

Underwater Unexploded Ordnance (UXO) Classification Using a Matched Subspace Classifier With Adaptive Dictionaries

John J. Hall , Member, IEEE, Mahmood R. Azimi-Sadjadi , Life Member, IEEE, Steven G. Kargl, Yinghui Zhao, and Kevin L. Williams

Abstract—This paper is concerned with the development of a system for the discrimination of military munitions and unexploded ordnance (UXO) in shallow underwater environments. Acoustic color features corresponding to calibrated target strength as a function of frequency and look angle are generated from the raw sonar returns for munition characterization. A matched subspace classifier (MSC) is designed to discriminate between different classes of detected contacts based upon the spectral content of the sonar backscatter. The system is exclusively trained using model-generated sonar data and then tested using the measured Target and Reverberation Experiment 2013 (TREX13) data sets collected from a synthetic aperture sonar system in a relatively low-clutter environment. A new *in situ* supervised learning method is also developed to incrementally train the MSC using a limited number of labeled samples drawn from the TREX13 data sets. The classification results of the MSC are presented using standard performance metrics, such as receiver operating characteristic curve and confusion matrices.

Index Terms—Incremental learning, K-SVD dictionary learning, matched subspace classifier (MSC), sparse coding, synthetic aperture sonar, unexploded ordnance classification.

I. INTRODUCTION

CLEARING of underwater sites that are contaminated with munitions, as a result of past military training and weapons testing activities, is a challenging task undertaken by many private and government supported groups. Among factors that complicate automatic detection and classification of these targets are the following:

- 1) highly variable operating and environmental conditions (e.g., lakes, ponds, rivers, gulf, or open ocean);
- 2) variations in target features as a function of target size and composition, as well as range,

Manuscript received February 8, 2016; revised April 7, 2017, November 21, 2017, and April 7, 2018; accepted May 3, 2018. Date of publication June 26, 2018; date of current version July 12, 2019. This work was supported by the Strategic Environmental Research and Development Program under Contracts W912HQ-14-C-0007 and W912HQ-17-C-0002. (Corresponding author: John Jack Hall.)

Associate Editor: J. Cobb.

J. J. Hall, M. R. Azimi-Sadjadi, and Y. Zhao are with the Department of Electrical and Computer Engineering, Colorado State University, Fort Collins, CO 80523 USA (e-mail: halljj2@engr.colostate.edu; azimi@engr.colostate.edu).

S. G. Kargl and K. L. Williams are with the Applied Physics Laboratory, University of Washington, Seattle, WA 98105 USA (e-mail: kargl@uw.edu; williams@apl.washington.edu).

Digital Object Identifier 10.1109/JOE.2018.2835538

grazing angle, and orientation with respect to the sonar platform;

- 3) targets may be partially or fully buried, or obscured by marine growth and vegetation.

Various methods have been developed for modeling the acoustic response of objects with geometries typically observed in munitions and unexploded ordnance (UXO) for the purpose of target classification [1]–[6]. The work by Zampolli [1] is concerned with backscatter modeling for spherical and cylindrical targets insonified by an incident signal in the range of 0.1–10 kHz. They used three different modeling methods, namely finite element (FE), boundary element, and semianalytical, and compared their abilities in capturing geometrical and structural-acoustic features of the targets. The FE method was shown to capture the complex internal structures better than the others. Dey *et al.* [2] introduced an FE-based 3-D time-harmonic model for capturing structural-acoustic features for both free-field and littoral environments. Numerical verifications were conducted using the experimental data for scattering from elastic objects. The role of volumetric acoustical imaging was examined in [3] for extracting scattering components attributed not only to the geometric and structural scattering but also to the sediment-water interface of buried UXO and non-UXO objects. They also illustrated good agreement between the measured data and those generated using an FE-based model, and more specifically, the ability to capture the salient geometric and elastic scattering features of the actual measured data. In [5], Kargl *et al.* used model data for shallow grazing angle situations to generate synthetic aperture sonar (SAS) images of simple buried targets and compared the results with those of the measured data. They also combined models for reverberation, acoustic penetration, and target scattering into a unified model. This is then used to generate pings suitable for SAS simulations over a wide range of environmental and operational conditions using computationally efficient FE modeling. Experimentally measured target scattering from proud and buried targets are then used to validate the model through simulations. In [6], Kargl *et al.* conducted monostatic and bistatic acoustic scattering measurements to investigate discrimination capabilities based on the acoustic response of underwater UXO. Results from this study showed that it is possible to use the acoustic templates generated using SAS processing techniques as acoustic color features [4]–[8] to classify a given target. Finally, Tucker *et al.* [8]

developed a method to remedy compositional noise in acoustic color data due to aspect angle uncertainties that occur during the data collection.

Considerable efforts have also been made to develop different feature extraction and classification methods for underwater UXO discrimination, which are presented in [9]–[12]. In [9], Bucaro *et al.* used an FE-based structural-acoustic model to simulate sonar returns from buried UXO at various orientations. A relevance vector machine (RVM) classifier [13] was then trained using these model-generated data. The trained system was shown to be successful in classifying different buried object types using their actual returns. Fischell and Schmidt [14] used 3-D bistatic scattered field data from spherical and cylindrical targets that were collected by an autonomous underwater vehicle in conjunction with a support vector machine (SVM) [15] to perform this two-class classification task. They used both experimental and simulated data to evaluate classifier performance. In [10], Wachowski and Azimi-Sadjadi developed new coherent-based feature extraction and SAS-like acoustic color for detection and classification of underwater UXO and non-UXO objects using the canonical coordinate analysis framework [16]. New multispect classification algorithms were also developed in [11] and [12] using a hidden Markov model and a collaborative multispect classifier to improve classification accuracy while reducing the false alarms.

The specific objective of this paper is the development and testing of an efficient method for the classification of military munitions in shallow underwater environments using sonar returns collected from an SAS system. Similar to the work in [9], we test the hypothesis that model-generated sonar data of various underwater objects can be used to train a classifier. The motivation here is related to the fact that collecting real UXO data in realistic settings is a difficult and an impractical task. The acoustic color data are synthetically generated via a fast ray model [17]. The matched subspace classifier (MSC) in [18] is then used to discriminate underwater munitions from other objects using the Target and REverberation EXperiment 2013 (TRESX13) data sets [19] collected for various target ranges and orientations to determine the generalization performance of the classifier. TRESX13 experiments were designed to not only measure reverberation in shallow water environments but also test for detection and classification of proud and buried UXO. Using an MSC, the class of an unknown object is determined by projecting its acoustic color feature vectors into two subspaces associated with the UXO and non-UXO classes and selecting the class corresponding to the subspace that contains the largest energy. In this paper, the UXO and non-UXO subspaces are constructed using a signal-specific subspace construction method, namely the K-SVD [20] method. The MSC, when employed in conjunction with the K-SVD subspace (or dictionary) learning, can be adapted *in situ* using a limited set of data collected in a new operating environment. To this end, the incremental K-SVD (IK-SVD) method presented in [21] is utilized to *in situ* update the MSC using a small randomly drawn subset of TRESX13 data set without the need to carry the previous training data sets. These important features are crucial if the system is to be used for actual UXO-hunting applications. The results of the proposed system were benchmarked against those

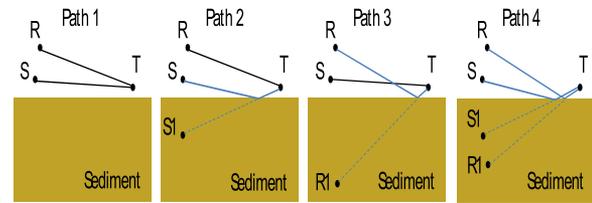


Fig. 1. Path 1 is a direct path. Paths 2 and 3 interact with the sediment once and scatter from the target in a bistatic direction. Path 4 is a backscattering path with two bottom interactions.

of an RVM classifier [13] trained and tested under the same conditions.

This paper is organized as follows. In Section II, we give a brief review of the fast ray model [17] used to generate synthetic data sets for training the MSC system. In Section III, we describe the methods used to generate the acoustic color data for the model-generated training data set as well as for the measured TRESX13 testing data sets. Section IV briefly describes the theory behind the MSC classification system. Section V gives the approaches chosen for subspace construction and its *in situ* adaptation using the IK-SVD method. In Section VI, we present the results of the MSC when tested on TRESX13 data sets. Finally, conclusions on the performance of the system and ideas for future work are discussed in Section VII.

II. FAST RAY MODELING OF THE ACOUSTIC RESPONSE FROM MUNITIONS

To construct template signals needed to reliably represent various UXO in our classification system, the work in [4], [17], [19], [22], and [23] on fast ray modeling of scattering from objects at a water–sediment interface has been utilized. The scattering model allows for monostatic SAS data sets to be simulated via a fast ray model that combines an acoustic ray approximation for propagation in a fluid-filled half-space with scattering from a target in a number of conditions and media. Kargl *et al.* [17] demonstrated that the fast ray model maintained a high-degree of fidelity when compared to FE model results. The fast ray model has the additional benefit that the large data sets required for dictionary construction are easily generated. In this section, we briefly discuss this scattering model as well as its utility in the development of the MSC.

