Coherence-Based Target Detection and Classification for Side Scan Sonar Imagery

*The Use of Canonical Correlation Analysis
For Sonar Target Detection and Classification*

**By J. Derek Tucker**
Graduate Student
and
Mahmood R. Azimi-Sadjadi
*Professor
Department of Electrical and
  Computer Engineering
Colorado State University
Fort Collins, Colorado*

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 and variations in target shapes, compositions and orientation. Moreover, bottom features, such as coral reefs, sand formations and vegetation, may totally obscure a target.

Various methods have been explored for target detection and classification in sonar imagery. In some cases, a nonlinear matched filter is utilized to identify mine-sized regions in the sonar image that match the target signature.[1,2] 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—a classifier using minimum distance from a current object feature vector to the training feature vector, 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 another method, an adaptive clutter filter is used, which
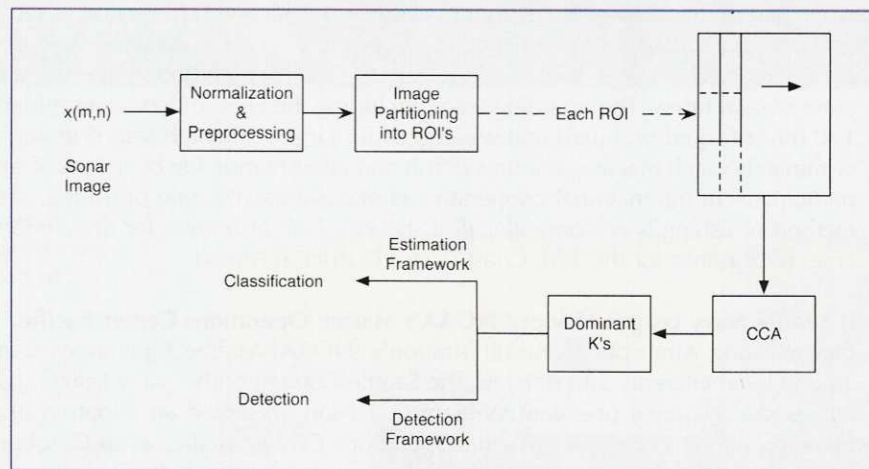
Block diagram of the CCA-based detection method.

exploits the difference in correlation characteristics between clutter and targets.[3] After detection, features are extracted and then orthogonalized prior to classification using a optimal Bayesian classifier.

The Canonical Coordinate Analysis (CCA) method has shown great promise in underwater target classification problems using sonar backscatter.[4] CCA allows one to quantify the changes between the returns from the bottom when target activities are present and, at the same time, provide via canonical correlations a set of features for target classification without the need to perform separate detection and feature extractions.

In this article, the coherent-based detection and classification method has been extended to high-resolution sonar imagery.[4,5] Using CCA, an optimal Neyman-Pearson detection scheme is developed that utilizes the canonical

correlations and coordinates extracted from regions of interest (ROIs) within the sonar image. From the canonical correlations, coherence can be measured and used to determine if a target is present in the processed ROI, while at the same time provide coherent-based features that can be used to classify the detected ROIs. The data set used in this study was provided by the Naval Surface Warfare Center (NSWC) in Panama City, Florida. The data set consists of high-resolution side-looking sonar imagery that contains either no targets, one target or multiple targets in varying clutter densities.

In the next section, a brief review of the CCA method and its application as a feature extraction (estimation framework) or detection tool for implementing the Neyman-Pearson detector is provided.

CCA Decomposition Review

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.

To see this process, consider two data channels: $x \in R^{m \times 1}$ and $y \in R^{n \times 1}$, where m is less than n. It is assumed that x and y are zero-mean random vectors with a composite covariance matrix shown in Equation 1 (shown on page 13), where R_{xx}, R_{yy} and $R_{xy} = R^{H}_{yx}$ are the auto-covariance and cross-covariance matrices of data channels x and y.[5] The singular value decomposition (SVD) of the coherence matrix C may then be written as Equation 2, where $R^{-1/2}_{xx} R_{xx} R^{-H/2}_{xx} = I$ and $R^{1/2}_{xx} R^{H/2}_{xx} = R_{xx}$.[5] Note that for this SVD, only the nonzero singular values of C and their corresponding nonzero vectors are considered.

