead of J-divergence in (11)) a threshold is determined experimentally to separate the ROI's that contain targets from those that do not. Following the detection, all the canonical correlations extracted from each ROI are used to classify target and non-target ROI's using a back propagation neural network (BPNN) classifier.

IV. Test Results on Sonar Imagery

CCA was applied to a data set of high-resolution side-looking sonar images provided by the NSWC, Panama City. More information on high-resolution side-looking sonars can be found in [10], [11]. The database contains 512 images with 293 images containing 310 targets with some of the images containing more than one target. The data set was broken up into easy, medium, and hard cases depending on the difficultly of the background clutter and bottom types, e.g. vegetation and coral reefs. Easy cases are considered to have low background variation and an overall smooth bottom with targets that are easily identifiable by a skilled operator. The medium cases contain background clutter and more difficult bottom conditions. However, the targets are still somewhat discernible to a skilled operator with some effort. Lastly, the hard cases are those where it is difficult to detect and classify the targets from a visual inspection due to a high variability of background clutter and very difficult bottom conditions. The data set was separated into these three classes based on a visual inspection of where the target was located and whether or not variation in the background and high density of clutter was present.

To show the separability of the dominant canonical correlations for ROI's that contain targets and background and those that contain only background, a test was conducted on the entire set of target ROI's and a random set of ROI's containing mainly background (for all three cases) of the same number of target ROI's. The plots of the 8 canonical correlations of ROI's containing targets and those containing

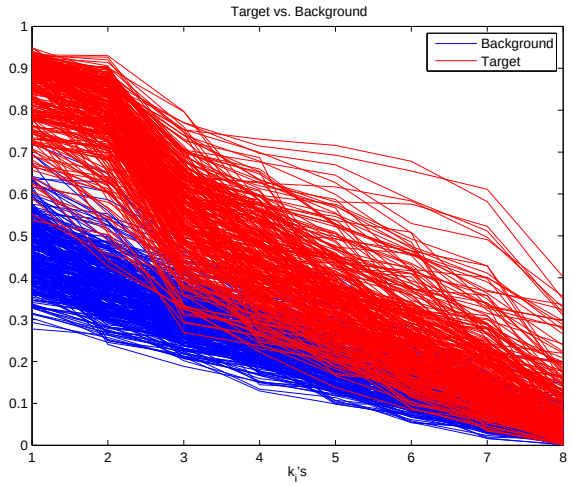

Fig. 3. Plot of Canonical Correlations for Target and Background.

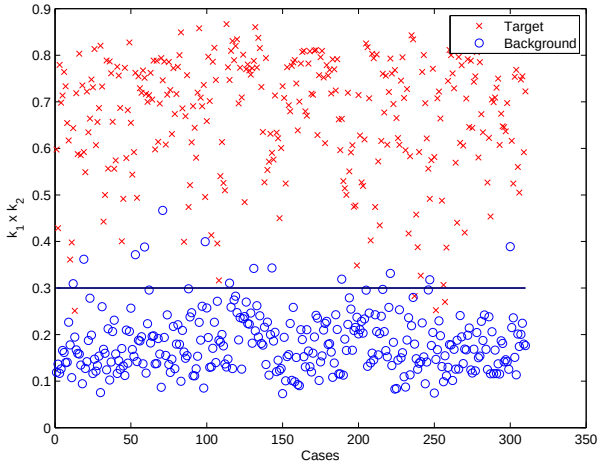

Fig. 4. Separation of Target versus Background for $k_1 \times k_2$.

background only are shown in Figure 3. As can be seen, there is good separation, especially for dominant canonical correlations, between targets and background, which can be attributed to the greater coherence in the channels across the targets versus the background where there is more pixel variation. This is attributed to the greater coherence between $\mathbf{x}$ and $\mathbf{y}$ over an ROI where a target is present.

Using the dominant canonical correlations, k_1 and k_2 a scalar decision measure of $k_1 \times k_2$ can be formed for the entire target set. This scalar value was formed and plotted for the entire target set and random set of backgrounds and is presented in Figure 4. A detection threshold was chosen based upon this measure to detect potential targets from the background. From this set of targets and limited number of backgrounds the optimal threshold value was chosen to be 0.3.