Under typical operation for a short-range SAS platform, air–water scattering paths can be ignored because paths that interact with the air–water interface are either removed by time gating the received signals or are naturally suppressed by the directivity of the source and receiver. In addition, the separation distance between the actual source and receiver is much smaller than the distance to the target, so the source and receiver can be considered to be co-located. Under these conditions, only the four ray paths shown in Fig. 1 contribute to the scattered pressure. The actual source, receiver, and target are denoted by S , R , and T , whereas $S1$ and $R1$ are an image source and receiver. In this model, the source, receiver, and target are located at \mathbf{r}_s , \mathbf{r}_r , and \mathbf{r}_t , and an image source is located at \mathbf{r}_{s_i} with an image receiver at \mathbf{r}_{r_i} . To distinguish path 2 and path 3, the source and receiver are shown at distinct locations; and with our assumption of co-located source and receiver, paths 2 and 3 are reciprocal

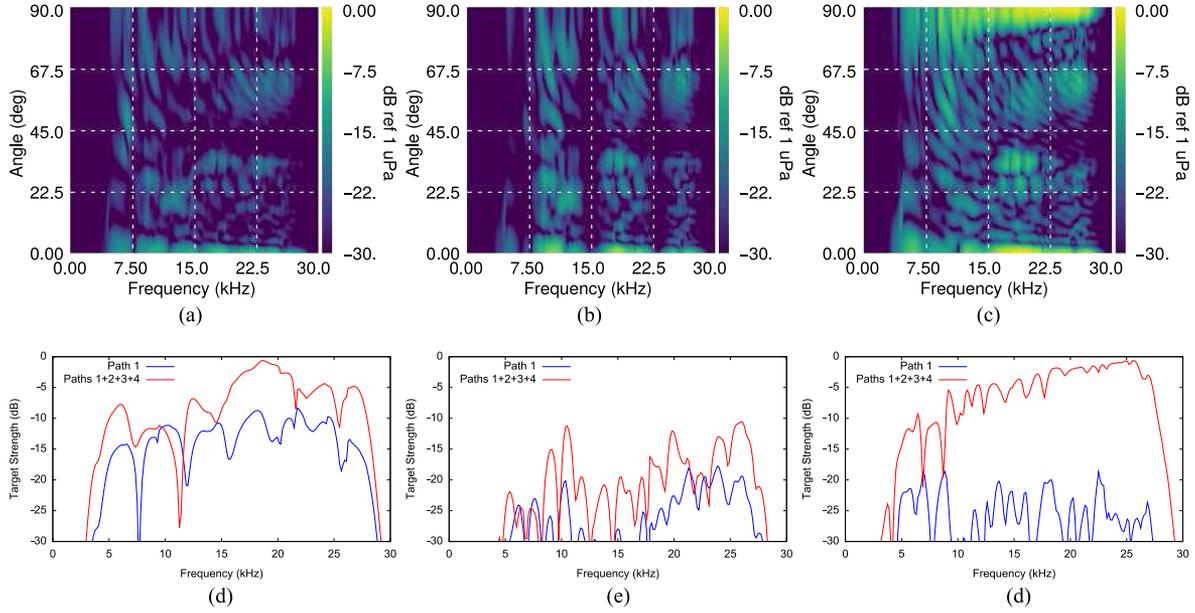


Fig. 2. Acoustic color templates, obtained from fast ray model simulations, are depicted for a 2:1 solid aluminum cylinder with (a) only the first term in (2), (b) the first and last terms in (2), and (c) all terms. (d)–(f) Comparisons of the target strength (TS) for a target in the free field (blue) and proud on the water–sediment interface (red) are shown for the indicated target aspect angles. (a) Path 1 only. (b) Paths 1 and 4. (c) Paths 1–4. (d) Target strength at 0° (broadside). (e) Target strength at 60° . (f) Target strength at 90° (end-on).

and paths 1 and 4 are backscattered. With the specification of a source and a receiver, the scattering from a target has been reduced to a superposition of four free-field scattering problems. Under operational conditions, the distance associated with each path satisfies $d \gg \lambda$, where λ is the wavelength of the pressure wave. The scattered steady-state pressure can then be written as

$$p_s = p_0 A(\mathbf{k}_s, \mathbf{k}_i, \omega) \frac{\exp(ikr)}{r} \quad (1)$$

where p_0 is the amplitude of an incident pressure wave, r is the range from a field point to the target, $A(\mathbf{k}_s, \mathbf{k}_i, \omega)$ is a scattering amplitude, $\exp(ikr)/r$ is a spherically diverging wave, $k = 2\pi/\lambda$ is the wave number, ω is an angular frequency, and \mathbf{k}_i and \mathbf{k}_s are the unit vectors associated with the direction of the incident and scattered fields, respectively. The scattering amplitude contains useful information concerning the material properties of the target and the directionality of the scattered field. The scattering amplitudes can be determined from analytic solutions to scattering problems (e.g., scattering from a spherical target), direct measurements from actual targets, or numerical simulations (e.g., an FE model for a given target).

Combining the ray model paradigm with free-field scattering as given in (1), the spectrum of the total scattered pressure can be written as follows:

$$P(\omega) = \left(\frac{A_1(\omega)}{d_1 d_2} e^{i\omega t_1} + \frac{V(\theta_g) A_2(\omega)}{d_2 d_3} e^{i\omega t_2} + \frac{V(\theta_g) A_3(\omega)}{d_1 d_4} e^{i\omega t_3} + \frac{V^2(\theta_g) A_4(\omega)}{d_3 d_4} e^{i\omega t_4} \right) \times r_0 P_{\text{src}}(\omega) \quad (2)$$

with $d_1 = |\mathbf{r}_s - \mathbf{r}_t|$, $d_2 = |\mathbf{r}_t - \mathbf{r}_r|$, $d_3 = |\mathbf{r}_{s_i} - \mathbf{r}_t|$, and $d_4 = |\mathbf{r}_t - \mathbf{r}_{r_i}|$. The time delays are then $t_1 = (d_1 + d_2)/c_1$, $t_2 =$

$(d_2 + d_3)/c_1$, $t_3 = (d_1 + d_4)/c_1$, and $t_4 = (d_3 + d_4)/c_1$; with c_1 being the speed of sound in water. The pressure spectrum $P_{\text{src}}(\omega)$ represents the frequency spectrum of the transmitted wave packet from the source and $r_0 = 1$ m is a reference distance. The scattering amplitudes $A_i(\omega)$ in (2) also depend on polar and azimuthal angles associated with the sources and receivers in a target-centered geometry. Note that the indices of A_i correspond to the path enumeration depicted in Fig. 1. The reflection coefficient at the water–sediment interface $V(\theta_g)$ is a function of the grazing angle θ_g , and is defined as follows:

$$V(\theta_g) = \frac{\rho \sin(\theta_g) - (\kappa^2 - \cos^2(\theta_g))^{1/2}}{\rho \sin(\theta_g) + (\kappa^2 - \cos^2(\theta_g))^{1/2}} \quad (3)$$

where $\rho = \rho_2/\rho_1$ and $\kappa = (1 + i\delta)/\nu$ with $\nu = c_2/c_1$. Here, ρ_1 is the density of water. The density, sound speed, and loss parameter for the sediment are ρ_2 , c_2 , and δ . An inverse Fourier transform of $P(\omega)$ thus gives a generated sonar signal that includes the four primary acoustical paths for a target near an interface.

Kargl *et al.* [17] discuss the importance of (2) in modeling the target-in-the-environment response. Although readers are referred to [17] for the details, here we have considered the scattering from a 2:1 solid aluminum cylinder resting on a water–sediment interface to illustrate results in this reference. The aluminum cylinder is placed at a 12-m horizontal range from the co-located source and receiver, which are elevated 3.8 m above the interface. The environmental properties are the same as those of TREX13: $\rho_1 = 1000$ kg/m³, $c_1 = 1530$ m/s, $\rho_2 = 2000$ kg/m³, $c_2 = 1694$ m/s, and $\delta = 0.008$. Fig. 2 depicts the predicted absolute TS for this target. The free-field TS as given by (1) or by including only the first term in (2) is shown in Fig. 2(a). In (2), the first and fourth terms are backscatter-

ing paths, and these paths can destructively interfere. Fig. 2(b) clearly shows the destructive interference of paths 1 and 4 across all aspect angles near $f \approx 7.5, 15.0,$ and 22.5 kHz. The superposition of all terms in (2) is depicted in Fig. 2(c). The second and third terms (i.e., paths 2 and 3) are associated with bistatic scattering, and as noted, these paths are reciprocal for the co-located source and receiver. That is, these paths constructively interfere and are the dominant contributions to the TS when the source and receiver are in a nearly end-on orientation with a target. In this orientation, for a finite cylinder, the flat end of the target forms a corner reflector with the water–sediment interface. The significance of the paths 2 and 3 contributions is easily seen by comparing Fig. 2(a) and (c). Fig. 2(d)–(f) displays the TS at target aspect angles of $0^\circ, 60^\circ,$ and 90° . For 0° and 60° , the TS for a proud target exceeds the TS for the target in free field by 5–10 dB over a significant portion of the displayed frequency range. For a target aspect angle of 90° , Fig. 2(f) shows an excess of 20–25 dB over nearly the entire frequency range. These results show the importance of including all four paths in the fast ray model of (2) to better approximate the behavior at all aspects when the target is proud on the bottom.

III. ACOUSTIC COLOR DATA GENERATION

The raw sonar returns generated by the fast ray model are processed to produce acoustic color features for training the MSC. Generation of acoustic color features amounts to forming the intensity of the returned spectral power over the entire range of aspect angles that are modeled in either linear path SAS (LSAS) or circular path SAS runs. This is accomplished by the following procedure: First, an FE model [23] is implemented to produce scattering amplitude information for the intended target. These scattering amplitudes are modeled for acoustic transmissions and returns in the frequency range of 3–30 kHz, which corresponds to the frequency range of the linear frequency-modulated (LFM) transmit signal used in the TREX13 experiments. Next, the half-space model including the four described ray paths in (2) is utilized to generate a raw sonar return data set by generating the modeled returns of a target using the inverse fast Fourier transform (FFT) of (2) over a pregenerated coordinate set representing the various positions along a linear path making soundings. The actual transmitted LFM was used as the source signal in the fast ray model. Next, these raw soundings are matched filtered (or pulse-compressed) with the transmit signal. The FFT is taken of the pulse-compressed data and the result is windowed to 3–30 kHz to remove the unused frequency portions and isolate the frequency range of interest. The mapping of the frequency-ping data to frequency aspect in acoustic color plot is accomplished by computing half of the total observable aspect range (in degrees) $\phi_h = (180/\pi) \times \tan^{-1}(\text{half path length/object range})$ and then linearly interpolating angles to $\phi_c \pm \phi_h$, where ϕ_c is the aspect known to be at the center of the path for a particular target orientation. The major benefit of utilizing the ray model developed in [17]–[23] is that after free-field scattering amplitudes for a desired object are collected or modeled via FE methods, the regeneration of the ray model simulating various aspects and

orientations is by far simpler and faster than rerunning these variations with the FE method [23].

Since an arbitrary trajectory of an SAS platform can be modeled with the fast ray model [17], here the model was used to generate synthetic LSAS sonar data with the same conditions used in TREX13 data collection where a target was located along the perpendicular bisector of the LSAS run. The synthesized data were generated for several object types at a subset of the ranges and with a sonar elevation of 3.8 m as in the TREX13 data set. Using the procedure described above, acoustic color data were generated for several simulated proud objects encountered in the TREX13 data set. Aspect-dependent spectral feature vectors (columns of the acoustic color matrices) at some randomly chosen aspect angles were then used to initialize the dictionary construction process for training of the MSC to be discussed later.