The canonical coordinates of x and y can now be defined as Equation 3, where elements of the u and v vectors
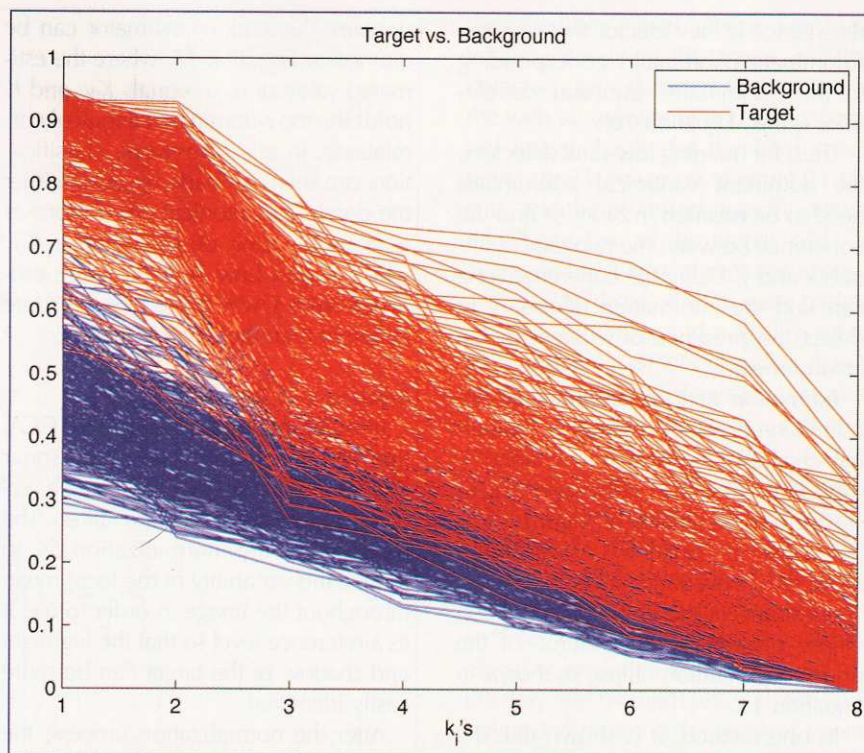
(Above) Plot of canonical correlations for targets and backgrounds.
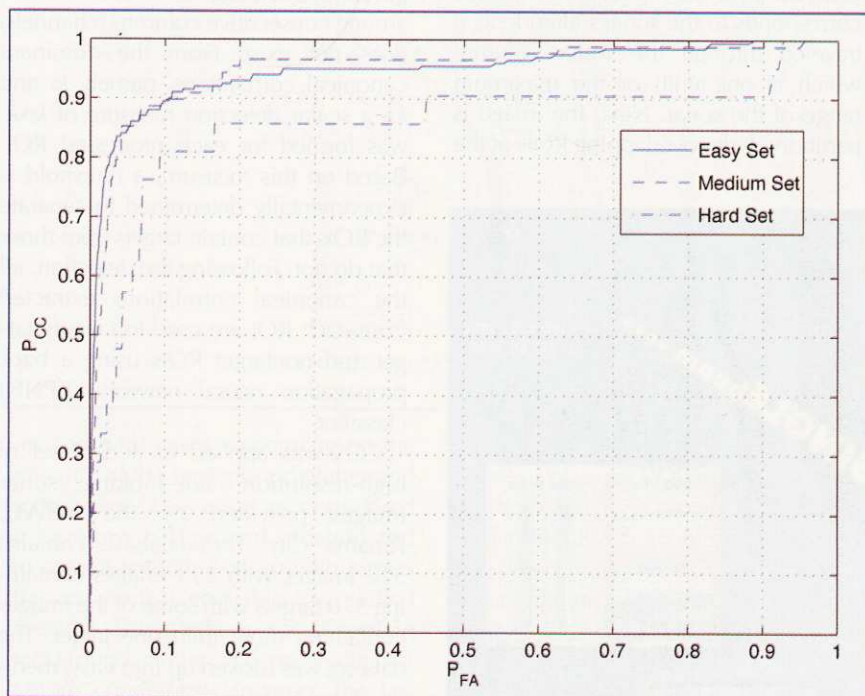
(Left) ROC curves for BPNN classifier in easy, medium and hard background cases.

