

Statistical Classification of Buried Unexploded Ordnance Using Nonparametric Prior Models

Alireza Aliamiri, *Student Member, IEEE*, Jack Stalnaker, and Eric L. Miller, *Senior Member, IEEE*

Abstract—We used kernel density estimation (KDE) methods to build *a priori* probability density functions (pdfs) for the vector of features that are used to classify unexploded ordnance items given electromagnetic-induction sensor data. This *a priori* information is then used to develop a new suite of estimation and classification algorithms. As opposed to the commonly used maximum-likelihood parameter estimation methods, here we employ a maximum *a posteriori* (MAP) estimation algorithm that makes use of KDE-generated pdfs. Similarly, we use KDE priors to develop a suite of classification schemes operating in both “feature” space as well as “signal/data” space. In terms of feature-based methods, we construct a support vector machine classifier and its extension to support *M*-ary classification. The KDE pdfs are also used to synthesize a MAP feature-based classifier. To address the numerical challenges associated with the optimal data-space Bayesian classifier, we have used several approximation techniques, including Laplacian approximation and generalized likelihood ratio tests employing the priors. Using both simulations and real field data, we observe a significant improvement in classification performance due to the use of the KDE-based prior models.

Index Terms—Classification, density estimation, electromagnetic induction (EMI), subsurface sensing, unexploded ordnance (UXO).

I. INTRODUCTION

COSTLY, time-consuming, and sometimes perilous excavations associated with false classification of buried unexploded ordnances (UXOs) demand robust and powerful processing algorithms. To this end, in the past decade, UXO detection and classification methods have evolved significantly, both with the development of more sophisticated sensors as well as more accurate physical models and signal processing algorithms [1]–[12]. Indeed, current electromagnetic-induction (EMI) sensors are capable of recording field response of targets with high spatial and temporal resolution. Physical models of the EMI sensors have been developed that enable us to model buried objects in terms of features that are relevant to classifica-

tion. From these data and models, two classes of problems are typically considered. The detection of UXO can be considered as a binary hypothesis problem where the object is present or not and was considered extensively in, e.g., [13]–[15]. Our focus here is on the related classification problem in which we want to infer information about the type of UXO being investigated.

Most model-based methods for the classification of buried objects from the EMI sensor data are composed of two steps. First, the data are used to estimate a set of parameters, some of which are of direct use in classification, and some of which are required to locate and orient the object relative to the sensor. These estimated parameters are then input to a classifier to determine the type of target under investigation. In many cases [11], [16], underlying both of these processing steps is an implicit assumption that the object-specific parameter values are independent of the orientation and location of the target. A broad array of modeling approaches is available that meets these criteria, from exact numerical models based on finite element and finite difference approaches to models that approximate the buried object as a magnetic dipole. We have chosen to implement the latter “dipole model,” which is a decision motivated by the practical need for rapid data processing and by the ubiquity of its use in the UXO detection and classification literature [16]–[22]. We note that more accurate models of the physics of UXO-like targets are available, and in some sense, the dipole model could be considered the first-order application of the techniques developed in this paper, which are extensible to any parameterized model representing UXO.

Under the dipole model, the information concerning the target class is encoded in the 3×3 magnetic polarizability tensor (MPT) whose three time-dependent eigenvalues consist of an infinite sum of decaying exponentials. The decay rates are the features of use in target classification. In theory, the veracity of this model requires illumination of an infinitesimally small target by a uniform field, which is a condition that is not generally met in practice. Moreover, given limited noisy data, at most, a single term in each of the three sums can be estimated reliably [21], leading some investigators to use forms of the model employing only two exponentials [21], [23] on the assumption that targets of interest are spheroidal and, therefore, have two equal eigenvalues. These theoretical and practical limitations of the dipole model impart a position and orientation dependence on the object-specific parameter estimates required for an accurate classification. Thus, the analysis of multiple instances (i.e., positions, orientations, and target example) of a target class results in not a single parameter value for each object but rather a cloud of values in parameter space.

Manuscript received June 6, 2006; revised February 17, 2007. This work was supported by the Strategic Environmental Research and Development Program through Project MM-1379.