The procedure for generating acoustic color data for the measured TREX13 sonar data is similar to that described for the synthesized sonar data. That is, raw sonar data time series collected from the TREX13 experiment are first pulse-compressed with the LFM transmit signal used in the data collection and the result is further processed to remove returns from the adjacent objects to isolate the object of interest. This processing utilizes a reversible SAS imaging process, a spatial filtering process using a 2-D Tukey window, and a pseudoinverse filtering [24]. This inverse filter maps the SAS image back to the pulse-compressed version that now has less interface scattering noise. These filtered pulse-compressed signals are then transformed to the frequency domain via FFT. This process is repeated for all aspects of LSAS and the power spectrum is then generated and plotted to display acoustic color for each object. Objects used in training and testing included: non-UXO objects, such as cylinders and pipes, and a variety of UXO objects, insonified at various ranges from 10 to 40 m. Examples of the acoustic color matrices for a non-UXO object generated for a 40-m-long simulated LSAS run at a target range of 25 m and with an interface elevation of 3.8 m are given in Fig. 3(a) and (b) for a single path, i.e., using only the first term in (2), and for all paths, i.e., using all four terms in (2), respectively. For this particular run configuration, the target was at 0° orientation (i.e., broadside view of the target). Note that although the maximum angle for this target range is $\pm 38.7^\circ$, the portion that contains nonzero signal is only $\approx \pm 20^\circ$. Fig. 3(c), on the other hand, shows the acoustic color matrix for the same non-UXO object under the same conditions in the experimental TREX13 data set. One can clearly see a closer match between the aspect-dependent spectral features in Fig. 3(b) and the corresponding ones in Fig. 3(c), especially in the frequency range of 10–20 kHz when compared with those of the single path case in Fig. 3(a).

IV. REVIEW OF THE MSC

In this section, we give a brief review of the theory behind the MSC and a simple modification of this classifier, which allows better utilization of a dictionary learned via K-SVD and sparse coding methods. This classifier operates on the aspect-dependent spectral feature vectors described in Section III with

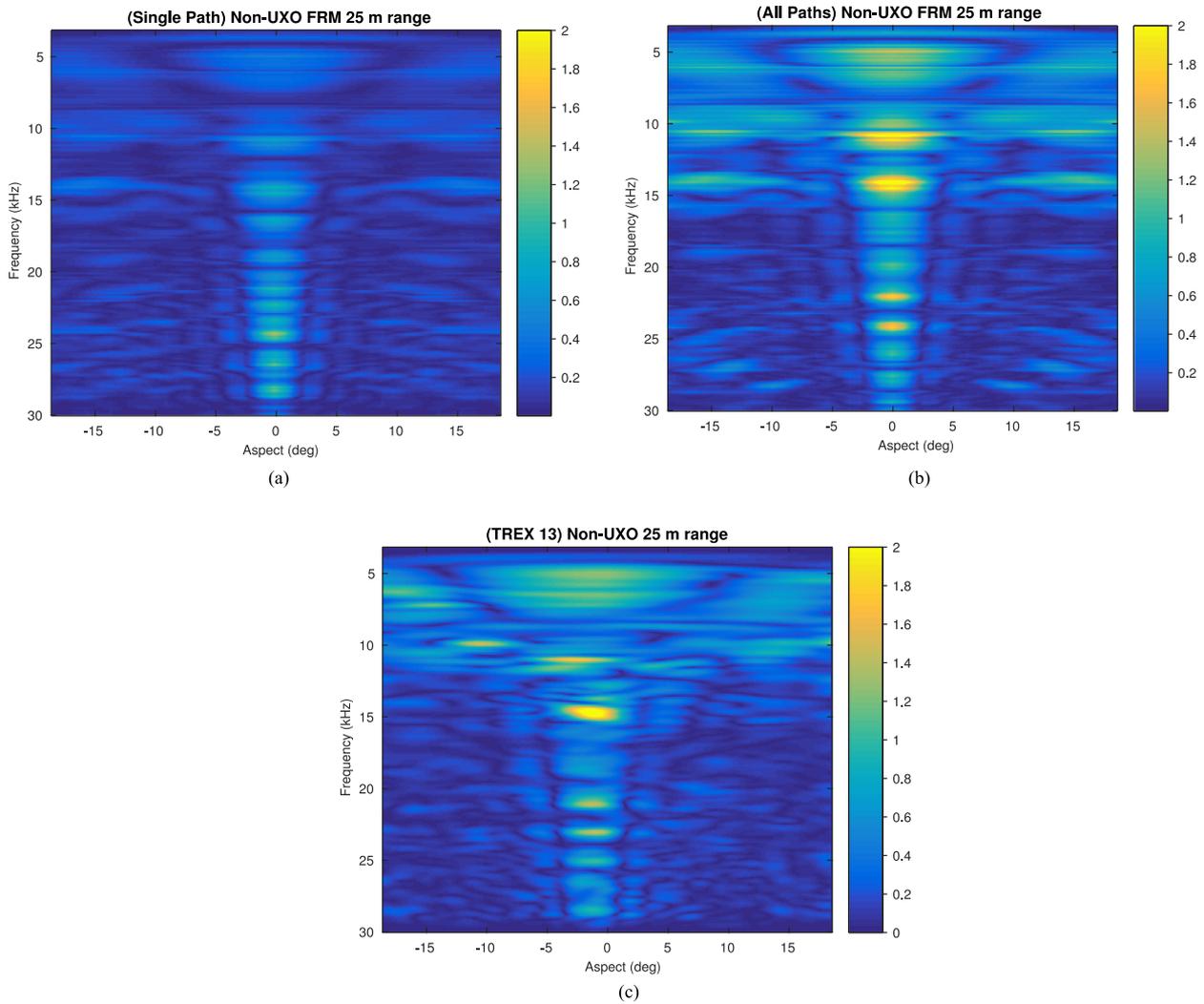


Fig. 3. Acoustic color plots for non-UXO object using LSAS-model-generated versus TREX data. (a) Fast ray model-single path. (b) Fast ray model-all paths. (c) TREX13 data.

the assumption that such vectors, each belonging to a given class of objects, can be accurately represented using a linear combination of a small subset of basis vectors associated with that class.

We consider a general M -ary classification problem in which the observations (feature vectors) can come from $m = 0, \dots, M - 1$ possible classes satisfying the signal model

$$\mathbf{Z} = \mathbf{H}_m \mathbf{X} + \mathbf{N} \quad m \in [0, M - 1] \quad (4)$$

where $\mathbf{Z} = [\mathbf{z}_1 \mathbf{z}_2 \dots \mathbf{z}_K]$ is the observation matrix of size $N \times K$ containing spectral feature vectors $\mathbf{z}_i \in \mathbb{R}^N$, $\mathbf{H}_m \in \mathbb{R}^{N \times L}$ is the dictionary (or subspace) matrix whose columns are the basis vectors \mathbf{h}_j that span the subspace associated with the m th object class, $\mathbf{X} = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_K]$ is an unknown parameter matrix with $\mathbf{x}_i \in \mathbb{R}^L$ being the parameter vector associated with \mathbf{z}_i , and \mathbf{N} denotes an additive zero-mean noise matrix that represents the inaccuracy in this representation. Since the distribution of the noise is typically unknown, it is not possible

to derive the maximum likelihood estimates of the unknown parameters [18] in this case.

The subspace matrix \mathbf{H}_m can be constructed utilizing the training data representing the objects belonging to class m using a subspace reconstruction method, such as K-SVD [20], which will be discussed later. Once these matrices are formed, the decision about the class membership of a particular unknown observation vector (e.g., aspect-dependent spectral feature vector) \mathbf{z}_k can be made using

$$m^* = \arg \min_{m \in [0, M - 1]} \{ \|\mathbf{z}_k - \mathbf{H}_m \hat{\mathbf{x}}_k\|^2 \} \quad (5)$$

where $\hat{\mathbf{x}}_k$ represents an estimate of the actual parameter vector \mathbf{x}_k . This decision rule assumes that the reconstruction error takes its smallest magnitude for the correct class m^* . Thus, $J_m = \|\mathbf{z}_k - \mathbf{H}_m \hat{\mathbf{x}}_k\|^2$ is the discriminant function for this classifier. Now, if the least squares (LS) estimate of $\hat{\mathbf{x}}_k$, i.e.,

$$\hat{\mathbf{x}}_k = (\mathbf{H}_m^T \mathbf{H}_m)^{-1} \mathbf{H}_m^T \mathbf{z}_k \quad (6)$$

is used in (5), this discriminant function becomes $J_m = \mathbf{z}_k^T \mathbf{P}_{\mathbf{H}_m}^\perp \mathbf{z}_k$ where $\mathbf{P}_{\mathbf{H}_m}^\perp = \mathbf{I} - \mathbf{P}_{\mathbf{H}_m}$ with $\mathbf{P}_{\mathbf{H}_m} = \mathbf{H}_m (\mathbf{H}_m^T \mathbf{H}_m)^{-1} \mathbf{H}_m^T$ being the orthogonal projection matrix onto the subspace spanned by the columns of the dictionary matrix \mathbf{H}_m . Thus, for this two-class (UXO versus non-UXO) discrimination problem, the test statistic becomes

$$\Lambda(\mathbf{z}_k) = \frac{J_0}{J_1} = \frac{\mathbf{z}_k^T \mathbf{P}_{\mathbf{H}_0}^\perp \mathbf{z}_k}{\mathbf{z}_k^T \mathbf{P}_{\mathbf{H}_1}^\perp \mathbf{z}_k} \underset{\text{Non-UXO}}{\overset{\text{UXO}}{\gtrless}} \gamma \quad (7)$$

where γ is the decision threshold which is varied to generate the receiver operating characteristic (ROC) curve of the classifier. An important property of this test statistic is its invariance to scaling of the observation vectors, i.e., $\Lambda(a\mathbf{z}_k) = \Lambda(\mathbf{z}_k)$ for any arbitrary scaling factor a .