Each image in the entire NSWC database is then blocked

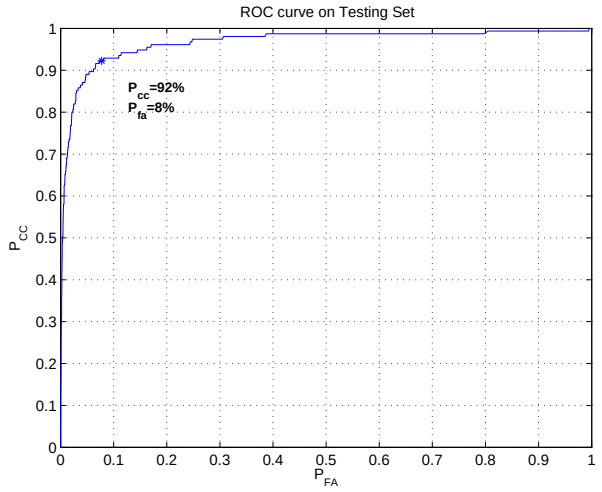

Fig. 5. Performance Curve for Two Layer BPNN Classifier on Testing Set.

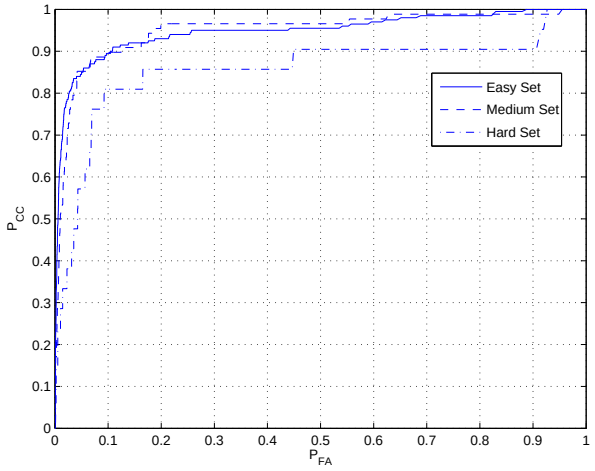

Fig. 6. Performance Curve for Two Layer BPNN Classifier on Target Sets.

into 12×34 ROI's with a 50% overlap for each block in both the range and cross-range directions. For each ROI the canonical correlations were formed as well as $k_1 \times k_2$ scalar detection measure and then compared against the optimal threshold. For the easy cases, there are 186 images containing 201 targets. The canonical correlation detector detected all but 1 of the targets in the set with an average of 127 detections per image. For the medium cases, there are 86 images containing 89 targets and the canonical correlation detector again detected all but 1 of the targets in the set with an average of 213 detections per image. Lastly, for the hard cases, there are 21 images, containing 21 targets, and the canonical correlation detector detected all but 2 of the targets and averaged 228 detections per image. It is important to note that with 50% overlap in the ROI's there will be at least 4 ROI's overlapping the object depending on the object length. Taking this into account into the number of detections, it was determined which of the detected ROI's were actually over the target which then gives an average of 116 detections on the easy set, 200 detections on the medium set, and 213 detections on the hard set.

The detected ROI's were then broken up into training and a testing set to train and test BPNN classifier. A network structure was determined experimentally and the structure that performed the best was a two-layer network with 8 inputs, 20 neurons in the first hidden layer, and 2 output neurons. The receiver operator characteristic (ROC) curve for the entire testing set is presented in Figure 5. The BPNN performs extremely well on the testing set with the knee point on the ROC gives 92% correct classification rate and 8% false alarm rate. This shows the power of the CCA-based feature extraction method.