are the canonical coordinates of x and y, respectively.[5]

Hence, x and y are mapped to their respective canonical coordinates using Equation 4, with $W^{H} = F^{H} R^{-1/2}_{xx}$ and $D^{H} = G^{H} R^{-1/2}_{yy}$. The canonical coordinates u and v share the composite covariance matrix shown in Equation 5, where the diagonal cross-covariance matrix shown in Equation 6 is the canonical correlation matrix of canonical correlations k_i; $i = 1, 2, \dots, m$, where $1 \geq k_1 \geq k_2 \geq \dots \geq k_m > 0$.[5] Note that $E[u_i v_i] = k_i \delta(i-j)$,

where $\delta(\cdot)$ represents the Kronecker delta function and i and j are indexes.

One of the most important properties of canonical correlations is that they are invariant under uncoupled nonsingular transformations of x and y.[5] Canonical correlations can also be used to determine useful properties regarding x and y, such as linear dependence (coherence) and mutual information.[5]

Detection in CCA Framework. The detection problem is stated as testing between two hypotheses of $H_0 : y : CN_n[0,$

$R_{nn}]$, (i.e., noise alone) versus $H_1 : y : CN_n[0, R_{yy} = R_{xx} + R_{nn}]$ (i.e. signal plus noise), where $CN_n[0, R_{xx}]$ denotes the n-variate proper complex normal distribution with zero-mean vector and covariance matrix R_{xx}. In the method depicted in this article, the Neyman-Pearson detector for testing H_0 and H_1 is formulated in the CCA framework, where the log-likelihood ratio is expressed in terms of canonical coordinates and correlations, as shown in Equation 7.

This is the standard Gauss-Gauss log-likelihood ratio, but in the canonical coordinates $v = G^{H} R^{-1/2}_{yy} y$ and $l(y)$ is the weighted sum of the magnitude-squared of the canonical coordinates weighted by canonical correlation-dependent weights. If G is defined as $G = [g_1, \dots, g_n]$, then $l(y)$ can be written as is shown in Equation 8.

The Jeffrey-divergence (j-divergence), or detectability measure, between H_1 and H_0 is found to be Equation 9, where EH_0 and EH_1 are the expected values of $l(y)$ under H_0 and H_1, respectively. The function $(k_i - 1/k_i)^2$ is non-increasing in the interval $(0, 1)$. Consequently, the rank (r) detector that maximizes the

divergence is the detector that uses the r-dominant coordinates corresponding to the r-dominant canonical correlations k (i.e., Equation 10).

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 x and y. Using the coherence, one can find the information necessary to detect the presence of a target in the environment.

Estimation in CCA Framework. The estimation problem is stated as estimating channel x (signal) from channel y (observation). The minimum mean square error estimator of x from y can be written as Equation 11, with minimum error covariance as Equation 12. The volume of the error concentration ellipse divided by the volume of the prior concentration ellipse is shown in Equation 13.

In one method, it is shown that the optimal rank (r) is less than or equal to n estimator of x_e (estimated value of x), of x, from y, is the one that retains only the dominant canonical coordinates and maximizes the coherence between x and y. In canonical coordinate domain, this rank (r) estimator can be defined as Equation 14, where the estimated value of u, u_e, equals $K_r v$, and 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 by using $u - K_r v$, which shows how well x can be estimated from y where the target is more coherent than the environment.

Method and Results

In order to prepare the data for CCA, the high-resolution side scan sonar images are first normalized using a serpentine forward-backward filter. 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.