For making a decision on several (P) aspect-dependent spectral feature vectors forming data matrix $\mathbf{Z} = [\mathbf{z}_1 \mathbf{z}_2 \cdots \mathbf{z}_P]$, the expression in (5) becomes

$$m^* = \arg \min_{m \in [0, M-1]} \{ \|\mathbf{Z} - \mathbf{H}_m \hat{\mathbf{X}}\|_F^2 \} \quad (8)$$

where $\|\mathbf{A}\|_F^2 = \text{tr}\{\mathbf{A}\mathbf{A}^T\}$ represents the squared Frobenius norm of matrix \mathbf{A} . Alternatively, for a two-class decision-making based upon P aspects, we can use

$$\Lambda(\mathbf{Z}) = \frac{\text{tr}\{\mathbf{Z}^T \mathbf{P}_{\mathbf{H}_0}^\perp \mathbf{Z}\}}{\text{tr}\{\mathbf{Z}^T \mathbf{P}_{\mathbf{H}_1}^\perp \mathbf{Z}\}} \underset{\text{Non-UXO}}{\overset{\text{UXO}}{\gtrless}} \gamma. \quad (9)$$

As can be seen from (7) or (9), the MSC measures the energy in each of the subspaces $\langle \mathbf{H}_m \rangle$ and selects the class label corresponding to the subspace that contains the largest amount of energy (or smallest residual).

Training of this classifier amounts to constructing class-dependent dictionary matrices \mathbf{H}_m from representative training data sets in each class m . In this paper, we used K-SVD [20] signal-specific dictionary-learning method using model-generated data sets. This dictionary-learning algorithm uses sparse coding methods, such as orthogonal matching pursuit (OMP) or basis pursuit [25], to reduce the data into sparse vectors as columns of $\hat{\mathbf{X}}$ that contain only a few nonzero elements. Since the estimates generated using these sparse coding methods differ from those of the standard LS solution [26] used to develop MSC decision rule in (8), we use the following modified decision rule

$$m^* = \arg \min_{m \in [0, M-1]} \{ \|\mathbf{Z} - \mathbf{H}_m \hat{\mathbf{X}}_{\text{OMP}}^{(m)}\|_F^2 \} \quad (10)$$

where $\hat{\mathbf{X}}_{\text{OMP}}^{(m)}$ is the estimate of \mathbf{X} generated using the OMP algorithm when dictionary \mathbf{H}_m is used. Note that comparing to the LS estimate $\hat{\mathbf{X}}$, each column of $\hat{\mathbf{X}}_{\text{OMP}}^{(m)}$ contains only a few principal elements while the rest of the elements are zero owing to the sparsity constraint. That is, only a small subset of L dictionary atoms are used to represent each observation. Again, for a two-class discrimination based upon P aspects, one can also use

$$\Lambda(\mathbf{Z}) = \frac{\|\mathbf{Z} - \mathbf{H}_0 \hat{\mathbf{X}}_{\text{OMP}}^{(0)}\|_F^2}{\|\mathbf{Z} - \mathbf{H}_1 \hat{\mathbf{X}}_{\text{OMP}}^{(1)}\|_F^2} \underset{\text{Non-UXO}}{\overset{\text{UXO}}{\gtrless}} \gamma. \quad (11)$$

Aside from its simplicity in structure, training, and decision-making, the proposed classification framework offers many other benefits including: ease in performing M -ary classification, certain invariance properties, and incremental *in situ* training in new environments (see Section V-B) without jeopardizing the prior training, i.e., offers flexibility without sacrificing the stability.

V. SIGNAL SUBSPACE CONSTRUCTION METHODS

In this section, we briefly describe the ideas behind the K-SVD and IK-SVD dictionary-learning methods used in this study.

A. K-SVD Method

The K-SVD subspace construction method [20] was employed to form dictionary matrices \mathbf{H}_m , $m = 0, 1$, for UXO versus non-UXO discrimination using acoustic color data. The purpose of K-SVD is to create a dictionary matrix that can optimally represent a signal vector as a sparse linear combination of relatively few basis vectors or atoms. K-SVD achieves this by solving a constrained optimization problem to minimize the reconstruction error given a set of training vectors. Let $\mathbf{Z}_{m,n} \in \mathbb{R}^{N \times K}$ be a matrix consisting of object n of class m training data vectors $\mathbf{z}_k^{(m,n)}$, $k \in [1, K]$, as its columns, $\mathbf{H}_{m,n} \in \mathbb{R}^{N \times L}$ be the dictionary matrix to be found, and $\mathbf{X}_{m,n} \in \mathbb{R}^{L \times K}$ be the sparse representation of $\mathbf{Z}_{m,n}$ in terms of its dictionary atoms. Dictionary matrix \mathbf{H}_m for class m is then formed by concatenating all matrices $\mathbf{H}_{m,n}$ of those objects belonging to class m . Note that it is desired that the number of nonzero elements of each $\mathbf{x}_k^{(m,n)}$ be less than or equal to $\tau \ll L$ to satisfy the sparsity constraint. The constrained optimization problem [20] for K-SVD is given by

$$\min_{\mathbf{H}_{m,n}, \mathbf{X}_{m,n}} \{ \|\mathbf{Z}_{m,n} - \mathbf{H}_{m,n} \mathbf{X}_{m,n}\|_F^2 \} \text{ s.t. } \|\mathbf{x}_k^{(m,n)}\|_0 \leq \tau \quad \forall k \quad (12)$$

where $\|\cdot\|_0$ denotes l_0 norm, which is the number of nonzero elements of a vector, and τ is the sparsity factor.

During the training, the K-SVD algorithm iterates between two phases. First, a sparse representation phase, where for each $\mathbf{z}_k^{(m,n)}$, the corresponding $\hat{\mathbf{x}}_k^{(m,n)}$ is computed based on a given $\mathbf{H}_{m,n}$ using a pursuit method such as OMP [25]. Second, dictionary update phase, where columns of $\mathbf{H}_{m,n}$ are updated one by one based on minimizing the reconstruction error using the singular value decomposition (SVD) of the restricted error matrix [20]. These two phases are repeated until convergence through monotonic mean-squared error (MSE) reduction. Additionally, we provide a summary of the essential steps in implementing the K-SVD algorithm. For more details, the reader is referred to [20]. For simplicity in notation, in the sequel, we dropped the class and object indices m and n , respectively, since the same sequence of operations is implemented irrespective of m or n .

- a) *Initialization*: Set the initial dictionary matrix \mathbf{H} with L randomly selected normalized columns of \mathbf{Z} . Set iteration index $t = 1$. Iterate the following steps until a stopping criterion is met.

- b) *Sparse coding phase*: Generate $\tilde{\mathbf{X}}$ by computing the sparse representation $\tilde{\mathbf{x}}_k$ for each training observation vector \mathbf{z}_k in training set \mathbf{Z} based on the current dictionary matrix \mathbf{H} obtained at iteration $(t - 1)$ using the OMP or fast OMP method in [27], which does not require any matrix inversion operation.
- c) *Dictionary update*: Each column \mathbf{h}_l , $l \in [1, L]$, in \mathbf{H} is updated individually by the following procedure.

- 1) For atom \mathbf{h}_i , compute the i -exclusive error matrix $\mathbf{E}_i = \mathbf{Z} - \sum_{j \neq i} \mathbf{h}_j \tilde{\mathbf{x}}_T^j$, which represents the error in reconstruction when the contribution of i th atom \mathbf{h}_i is excluded, i.e., $\|\mathbf{Z} - \mathbf{H}\tilde{\mathbf{X}}\|_F^2 = \|\mathbf{E}_i - \mathbf{h}_i \tilde{\mathbf{x}}_T^i\|_F^2$. Here $\tilde{\mathbf{x}}_T^i$ is the i th row vector of matrix $\tilde{\mathbf{X}}$, i.e., all the coefficients corresponding to the i th atom.
- 2) To impose sparsity, define a subset $\omega_i = \{n | \hat{x}_T^i(n) \neq 0\}$ of indices of training samples in \mathbf{Z} that use dictionary atom \mathbf{h}_i in their reconstruction via \mathbf{H} , i.e., those with $\hat{x}_T^i(n) \neq 0$, where $\hat{x}_T^i(n)$ is the n th element of $\tilde{\mathbf{x}}_T^i$.
- 3) Form the restricted error matrix \mathbf{E}_i^R and coefficient vector $\tilde{\mathbf{x}}_R^i$ by selecting columns of \mathbf{E}_i corresponding to those indices in ω_i set and likewise for entries of $\tilde{\mathbf{x}}_T^i$ (i.e., discard zero entries in the row vector). More specifically, $\mathbf{E}_i^R = \mathbf{E}_i \mathbf{\Omega}_i$ and $\tilde{\mathbf{x}}_R^i = \tilde{\mathbf{x}}_T^i \mathbf{\Omega}_i$, where $\mathbf{\Omega}_i$ is a matrix of size $L \times Q$, with Q being the cardinality of set ω_i , whose $(\omega_i(n), n)$ entries are ones and zeroes elsewhere.
- 4) Apply SVD to restricted error matrix $\mathbf{E}_i^R = \mathbf{U}\mathbf{S}\mathbf{V}^T$, the updated dictionary column $\hat{\mathbf{h}}_i = \mathbf{u}_1$, i.e., the first column of \mathbf{U} and the updated coefficient vector $\tilde{\mathbf{x}}_R^i = \mathbf{v}_1 \mathbf{S}_{1,1}$, where $\mathbf{S}_{1,1}$ is the first and largest singular value in the SVD of \mathbf{E}_i^R and \mathbf{v}_1 is the first row of \mathbf{V}^T . These choices of $\tilde{\mathbf{x}}_R^i$ and $\hat{\mathbf{h}}_i$ give the best rank one solution [28] to minimize total error in reconstruction.

Set $t = t + 1$ and repeat b) and c) until stopping criterion is met.

B. IK-SVD Method and In Situ Training

The aspect-dependent spectral features for a specific target vary significantly depending on the object's burial condition and orientation, seafloor properties and roughness, range, and grazing angles with respect to the sonar, etc. Although it is unrealistic to expect model-generated data to capture all such variations for target characterization, training based upon such plentiful data provides us with a classifier that is trained using a "baseline" training set. This baseline-trained classifier can be used to determine how to augment the training data sets using a limited set of sonar returns from actual objects drawn from different environmental conditions to improve the classifier's robustness and generalization. This calls for an *in situ* learning scheme that can be used to update the dictionary matrices without erasing the prior learning.