A. Aliamiri is with Airvana, Inc., Chelmsford, MA 01824 USA (e-mail: aaliamir@airvana.com).

J. Stalnaker is with the Textron Systems Corporation, Wilmington, MA 01887 USA (e-mail: jstalnak@systems.textron.com).

E. L. Miller is with the Department of Electrical and Computer Engineering, Tufts University, Medford, MA 02155 USA (e-mail: elmiller@ece.tufts.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2007.900681

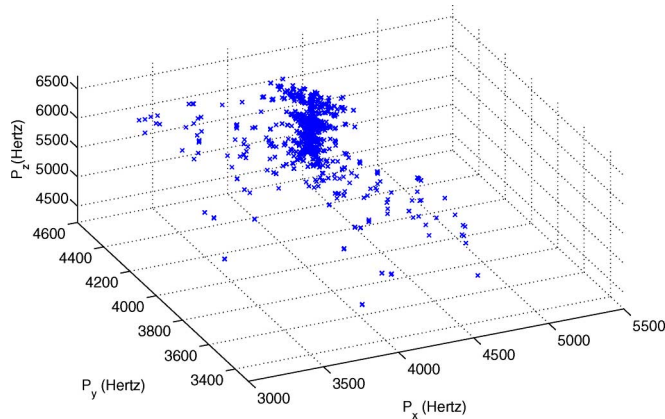


Fig. 1. Pole library for a cylindrical steel object.

To capture the dependence of the features on the location and orientation of the object, several methods have been proposed in the past. For example, in [23], the authors used Laplace plane poles located on the real s -axis for classification purposes. To account for the fact that the poles differed according to the orientation of the object in space, a single object was regarded as three different targets, one for each of the three orthogonal orientations of the true UXO. Identification for arbitrary orientations and the issue of object localization were not considered in [23]. In [10], instead of a sum of decaying exponentials model, a frequency-domain processing approach was considered. Here, the data at each frequency were used to estimate all nine elements of the MPT. An eigenanalysis of the estimated MPTs was used to determine a rotation matrix which implicitly described the orientation of the object in space as well as a collection of frequency-dependent features required for classification.

In our previous work [24], [25], a processing algorithm was developed based on the fusion of the dipole scattering model and a parametric model for the eigenvalues of the MPT which are referred to here as principle axis polarizability functions (PAPFs). The model directly captures both the features for classification and the parameters required to infer the location and orientation of the buried object. The model uses a single-pole per axis version of the PAPF expansion introduced in [21] which results in a 3-D feature space with one-pole per body axis. Because of this approximation, the features are no longer independent of the object orientation and location. This dependence is captured by building a library of poles for the different orientations and locations. That is, each object in the library was represented in terms of a “cloud” of pole values as shown in Fig. 1 for a 3-in long by a 3-in diameter stainless steel cylinder.

In terms of processing, a simple nonlinear least square method was used in [24] and [25] to extract poles, location, and orientation information from the data. Using the estimated poles, classification was performed using Mahalanobis distance metric. More specifically, for each object in the library, this distance metric was constructed using the mean pole vector and covariance matrix for the associated pole cloud. In essence then, our work in [24] and [25] was predicated on the implicit

assumption that the pole cloud was Gaussian. As is evident from Fig. 1 and, indeed, as we have found to be true in general, such an assumption is not, in fact, warranted in practice [16].

Here, we consider new processing methods that account for this variability in ways that improve both the accuracy of the parameter estimates as well as the final classification. We interpret these clouds as prior probability density functions (pdfs) over the object-specific parameters and use kernel-density-estimation (KDE) techniques to determine these pdfs. The resulting prior models in parameter space are used to develop a new collection of estimation and classification algorithms. In fact, it should be noted that the variability in the estimated poles is, in part, deterministic in origin. In the absence of additive noise, variation in the estimated poles is due entirely to model mismatch, i.e., the fact that the dipole model is only an approximation to the true underlying physics. Conceivably, the spread of the pole cloud could be minimized given a physical model that more accurately represented the EMI system. Still, in the absence of an exact model (which would be intractable as a tool for processing), one would have some degree of parameter spreading that is necessitating the use of the type of method we explore here. Thus, we concern ourselves here exclusively with the use of the dipole-type model. First, we feel that this choice allows us to concentrate on the technical issues at the heart of this paper, which are the processing methods. Second, given the widespread use of the dipole model both in the research community as well as in the fielded systems, there is value in the development of processing methods that extend the utility of this tool. Thus, we have chosen to treat the diffuse nature of the pole cloud using the machinery of probability theory due to the computational expense incurred by a more accurate model, approaching an otherwise intractable yet deterministic problem as a probabilistic one [26].