After the normalization process, the first 120 pixels are ignored. This value corresponds to the sonar's altitude as it traveled through the water column, which is one-tenth of the maximum range of the sonar. Next, the image is partitioned into overlapping ROIs of the size $M \times N$. For this data set, the ROI size of 12 by 34 pixels was experimentally determined to be optimal, considering the average size of the targets in the data set. The overlap along the horizontal and vertical directions was 50 percent in order to ensure that a target would be covered by more than one ROI. Each ROI is then channelized in a column-wise fashion for CCA. The x and y channels consist of the first eight pixels in one column, x, and the first eight pixels in the adjacent column, y. This process is continued, moving in the horizontal direction across the ROI. On the next pass through the ROI, the channels were given a 50 percent overlap in the vertical direction to ensure complete coverage of the target in the ROI. 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. Clearly, for background ROIs, a high level of coherence among consecutive columns (channels) does not exist. 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, a threshold is experimentally determined to separate the ROIs 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 nontarget ROIs using a back propagation neural network (BPNN) classifier.

CCA was applied to a data set of high-resolution side-looking sonar images provided by the NSWC, Panama City. 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, medi-

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

where $E[\cdot]$ is the expectation operation and $(\cdot)^H$ is the Hermitian operator or complex transpose.

$$C = R_{xx}^{-1/2} R_{xy} R_{yy}^{-H/2} = FKG^H \quad \text{and} \quad F^H CG = K. \tag{2}$$

where F, G are orthogonal matrices, i.e. $F^H F = I$ and $G^H G = I$ and I is the Identity matrix. K is a diagonal matrix containing the singular values of C, i.e. $K = \text{diag}[k_1, k_2, \ldots, k_m]$

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

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

where $W^H = F^H R_{xx}^{-1/2}$ and $D^H = G^J R_{yy}^{-1/2}$

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

$$K = R_{uv} = F^H CG = W^H R_{xy} D \tag{6}$$

$$l(\mathbf{y}) = (G^H R_{yy}^{-1/2}\mathbf{y})^H([I - K^2]^{-1} - I)(G^H R_{yy}^{-1/2}\mathbf{y}). \tag{7}$$

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

$$J = E_{H_1} l(\mathbf{y}) - E_{H_0} l(\mathbf{y}) = tr(CC^H + (CC^H)^{-1} - 2I)$$
$$= tr(KK^H + (KK^H)^{-1} - 2I) = \sum_{i=1}^{m^2}(k_i - \tfrac{1}{k})^2 \tag{9}$$

$$J_r = \sum_{i=1}^{r^2}(k_i - \tfrac{1}{k_i})^2 \tag{10}$$

$$\mathbf{x}_e = R_{xx}^{-1/2} FKG^H R_{yy}^{-1/2}\mathbf{y}. \tag{11}$$

$$Q_{xx} = R_{xx}^{-1/2} F(I - K^2) F^H R_{xx}^{H/2}. \tag{12}$$

$$V = \frac{\det(Q_{xx})}{\det(R_{xx})} = \det(I - K^2). \tag{13}$$

$$\mathbf{x}_e = R_{xx}^{-1/2} FK_r G^H R_{yy}^{-1/2}\mathbf{y} = W\mathbf{u}_e. \tag{14}$$

um and hard cases, depending on the difficulty of the background clutter and bottom types. Easy cases are considered to have low background variation and an overall smooth bottom with targets that are easily identifiable to 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.

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

There was good separation, especially for dominant canonical correlations, between targets and background. Using the dominant canonical correlations, k_1 and k_2, the scalar decision measure of $k_1 \times k_2$ can be formed and analyzed for the entire target set and a random set of backgrounds. From this set of targets and the limited number of backgrounds, the optimal threshold value was chosen to be 0.3.

Each image in the entire NSWC database was then blocked into 12 by 34 ROIs. For each ROI, the canonical correlations were formed and the scalar detection measure $k_1 \times k_2$ was compared against the threshold. For the easy cases, there were 186 images containing 201 targets. The detector detected all but one of the targets, with an average of 116 false alarms per image. For

the medium cases, there were 86 images containing 89 targets, and the detector again detected all but one of the with an average of 200 false alarms per image. Lastly, for the hard cases, there are 21 images containing 21 targets, and the proposed detector detected all but two of the targets with an average of 213 false alarms per image.