When the model-generated data will no longer be adequate to represent the data in new environments, the dictionary ma-

trix \mathbf{H} need to be adapted to maintain the classification performance. To accomplish this dictionary updating, we use the IK-SVD algorithm in [21], which can introduce new atoms into the existing dictionary using the collected *in situ* data set $\tilde{\mathbf{Z}} = [\mathbf{z}_{K+1} \cdots \mathbf{z}_{K+\tilde{K}}]$ from the new environment. The objective function for this incremental learning model is given by

$$\min_{\tilde{\mathbf{H}}, \tilde{\mathbf{X}}} \|\tilde{\mathbf{Z}} - \tilde{\mathbf{H}}\tilde{\mathbf{X}}\|_F^2 \quad \text{s.t.} \quad \|\tilde{\mathbf{x}}_k\|_0 \leq \tau \quad \forall k \in [K+1, K+\tilde{K}] \quad (13)$$

where $\tilde{\mathbf{X}}$ is the associated sparse representation of $\tilde{\mathbf{Z}}$ and $\tilde{\mathbf{H}} = [\mathbf{H} \tilde{\mathbf{H}}]$ represents the new dictionary matrix consisting of the old dictionary matrix $\mathbf{H} = [\mathbf{h}_1 \cdots \mathbf{h}_L]$ and the incremental one $\tilde{\mathbf{H}} = [\mathbf{h}_{L+1} \cdots \mathbf{h}_{L+\tilde{L}}]$ with \mathbf{h}_{L+i} , $i = 1, \dots, \tilde{L}$, being the newly added atoms. In this algorithm, the old dictionary \mathbf{H} remains unchanged while only the new atoms are updated by the K-SVD algorithm. Hence, the error matrix for \mathbf{h}_i , $i \in [L+1, L+\tilde{L}]$, is

$$\tilde{\mathbf{E}}_i = \tilde{\mathbf{Z}} - \sum_{j=1}^L \mathbf{h}_j \tilde{\mathbf{x}}_T^j - \sum_{j=L+1, j \neq i}^{L+\tilde{L}} \mathbf{h}_j \tilde{\mathbf{x}}_T^j. \quad (14)$$

This error matrix is used in the original K-SVD algorithm in Section V-A to find the new dictionary atoms.

In [21], an entropy-based criterion is employed to select the initial values of the new atoms by first sparsely coding $\tilde{\mathbf{Z}}$ using the old dictionary matrix \mathbf{H} . Then, the entropy of each sparse coefficient vector in $\tilde{\mathbf{X}}$ is calculated, and the samples with the largest entropy (maximal disagreement) are used to initialize the new atoms. These samples correspond to those that are least sparse and cannot be accurately represented by the old dictionary matrix.

VI. RESULTS AND DISCUSSION

In this section, we provide results of the MSC-based classification system when trained on model-generated sonar data and subsequently tested on the experimental TREX13 data sets for various UXO and non-UXO objects.

A. Synthesized Training Data Set and Procedure

The objective here is to use the model-generated aspect-dependent spectral feature vectors for several UXO and non-UXO objects with known geometrical and physical characteristics at different ranges and orientations to construct dictionary matrices suitable for the MSC-based classification system. The fast-ray model was utilized to create acoustic color matrices for different environments and simulated runs. Using the procedures described in Sections II and III, simulated sonar runs for eight different objects¹ (six UXO and two non-UXO) were generated. These simulated runs were designed to replicate the conditions of the experimental TREX13 data collection. In particular, these runs were generated for a 40 m LSAS path length at ranges of 10, 15, 20, 25, 30, 35, and 40 m to the target and with a sonar interface elevation of 3.8 m. Additionally, object orientations ranged

¹At the time of writing this paper, the fast ray models existed only for these 8 objects out of 25 objects in the TREX13 data set.

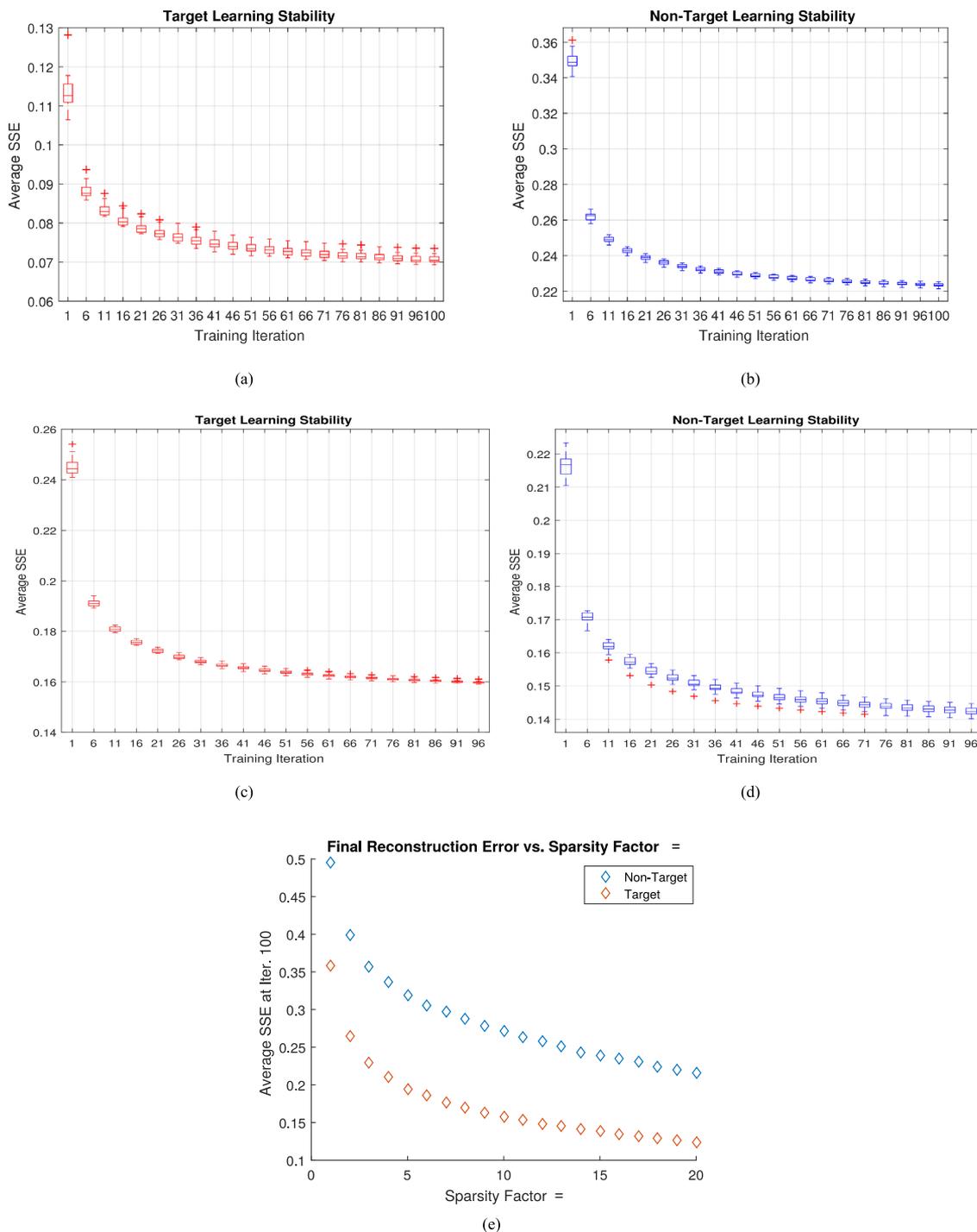


Fig. 4. K-SVD learning and sparsity. (a) Learning curve for a UXO object 4. (b) Learning curve for a non-UXO object 1. (c) Learning curve for a UXO object 8. (d) Learning curve for a non-UXO object 2. (e) Different choices of τ -all objects.

from -80° to $+80^\circ$ in 20° increments, with 0° orientation corresponding to broadside view of the object and 90° corresponding to an end-on view of objects. These synthetic sonar data sets and their corresponding acoustic color matrices were generated for two different environments with sound speeds matching those conditions in TREX13 experiments. In particular, one environ-

ment used the sediment sound speed of sand and the other used that of a slightly denser material simulating a mixture of sand and silt or clay.

For each object, the aspect-dependent spectral features in the associated acoustic color matrix were decimated along the frequency dimension to have $N = 271$ frequency bins spanning

the 3–30kHz frequency range (i.e., 100-Hz separation between frequency bins). However, along the aspect dimension, the original ping separation was used resulting in aspect resolution that changed depending on target range. The training data matrix $\mathbf{Z}_{m,n}$ for object n of class m will then contain all the model-generated aspect-dependent spectral features as its columns for all the above-mentioned ranges and orientations of the UXO and non-UXO objects.

Using this training database, the K-SVD subspace construction method was applied to form dictionary matrices $\mathbf{H}_{m,n}$ for all objects and classes. Since the maximum rank of each data matrix $\mathbf{Z}_{m,n}$ is 271, to initialize the dictionary atoms, we randomly selected $L = 271$ aspect-dependent spectral feature vectors from this matrix. Thus, after K-SVD training is completed, each matrix $\mathbf{H}_{m,n}$ would be of size 271×271 . As mentioned previously, dictionary matrix \mathbf{H}_m for class m is then formed by concatenating all matrices $\mathbf{H}_{m,n}$ of those objects belonging to class m . Fig. 4(a)–(d) demonstrates the K-SVD learning curves (boxplots) for two different UXO and non-UXO objects for fixed sparsity factor of $\tau = 15$ and for 20 random initializations. As can be observed, the learning curves are relatively insensitive to initialization. Additionally, though these learning curves illustrate different learning behavior for these UXO and non-UXO objects, our experiments with all the object types indicated that 100 iterations are generally sufficient to meet an average sum-squared error (SSE) goal of 0.25. Fig. 4(e), on the other hand, shows the plots of the average SSE of the reconstruction error for different values of sparsity factor τ averaged for all UXO and non-UXO objects, respectively. Based upon these results, a sparsity factor of $\tau = 15$ at which the average SSE ≤ 0.25 for the non-UXO class was selected throughout all the experiments.