In terms of estimation, the priors that we build allow for the use of maximum *a posteriori* (MAP) methods for feature extraction rather than maximum-likelihood (ML) techniques. For classification, we consider the schemes that function both in feature space (i.e., exclusively using the estimated parameters themselves) as well as data space (i.e., using these features along with a sensor model to simulate data and compute residuals). In the former case, we developed the feature-space MAP classifiers built around our KDE pdfs and a more generic support vector machine (SVM) classifier. For the data-space classifiers, we have developed a pair of generalized likelihood ratio-type tests as well as an approximation to the Bayesian optimal classifier that makes use of the Laplacian approximation to the multidimensional integral encountered in this approach to classification.

The performance of these methods is evaluated at different signal-to-noise (SNR) ratios and compared in terms of complexity and probability of error. Results using both simulations and real field data indicate that the incorporation of these prior pdfs into the classification processing can substantially improve the accuracy relative to our previous mean-covariance classifiers.

This paper is organized as follows. In Section II, the physical model of the problem and basic signal processing are reviewed. In Section III, the KDE and its applications to our work are

examined. In Section IV, the proposed classification algorithms are discussed. Section V compares and discusses the results.

II. PROCESSING MODEL AND PRIOR WORK

The processing model used here is based on [24], which is a truncated version of the EMI physical model in [9] and [21]. Using this model, the scattered signal collected at M time gates or frequencies (depending on the sensor) at each of L locations in space can be written as

$$y_{l,m} = H_{R,l}^T(r_0)R^T(\alpha)\Lambda_m(p, c, \delta)R(\alpha)H_{X,l}(r_0) + n_{l,m},$$

$$l = 1, \dots, L, \quad m = 1, \dots, M \quad (1)$$

where

$y_{l,m}$	received signal at location l and frequency or time sample m ;
$H_{X,l}$ and $H_{R,l}$	3×1 vectors holding the x , y , and z components of the transmitted fields and the hypothetical field of the receiver [27]; both of these vectors are functions of the 3×1 vector r_0 containing the x , y , and z locations of the UXO item;
$n_{l,m}$	additive Gaussian noise at location l and frequency/time m ;
$R(\alpha)$	3×3 rotation matrix orienting the object in space and dependent on a 3×1 vector α of Euler angles.

The matrix $\Lambda_m(p, c, \delta)$ is a 3×3 diagonal matrix taking the form

$$\Lambda_m(p, c, \delta) = \begin{bmatrix} \lambda_{m,1}(p, c, \delta) & 0 & 0 \\ 0 & \lambda_{m,2}(p, c, \delta) & 0 \\ 0 & 0 & \lambda_{m,3}(p, c, \delta) \end{bmatrix}. \quad (2)$$

Each of the diagonal elements of Λ_m , $\lambda_{m,i}(i=1,\dots,3)$, is a PAPF which contains orientation-invariant spectral signatures and can be written as infinite sum of one pole transfer functions in the frequency domain or sum of exponentials in the time domain

$$\lambda_{m,i}(p, c, \delta) = \begin{cases} d_i + \sum_{l=0}^{\infty} \frac{c_{i,l} j f_m}{p_{i,l} + j f_m} & \text{(Frequency)} \\ d_i \delta(t_m) + \sum_{l=0}^{\infty} -p_{i,l} c_{i,l} e^{-p_{i,l} t_m} & \text{(Time)}. \end{cases} \quad (3)$$

Going forward, we refer to the parameters p_i in (3) as “poles” in recognition to the fact that they are the roots of the transfer functions in the frequency-domain form of the $\lambda_{m,i}$. Combining the data from all spatial locations and time or frequency samples gives the model

$$y(\theta) = [y_1^T(\theta) \quad y_2^T(\theta) \quad \dots \quad y_M^T(\theta)]^T$$

$$y_i(\theta) = m_i(\theta) + n_i \quad (4)$$

where

- $\theta = [p^T, \alpha^T, r_0^T, c^T, \delta^T]^T$ with, e.g., p the vector whose i th element is p_i from (3);
- y_i $L \times 1$ received vector signal at frequency/time m ;
- m_i $L \times 1$ vector model in (1);
- n_i $L \times 1$ vector of Gaussian noise which is taken to have zero mean and covariance matrix C independent of i .

In our processing, the feature vector to be used for classification is composed of the three poles, with one for each of the principal axes of the target. That is, only the $l = 0$ terms in the summations of (3) are retained. It then proves convenient for us to partition the parameter vector θ into features (i.e., the poles) and nuisance parameters, where the former are relevant to classification and the latter consist of location, orientation, coefficient factors, and dc parameter which are used to infer the geometry of the object rather than its type. The parameter vector will be denoted as $\theta = [p^T, \nu^T]^T$, where p is the 3×1 vector of poles, and ν is the vector of the nuisance parameters.