The trained classifier was then applied to all the detected ROI's in the easy, medium, and hard data sets that were mentioned earlier. The classifier ROC curves are presented in Figure 6 for the three cases. The classifier performs extremely well on the easy and medium sets with the knee points at 90% correct classification rate and 10% false alarm rate for both cases. The classifier did not perform as well on the hard data set with only 84% correct classification and 16% false alarm rate. Nonetheless, considering the difficulty of the bottom conditions and background clutter the classification rate is acceptable. Overall, the classifier performs well on all the data sets by keeping the correct classification rate high and the false alarm rate low using just the canonical correlations as the feature vector.

V. Conclusions and Observations

In this paper, CCA was used as an optimum Neyman-Pearson detector to detect underwater targets and as a feature extraction tool for classification of underwater targets in high-resolution side-looking sonar imagery. The basic idea is that when an object (target or non-target) is present in an ROI in the sonar image there will be a change in the coherence level compared to the case when there is no object. Further it has been shown that using the dominant canonical correlations a scalar decision measure can be formed in order to separate the ROI's that contain a target from those that contain background only and can also be used to classify the ROI's that contain a target from those do not. The extracted features, or canonical correlations, can then also be used as a classification tool to classify the ROI's which contain a target and those that do not without the need to perform a separate anomaly feature extraction for classification. Our experiments on the data set provided by the NSWC demonstrated that there is good separability of the canonical correlations, especially the dominant ones, extracted over ROI's containing targets from those extracted from ROI's that do not contain targets. Overall, CCA did well in detecting all of the targets in all of the images missing only 4 of the possible 310 targets in the set, while keeping the probability of detection high and false alarm rate low. Classification using CCA features gave good results given

the limited size of the training set. The results showed an average of 88% correct classification rate with an average of 12% false alarm rate. Thus, CCA shows great promise as a primary detection and classification tool for underwater targets in sonar imagery.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research, Code 321OE under contract #N61331-06-C0027

REFERENCES

[1] G. J. Dobeck, J. Hyland, and L. Smedley, "Automated detection/classification of sea mines in sonar imagery," *Proc. SPIE*, vol. 3079, pp. 90–110, April 1997.

[2] G. J. Dobeck, "Fusing sonar images for mine detection and classification," *Proc. SPIE*, vol. 3710, pp. 602–614, April 1999.

[3] C. Ciany and J. Huang, "Data fusion of VSW CAD/CAC algorithms," *Proc. SPIE*, vol. 4038, pp. 413–420, April 2000.

[4] T. Aridges, P. Libera, M. Fernandez, and G. J. Dobeck, "Adaptive filter/feature orthogonalization processing string for optimal LLRT mine classification in side-scan sonar imagery," *Proc. SPIE*, vol. 2765, pp. 110–121, April 1996.

[5] V. Chandran, S. Elgar, and A. Nguyen, "Detection of mines in acoustic images using higher order spectral features," *IEEE Journal of Oceanic Engineering*, vol. 27, no. 3, pp. 610–618, July 2002.

[6] A. Pezeshki, M. Azimi-Sadjadi, L. Scharf, and M. Robinson, "Underwater target classification using canonical correlations," *Proceedings of MTS/IEEE Oceans 2003*, vol. 4, pp. 1906–1911, Sept 2003.

[7] L. Scharf and C. Mullis, "Canonical coordinates and the geometry of inference, rate, and capacity," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 824–891, March 2000.

[8] A. Pezeshki, L. Scharf, J. K. Thomas, and B. D. Van Veen, "Canonical coordiantes are the right coordinates for low-rand gauss-gauss detection and estimation," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4817–4820, Dec 2006.

[9] G. J. Dobeck, "Image normalziation using the serpentine forward-bckward filter: Application to high-reoslution sonar imagery and its imact on mine detection and classification," *Proc. SPIE*, vol. 5734, pp. 90–110, April 2005.

[10] N. Rimski-Korsakov, Y. Russak, and R. Pavlov, "Simple digital system for side scan sonar data imaging," *Proc. Oceans'94*, vol. 1, pp. 643–645, Sept 1994.

[11] W. Key, "Side scan sonar technology," *Proc. Oceans'00*, vol. 2, pp. 1029–1033, Sept 2000.