B. Experimental TREX13 Testing Data Set

Fig. 5 shows a pictorial representation of a TREX13 target field where the targets were in the field of view of the sonar tower. Five additional target field configurations, similar to the one depicted in Fig. 5, were employed during the data collection process. The target field contained several UXO and non-UXO objects with varying shapes, sizes, and compositions, all of which were located between 10 and 40 m horizontally from the rail system and are proud on the bottom. The sonar system was mounted on a mobile tower to minimize platform motion as the sonar tower traversed along the rail. The length of the rail was approximately 40 m. The sonar transmitted a 6-ms LFM signal over 3–30 kHz with a 10% taper between the leading and trailing edges to minimize ringing effects. Sonar backscatter was received by a six-element linear array. For the formation of acoustic color matrices, data from only the third hydrophone element of this array was used.

There were a total of nine runs through the target field in which the physical orientation of all the objects in the scene differed with each run, with each object having the same orientation for a given run. The object orientations varied from -80° to $+80^\circ$ in 20° increments. Each run consisted of 1600 pings in which the sonar platform moved along the fixed rail in increments of 0.025 m, transmitting and receiving once for each sonar position. The data were sampled at 100 kHz and the sonar platform was

tilted at either a 10° or 20° grazing angle depending on range to the targets being observed (angle of the sonar main response axis with respect to the horizontal plane). Acoustic color data were generated using the procedure described in Section III.

For analyzing the performance of the MSC, several experiments were conducted which included observations from eight objects at all ranges and orientations present in the TREX13 data set described previously. The results are given in the next section.

C. Classification Results and Analysis

1) *Two-Class Classification Results*: Using the decision rule given in (11), a class membership (UXO versus non-UXO) was decided for each observation matrix \mathbf{Z} containing several aspect-dependent spectral feature vectors extracted from the acoustic color matrices. Multiaspect classification is more amenable to actual operational situations where underwater objects are typically viewed from multiple aspects. Moreover, it is expected that multiaspect decision-making provides a much better opportunity to discriminate between UXO and non-UXO objects with minimal confusion. This is because discrimination of target types is impossible from some viewing angles though clear from others. Thus, here we conducted an experiment where $P = 7$ number of aspects were used to perform UXO versus non-UXO discrimination. In the TREX13 experiments, aspect separation was approximately uniform for a given target range due to the sonar being on the rail system. Thus, to simulate nonuniform platform motion in realistic search scenarios, aspect separation corresponding to $10 + s$ pings was used, where for every pair of aspect angles, a new realization of discrete random variable s is drawn from a uniform pdf $s \sim \text{unif}[-3, 3]$ representing aspect jitter. Note that this process corresponds to approximately 0.175–0.325 m of platform motion between pings. Additionally, to consider all possible aspect and target orientation combinations, the acoustic color matrices for different orientations were arranged in increasing order starting from -80° to $+80^\circ$ in such a way to mimic a continuous circular path over the object in a particular range. This process is then repeated for each object type at all ranges. In this manner, the starting aspect was varied from 0 to 9 pings for each trial starting with the data at -80° orientation to provide many different aspect combinations.

As a benchmark, an RVM classifier [13] was also trained using the same training data set and the results were compared with those of the MSC-based two-class classifier. An RVM classifier has been previously applied to similar applications in many references including [9] and [22]. The training of the RVM was done using MacKay approximation procedure based upon Laplace's method [29], [30], which maps the classification problem to the regression. In the RVM classifier with logistic sigmoid activation function $g(y) = 1/(1 + e^{-y})$, the Bernoulli conditional distribution $p(t|w) = g(y)^t(1 - g(y))^{1-t}$ is used for the class label (or target value) $t \in \{0, 1\}$. The output takes the form $y = \mathbf{w}^T \mathbf{k}(\mathbf{z})$, where $\mathbf{z} \in \mathbb{R}^N$ is the input vector, $\mathbf{k}(\mathbf{z})$ is a kernel function vector $\mathbf{k}(\mathbf{z}) = [1, k(\mathbf{z}_1, \mathbf{z}), k(\mathbf{z}_2, \mathbf{z}), \dots, k(\mathbf{z}_K, \mathbf{z})]^T$ with $k(\mathbf{z}_i, \mathbf{z})$ being an acceptable kernel function centered at \mathbf{z}_i , and the 1 at the beginning of $\mathbf{k}(\mathbf{z})$ vector accounts for the bias as the first entry of \mathbf{w} vec-

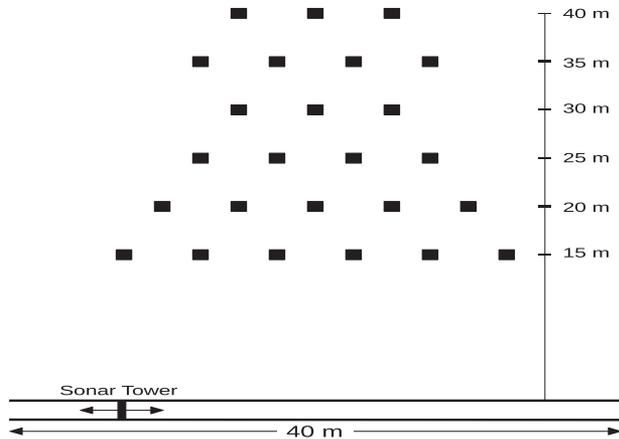


Fig. 5. Schematic of TRESX13 target field. Each black square marks the location of a target placed into the target field by divers.

tor. Furthermore, zero mean normal prior distributions are also assumed for the individual components of the weight vector \mathbf{w} with the corresponding precision hyperparameters (inverse of variance) $\alpha = [\alpha_0 \alpha_1 \dots \alpha_K]^T$ leading to the prior distribution for the weight vector \mathbf{w} in the form

$$p(\mathbf{w}|\alpha) = \prod_{i=0}^K \mathcal{N}_{w_i}(0, \alpha_i^{-1}).$$

Given K independent identically distributed training samples \mathbf{z}_k , learning process in RVM constitutes choosing initial values for the precision hyperparameters α_i and alternately until convergence the re-estimation of the weight vector \mathbf{w} , the posterior covariance matrix, and the hyperparameters. Note that the Laplace's method leads to Gaussian approximation of the posterior distribution $p(\mathbf{w}|\mathbf{t}; \alpha) = p(\mathbf{t}|\mathbf{w})p(\mathbf{w}|\alpha)/p(\mathbf{t}|\alpha)$, where $p(\mathbf{t}|\mathbf{w}) = \prod_{k=1}^K \mathbf{g}(\mathbf{y}_k)^t (\mathbf{1} - \mathbf{g}(\mathbf{y}_k))^{1-t}$ with \mathbf{Z} the training data matrix and \mathbf{t} the associated target (or label) vector. Compared to the MSC-based classifier, which makes no assumption about the data/parameters distributions, the Gaussian approximation in RVM could be problematic particularly when the actual data do not support such an assumption.

Here, we initialized the weight precision hyperparameters $\alpha_i = 0.5, \forall i$. Both linear and Gaussian (smoothing parameter $\sigma = 20$) kernel functions were used. As mentioned previously, during the training RVM, the logistic log-likelihood function is maximized iteratively with respect to \mathbf{w} . The RVM training used about 100 iterations to reach a comparable SSE to that for the MSC-based method on a validation data set. As with the MSC in (10) or (11), the decisions of the RVM classifiers at multiple aspects were fused together using the accumulated log-likelihood ratios at each aspect using the assumption of independence among consecutive observations.

Fig. 6 illustrates the ROC curves of the MSC-based and both RVM classifiers tested on all ranges and orientations of the TRESX13 data set for $P = 7$ aspects. These ROC curves represent the averaged classification rates for ten trials, each corresponding to a different starting aspect as explained earlier. Comparison of these ROC curves illustrates the fact that

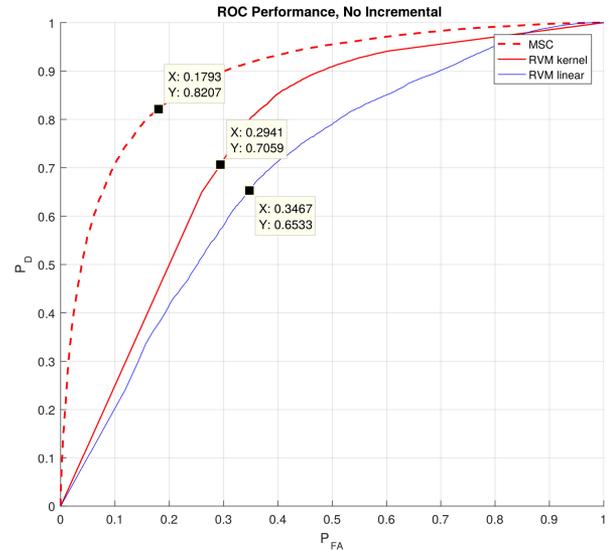


Fig. 6. ROC curves for two-class classifiers—MSC and RVM (linear and nonlinear).

TABLE I
CONFUSION MATRIX FOR TWO-CLASS MSC

Truth \ Dec.	UXO	Non-UXO
UXO	0.716	0.284
Non-UXO	0.106	0.894

TABLE II
CONFUSION MATRIX FOR TWO-CLASS RVM (LINEAR) CLASSIFIER

Truth \ Dec.	UXO	Non-UXO
UXO	0.658	0.342
Non-UXO	0.347	0.653

both RVM classifiers had difficulty in correctly classifying (or generalizing) the experimental TRESX13 data when trained exclusively on model-generated data. The knee points of the ROC (where $P_{cc} + P_{fa} = 1$) curves exhibit approximately $P_{cc} = 82\%$, $P_{cc} = 64\%$, and $P_{cc} = 71\%$ for the MSC, linear RVM, and Gaussian RVM classifiers, respectively. The rather poor performance of the RVM classifier in this particular case may be attributed to: (1) sparsity of the weight solution which in case of training with large amount of synthesized data can result in lack of generalization on the TRESX13 data sets; and (2) the Gaussian approximation of the posterior distribution which in case of mismatch can have a negative effect on the performance. The former implies that the relevance vectors established using exclusively the synthesized data are incapable of representing the TRESX13 data samples. Further investigation of the MSC-based classifier results revealed that misclassifications occurred almost equally for all orientations of a given troublesome object. Tables I–III display the confusion matrices for the MSC and the two RVM classifiers (linear and Gaussian kernels), respectively, when hard-limiting threshold was used.