Then, the detected ROIs were broken up into a training and a testing set to train and test a BPNN classifier. A network structure was determined experimentally, and the structure that performed the best was a two-layer network with eight inputs, 20 neurons in the first hidden layer and two output neurons. The trained classifier was then applied to all the detected ROIs in the easy, medium and hard data sets. The receiver operator characteristic (ROC) classifier performs extremely well on the easy and medium sets with the knee points (the point on the receiver operator characteristics curve) where $Pcc + Pfa = 1$ at $Pcc = 90$ percent and $Pfa = 10$ percent for both cases (Pcc is probability of correct classification and Pfa is probability of false alarm). The classifier did not perform as well on the hard data set, with only an 84 percent correct classification rate and a 16 percent false alarm rate at the knee point of the ROC. Nonetheless, considering the difficulty of the bottom conditions and background clutter, as well as the simplicity of the overall detection and classification system, the classification rates are indeed impressive.

Conclusions and Observations

This article presented a CCA-based Neyman-Pearson detector for detection of underwater targets in high-resolution side-looking sonar imagery. The method exploits coherence properties in a ROI of a sonar image to detect the presence of an object. The dominant canonical correlations that carry this coherence information are used for definition, a simple measure for detection of targets in different background conditions. The extracted features' canonical correlations are subsequently used to classify the detected ROIs without the need to perform a separate feature extraction for classification. The experiments on the side scan sonar data sets with varying density of background clutter demonstrated good separability of the dominant canonical correlations extracted over ROIs containing targets

from those extracted from background only. Overall, CCA did well in detecting all of the targets in all of the images, missing only four of the possible 310 targets in the entire set, while keeping the probability of false alarms low.

Classification using canonical correlation features gave good results given the limited size of the training set. The results showed an average rate of correct classification of 88 percent with an average false alarm rate of 12 percent. These results attest to the great promise of CCA-based underwater target detection and classification from sonar imagery.

References

1. Dobeck, G.J., J. Hyland and L. Smedley, "Automated Detection/Classification of Sea Mines in Sonar Imagery," *SPIE Conference Proceedings*, vol. 3,079, pp. 90–110, 1997.
2. Dobeck, G.J., "Fusing Sonar Images for Mine Detection and Classification," *SPIE Conference Proceedings*, vol. 3,710, pp. 602–614, 1999.
3. Aridges, T., P. Libera, M. Fernandez and G.J. Dobeck, "Adaptive Filter/Feature Orthogonalization Processing String for Optimal LLRT [Log-Likelihood Ratio Test] Mine Classification in Side Scan Sonar Imagery," *SPIE Conference Proceedings*, vol. 2,765, pp. 110–121, 1996.
4. Pezeshki, A., M. Azimi-Sadjadi, L. Scharf and M. Robinson, "Underwater Target Classification Using Canonical Correlations," *Marine Technology Society/IEEE Oceans 2003 Conference Proceedings*, vol. 4, pp. 1,906–1,911, 2003.
5. Scharf, L., 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, 2000.

For a more complete list of references, please contact J. Derek Tucker at dtucker@engr.colostate.edu. ∎

Visit our Web site at www.sea-technology.com, and click on the title of this article in the Table of Contents to be linked to the respective company's Web site.

J. Derek Tucker received his B.S. degree in electrical engineering from Colorado State University. Currently, he is pursuing an M.S. degree. His research interests include digital signal/image processing, target detection, classification, neural networks and feature extraction methods.
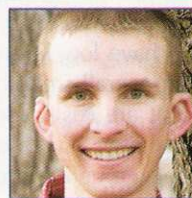
Mahmood R. Azimi-Sadjadi received his Ph.D. in electrical engineering from the Imperial College of Science and Technology at the University of London. He is currently a professor in the electrical and computer engineering department at Colorado State University.