TABLE III
CONFUSION MATRIX FOR TWO-CLASS RVM (GAUSSIAN) CLASSIFIER

Truth \ Dec.	UXO	Non-UXO
UXO	0.650	0.350
Non-UXO	0.259	0.741

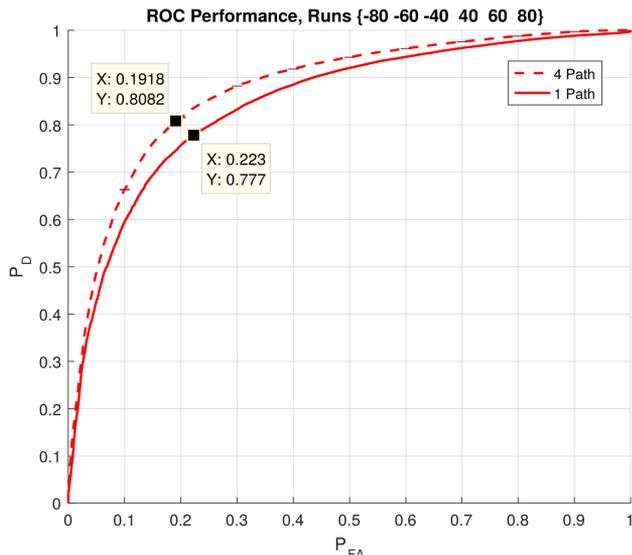


Fig. 7. ROC curves for two-class classifiers—Single path versus all paths near end-on.

To show the importance of including all four paths versus only single path (i.e., the first one) in the fast ray model of (2), an experiment was conducted where the MSC was trained using the model-generated data for the single-path case for all target ranges, orientations, types, and for $P = 7$ aspects, as in the four path case described before. However, the trained classifiers were tested only on a subset of the experimental TREX13 data set corresponding to object orientations -80° , -60° , -40° , $+40^\circ$, $+60^\circ$, and $+80^\circ$, i.e., off-broadside. Fig. 7 shows the ROC curves of these MSC-based classifiers for this experiment. As can be seen, the knee point of the ROC curve for the four-path MSC shows an improvement of about 3% in correct classification rate over that of the single-path case. Additionally, comparison of the ROC curve for the four-path classifier in Fig. 7 tested only for the near end-on orientations with that of the classifier in Fig. 6 for all orientations in the TREX13 data set shows less than 2% degradation in the classification performance due to removal of those orientations near broadside. This experiment not only validates the importance of including paths 2 and 3 in the model-generated data to capture contributions to the TS when the source and receiver are in a nearly end-on orientation with a target but also shows that there is indeed exploitable discriminatory information in the nonbroadside aspects. The latter is very important since in an actual operational setting, the sonar may not have a broadside look at the target.

2) *Four-Class Classification Results*: In this section, the utility of the MSC-based system is tested for $M = 4$ class discrimination. The eight objects in the TREX13 data set are

TABLE IV
CONFUSION MATRIX FOR FOUR-CLASS DISCRIMINATION PROBLEM

Truth \ Dec.	Class 1	Class 2	Class 3	Class 4
Class 1	0.676	0.052	0.058	0.214
Class 2	0.026	0.710	0.238	0.026
Class 3	0.015	0.449	0.482	0.054
Class 4	0.277	0.201	0.176	0.346

arranged into groups of two objects based upon similarity of their shapes and object types. Class 1 represents non-UXO objects, whereas classes 2–4 correspond to UXO objects. This experiment will also shed some light on the reasons for misclassifications among different object types. Table IV gives the confusion matrix for the $P = 7$ aspect-MS-C-based classifier when used to discriminate these four classes at all ranges and conditions as explained previously. As can be noted, most false alarm cases ($\approx 20\%$) happened because of confusion with class 4 UXO objects at some ranges and orientations. About the same percentage of UXO class 4 objects are also misclassified as non-UXO. We believe this is largely caused by the similarity in shape and size of object types in classes 1 and 4. As far as the within UXO class confusion is concerned, misclassification mostly occurred among UXO objects of classes 2 and 3.

3) *In Situ Dictionary Updating and Results*: As mentioned previously, one of the important benefits of the MSC-based classifier is that the dictionary matrices \mathbf{H}_m s can easily be updated *in situ* to incorporate some limited labeled data drawn from the new environment in which the sonar is operating. This can be done without jeopardizing the performance on the previously learned data using the IK-SVD method described in Section V-B and also without the need to carry over the previous training data sets. Here, we adopted this algorithm by keeping the old dictionaries trained using the model-generated data for UXO and non-UXO classes intact while augmenting them with the incremental dictionary matrices $\tilde{\mathbf{H}}_m$ that are *in situ* trained using less than 15% of the randomly drawn TREX13 data set. For this experiment, we allowed adding an incremental dictionary with only five columns (i.e., $\tilde{L} = 5$) while keeping the sparsity factor $\tau = 15$ as before. The number of incrementally added atoms may be decided by monitoring the MSE as well as the entropy of the features [21] during the training. It was experimentally determined that adding five dictionary atoms is adequate for this particular application. Fig. 8 shows the ROC curve (dashed line) for this system when tested on the remaining TREX13 data set using the same conditions as stated in Section VI-C1, i.e., all target ranges, orientations, and $P = 7$ aspects. As expected, *in situ* training in the new environment significantly improved ($\approx 9\%$) the classification performance. Furthermore, this *in situ* training can easily be implemented using the MSC-based classifier in conjunction with the IK-SVD algorithm using a limited data set. Table V gives the confusion matrix for this two-class MSC-based classifier when hard-limiting threshold was used. An improvement of about 7% can be noted when compared to the results in Table I. Table VI, on the other hand, gives the

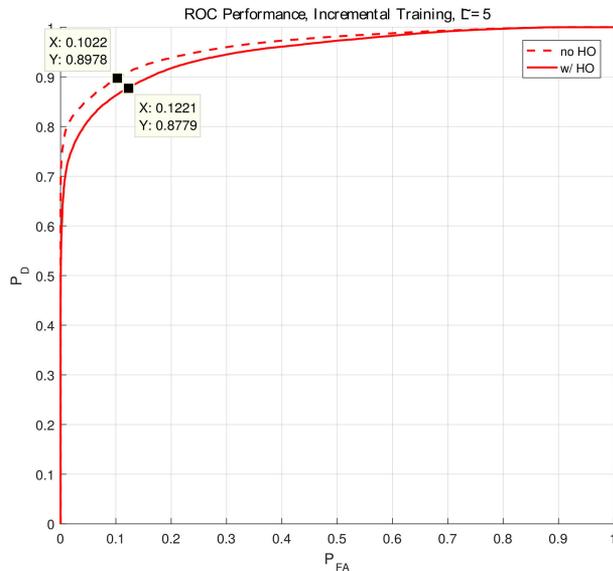


Fig. 8. ROC curves for *in situ* trained MSC with and without holdout.

TABLE V
CONFUSION MATRIX FOR TWO-CLASS MSC USING *In Situ* TRAINING
WITH IK-SVD METHOD

Truth \ Dec.	UXO	Non-UXO
UXO	0.728	0.272
Non-UXO	0.001	0.999

TABLE VI
CONFUSION MATRIX FOR FOUR-CLASS MSC USING *In Situ* TRAINING

Truth \ Dec.	Class 1	Class 2	Class 3	Class 4
Class 1	0.797	0.023	0.018	0.193
Class 2	0.023	0.698	0.240	0.051
Class 3	0.006	0.134	0.828	0.032
Class 4	0.256	0.122	0.148	0.473

confusion matrix for the four-class MSC-based classifier when incrementally trained with the IK-SVD algorithm using similar conditions as in Section VI-C2. Compared to the results presented in Table IV, the overall improvement in the correct classification rate is close to 26%, which is indeed significant. These results show the flexibility and ease of the MSC-based system for adaptive classification problems.

To further examine the generalization ability of the MSC-based classifier after *in situ* training on unseen objects, a separate experiment was conducted where all samples from the two UXOs were left out of the *in situ* training set. However, all samples (at all ranges and orientations) of these unseen objects were included in the testing data set to evaluate the classification performance using the *in situ* updated dictionaries. Fig. 8 also illustrates the ROC curve (solid line) of the MSC for this experiment. Compared to the ROC curve associated with the no holdout case (dashed line), the degradation in performance

as a result of excluding two UXO objects during the *in situ* training is found to be about 2%. These results demonstrate the generalization ability of the MSC when *in situ* trained using only a subset of UXO objects found in the TREX13 operating environment.

It must be pointed out that although the RVM (linear or non-linear) classifier was found to provide comparable results to those of the MSC-based classifier, these results are not included here. This is because the RVM classifier is not suitable for *in situ* training for several reasons. First, it will require complete retraining on the augmented data set. Second, it requires carrying potentially large data sets from the previous training environments, which makes it unsuitable for real operating environments. Third, the computational requirements are by far more significant than those of the MSC-based classifier.

VII. CONCLUSION

In this paper, a new system for discriminating underwater UXO and non-UXO objects is proposed using an adaptive MSC-based classifier. The system was trained exclusively on model-generated data using a fast ray model and subsequently tested on the experimental TREX13 data set to determine the generalization ability of the classifier. The motivation behind this training was that one cannot easily collect an abundance of data for different UXO shapes and sizes in realistic settings. The signal-dependent K-SVD dictionary-learning method was employed to build the dictionary matrices for different object types based upon their acoustic color data generated using the fast ray model. It was shown that using these dictionaries one can train the MSC-based classifier to successfully perform two-class or multiclass decision-making on field-collected experimental data with good generalization, especially when the objects are viewed from several aspect angles. This is mainly due to lack of useful discrimination information for some object types at certain aspect angles. Performance benchmarking against RVM classifiers was also carried out on the same test data set. Finally, to show the adaptability of the MSC system in new operating environments, an *in situ* learning algorithm was proposed to incrementally update the dictionary matrices using the IK-SVD method. This can be done without impacting the prior training. The MSC-based classifier was then *in situ* trained using a small subset of the TREX13 data (with and without holding out some UXO objects) and tested on the remaining test set. Substantial improvements in the classification performance were noted under all scenarios. This study not only showed the importance of *in situ* training in the new environment but also reveals the generalization ability of the proposed adaptive classification system. Additionally, unlike the RVM classifier that assumes Gaussian density for the weight vector and uses Gaussian approximation of the posterior distribution, no knowledge of the distributions is assumed in the MSC-based classifier. As a result, in situations where there is a significant mismatch between the assumed model and the actual data, the MSC-based-type classifiers perform much better.

Future work could include: testing this MSC-based UXO classification framework on more challenging sonar data sets that contain high density of natural clutter; and further investigation

of the adaptive dictionary-learning method and development of an information-theoretic method for determining which samples drawn from the new operating environment require querying the sonar operator for labels.

ACKNOWLEDGMENT

The TREX13 data used in this study were collected during a joint collaboration between NSWC-Panama City, Washington Navy Yard, DC, USA, and the Applied Physics Laboratory, University of Washington, Seattle, WA, USA. The authors would like to thank them for their support.

REFERENCES

- [1] M. Zampolli, F. Jensen, and A. Tesei, "Benchmark problems for acoustic scattering from elastic objects in the free field and near the seafloor," *J. Acoust. Soc. Amer.*, vol. 125, pp. 89–98, Jan. 2009.
- [2] S. Dey *et al.*, "Structural-acoustic modeling for three-dimensional freefield and littoral environments with verification and validation," *J. Acoust. Soc. Amer.*, vol. 129, pp. 2979–2990, May 2011.
- [3] Z. Waters *et al.*, "Bistatic, above-critical angle scattering measurements of fully buried unexploded ordnance (UXO) and clutter," *J. Acoust. Soc. Amer.*, vol. 132, pp. 3076–3085, Nov. 2012.
- [4] K. Williams *et al.*, "Acoustic scattering from a solid aluminum cylinder in contact with a sand sediment: Measurements, modeling, and interpretation," *J. Acoust. Soc. Amer.*, vol. 127, pp. 3356–3371, 2010.
- [5] S. Kargl, K. Williams, and E. Thorsos, "Synthetic aperture sonar imaging of simple finite targets," *IEEE J. Ocean. Eng.*, vol. 37, no. 3, pp. 516–532, Jul. 2012.
- [6] S. Kargl, K. Williams, T. Marston, and J. Kennedy, "Acoustic response of unexploded ordnance (UXO) and cylindrical targets," in *Proc. MTS/IEEE OCEANS Conf.*, Seattle, WA, USA, Sep. 2010, pp. 1–5.
- [7] N. Intrator, N. Neretti, and Q. Huynh, "Sonar object discrimination via spectral density analysis," in *Proc. MTS/IEEE TECHNO-OCEAN'04*, Nov. 2004, pp. 740–742.
- [8] J. Tucker, W. Wu, and A. Srivastava, "Analysis of signals under compositional noise with applications to SONAR data," *IEEE J. Ocean. Eng.*, vol. 39, no. 2, pp. 318–330, Apr. 2014.
- [9] J. Bucaro *et al.*, "Acoustic identification of buried underwater exploded ordnance using a numerically trained classifier," *J. Acoust. Soc. Amer.*, vol. 132, pp. 3614–3617, Dec. 2012.
- [10] N. Wachowski and M. Azimi-Sadjadi, "A new synthetic aperture sonar processing method using coherence analysis," *IEEE J. Ocean. Eng.*, vol. 34, no. 4, pp. 665–678, Oct. 2011.
- [11] P. Runkle, L. Carin, L. Couchman, J. A. Bucaro, and T. J. Yoder, "Multi-aspect identification of submerged elastic targets via wave-based matching pursuits and hidden Markov models," *J. Acoust. Soc. Amer.*, vol. 106, pp. 605–616, Aug. 1999.
- [12] J. Cartmill, M. Azimi-Sadjadi, and N. Wachowski, "Buried underwater object classification using a collaborative multi-aspect classifier," in *Proc. Int. Joint Conf. Neural Netw.*, 2007, pp. 1807–1812.
- [13] M. Tipping, "The relevance vector machine," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 652–658.
- [14] E. Fischell and H. Schmidt, "Classification of underwater targets from autonomous underwater vehicle sampled bistatic acoustic scattered fields," *J. Acoust. Soc. Amer.*, vol. 138, pp. 3773–3784, Dec. 2015.
- [15] S. S. Haykin, *Neural Networks and Learning Machines*, vol. 3. Upper Saddle River, NJ, USA: Pearson, 2009, ch. 6, pp. 269–269.
- [16] L. Scharf and C. Mullis, "Canonical coordinates and the geometry of inference, rate, and capacity," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 824–831, Mar. 2000.
- [17] S. Kargl, A. España, K. Williams, J. Kennedy, and J. Lopes, "Scattering from objects at a water-sediment interface: Experiment, high-speed and high-fidelity models, and physical insight," *IEEE J. Ocean. Eng.*, vol. 40, no. 3, pp. 632–642, Jul. 2015.
- [18] A. Salberg, A. Hanssen, and L. Scharf, "Robust multidimensional matched subspace classifiers based on weighted least-squares," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 873–880, Mar. 2007.
- [19] J. Yang, D. Tang, B. Hefner, K. Williams, and J. Preston, "Overview of reverberation measurements in target and reverberation experiment 2013," in *Proc. TREX13 Workshop*, Nov. 2014, pp. 784–794.
- [20] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [21] L. Wang, K. Lu, P. Liu, R. Ranjan, and L. Chen, "IK-SVD: Dictionary learning for spatial big data via incremental atom update," *IEEE Comput. Sci. Eng.*, vol. 16, no. 4, pp. 41–52, Jul./Aug. 2014.
- [22] S. Kargl and K. Williams, "Full scale measurement and modeling of the acoustic response of proud and buried munitions at frequencies from 1–30 kHz," Univ. Washington, Seattle, WA, USA, Final Rep. SERDP Project MR-1665, May 2012, pp. 18–19.
- [23] A. L. España, K. L. Williams, D. Plotnick, and P. Marston, "Acoustic scattering from a water-filled cylindrical shell: Measurements, modeling, and interpretation," *J. Acoust. Soc. Amer.*, vol. 136, no. 1, pp. 109–121, Jul. 2014.
- [24] T. Marston, P. Marston, and K. Williams, "Scattering resonances, filtering with reversible SAS processing, and applications of quantitative ray theory," in *Proc. MTS/IEEE OCEANS Conf.*, Seattle, WA, USA, Sep. 2010, pp. 1–9.
- [25] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
- [26] L. Scharf, *Statistical Signal Processing Detection, Estimation, and Time Series Analysis*. Reading, MA, USA: Addison-Wesley, 1991, chs. 4, 9, pp. 145–146, 365.
- [27] M. Azimi-Sadjadi, J. Kopacz, and N. Klausner, "K-SVD dictionary learning using a fast OMP with applications," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 1599–1603.
- [28] I. Markovsky, *Low-Rank Approximation: Algorithms, Implementation, and Applications*. Berlin, Germany: Springer-Verlag, 2011, ch. 2, pp. 62–67.
- [29] C. M. Bishop, *Pattern Recognition*, vol. 128. New York, NY, USA: Springer-Verlag, 2006, ch. 7, pp. 345–356.
- [30] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.



John Jack Hall (M'13) received the B.S. degree in electrical engineering from Miami University, Oxford, OH, USA, in 2014 and the M.S. degree in electrical and computer engineering from Colorado State University, Fort Collins, CO, USA, in 2016, where he is currently working toward the Ph.D. degree in electrical engineering.

His research interests include digital and statistical signal and image processing, adaptive filtering, artificial intelligence, and machine learning.



Mahmood R. Azimi-Sadjadi (S'81–M'81–SM'89–LSM'18) received the M.S. and Ph.D. degrees from the Imperial College of Science and Technology, University of London, London, U.K., in 1978 and 1982, respectively, both in electrical engineering with specialization in digital signal/image processing.

He is currently a Full Professor at the Electrical and Computer Engineering Department, Colorado State University (CSU), Fort Collins, CO, USA. He is also the Director of the Digital Signal/Image Laboratory, CSU. His research interests include statistical signal

and image processing, machine learning and adaptive systems, target detection, classification and tracking, sensor array processing, and distributed sensor networks.

Dr. Azimi-Sadjadi was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON NEURAL NETWORKS.



Steven G. Kargl received the B.S. degree in physics and mathematics from the University of Dayton, Dayton, OH, USA, in 1985 and the M.S. and Ph.D. degrees in physics from Washington State University, Pullman, WA, USA, in 1987 and 1990, respectively.

For the next three years, he was a Research Physicist at the Coastal Systems Station, Naval Surface Warfare Center, Panama City, FL, USA. In 1993, he joined the Applied Physics Laboratory, University of Washington, Seattle, WA, USA, where he continues to investigate a variety of topics in acoustics. He is currently a Senior Principal Physicist in the Applied Physics Laboratory. His research interests include wave propagation in porous media and ambient noise in littoral waters. His recent research efforts have focused on the application of acoustic techniques to aid in the remediation of underwater unexploded ordnance.

Dr. Kargl is a Fellow of the Acoustical Society of America. He was an Associate Editor for the *Journal of the Acoustical Society of America* from 2000 to 2003 with editorial responsibilities in the areas of ultrasonics and physical acoustics.



Yinghui Zhao received the B.S. degree in electrical and information engineering from the University of Science and Technology of China, Hefei, China, in 2002 and the Ph.D. degree in communication and information system from Chinese Academy of Sciences, Beijing, China, in 2007.

She was a Postdoctoral Researcher (2008–2009) and a Visiting Scholar (2016–2017) at the Colorado State University, Fort Collins, CO, USA. She is currently an Associate Professor with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. Her research interests include synthetic aperture radar/sonar signal processing, remote sensing signal processing, and deep learning.



Kevin L. Williams received the B.S., M.S., and Ph.D. degrees in physics from Washington State University, Pullman, WA, USA, in 1979, 1983, and 1985, respectively.

From 1985 to 1988, he was with the Naval Coastal Systems Center, Panama City, FL, USA, where his primary focus was in acoustic scattering from finite bodies and propagation into ocean sediments. In 1988, he moved to the Applied Physics Laboratory, University of Washington, Seattle, WA, USA, where he was involved in the area of high-frequency environmental acoustics (studying propagation through ocean internal waves and arctic ice and propagation in and scattering from ocean sediments) and understanding the response of targets in the ocean environment.

Dr. Williams is a Fellow of the Acoustical Society of America.