Here, the tacit assumption has been made that the data have been preprocessed in order to remove the systematic noise sources. For example, it is a standard practice to subtract the response of the background in the absence of the target from the data [14], [19]. Additional noise deriving, for example, from sensor position uncertainty and mutual coupling in the presence of other metallic targets must also be considered. However, it may be assumed that the classification approach presented in this paper is part of a larger data-processing strategy that includes tactics for dealing with these circumstances.

As noted previously, the approximate nature of this model will, in fact, impart a dependence of these parameters on the location and orientation of the object relative to the sensor. In order to account for the dependence of features on object location and orientation in the processing, this paper makes use of a “pole library” for each object. This library is composed of the estimated values of the poles as a function of object position and orientation. Given either measured or simulated data for an object in one of the N possible positions/orientations, we build samples from this library via an ML approach¹

$$\hat{p}_{ML,j} = \arg \min_{p,\nu} \left\| C^{-1/2} (y(\alpha_j, r_{0,j}) - m(p, \nu)) \right\|_2^2 \quad (5)$$

for $j = 1, 2, \dots, N$. Fig. 1 shows an example of a library (feature space) built for a steel cylindrical object corresponding to five possible object depths, seven possible values for each of the three Euler angles, and no horizontal variation of target. Data were generated using a four-pole per axis dipole model in which the sums in (3) are terminated after four terms [21], [24].

In our previous work [24], classification was done in feature space using a pole library in a fairly simplistic manner. Assuming that we had a library of $i = 1, 2, \dots, N$ objects, we generated a pole cloud like the one in Fig. 1 for each target. For each cloud, we computed the mean pole vector \bar{p}_i as well as the associated covariance matrix C_i . Given a data set containing an object to be classified, we first estimated the best fit model

¹See Section IV-A for a discussion as to how the optimization problem is solved.

parameters using an ML scheme that is much like (5) with $y(\alpha_j, r_{0j})$ replaced by data containing an object of unknown structure. Classification was then done using a Mahalanobis distance classifier based on the statistics of the different pole clouds [28]

$$\hat{i} = \arg \min_i (\hat{p}_i - \bar{p}_i)^T C_i^{-1} (\hat{p}_i - \bar{p}_i) \quad (6)$$

where \hat{p}_i is the ML pole estimate.

Fig. 1 shows that the shape of pole cloud is far from an ellipsoid, thereby invalidating the Gaussian assumption used in [24]. This leads us to believe that a more accurate model of this distribution could improve the overall performance of the processing. Substantiation of this claim and exploitation of the resulting structure to improve classification are the essential contributions of this paper. In the following sections, we will apply nonparametric probability-distribution estimation to build a more accurate representation of the pole distribution. These nonparametric distributions, therefore, represent the enhanced prior models for the pole clouds associated with each object. Using this prior information, the processing scheme can be set into a Bayesian framework both for estimation and classification. The results show significant improvement over the simple Mahalanobis approach, particularly for Bayesian classifiers at low SNR ratios, as will be explained more in Section V.

III. PRIOR MODEL

The assumption of parametrically defined probability distributions for inference and hypothesis testing fails in many practical engineering problems; therefore, the techniques which make less rigid assumptions about the distribution of data are required. Nonparametric estimation schemes have been considered for this purpose [29]–[31]. Given a set of N d -dimensional data samples $\{x_n, n = 1, \dots, N\}$ of an unknown distribution, the problem is to estimate the probability distribution in some flexible manner. Kernel methods proceed by locating a function ϕ with “width” parameter h at each observed datum to generate the estimated density $\hat{f}(x)$ as [29]

$$\hat{f}(x) = \frac{1}{Nh^d} \sum_{n=1}^N \phi\left(\frac{1}{h}(x - x_n)\right). \quad (7)$$

The kernel function ϕ is chosen to satisfy

$$\phi(x) \geq 0 \quad \text{and} \quad \int_{\mathbb{R}^d} \phi(x) dx = 1. \quad (8)$$

Here, we use Gaussian kernels which take the form

$$\phi(x) = (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}x^T x\right). \quad (9)$$

In order to obtain a probability distribution over feature space, the pole estimates comprising the cloud are considered as samples of an unknown probability distribution function, and

KDE is used to estimate the corresponding pdf. As the poles are not equally spread in all directions, the data are first whitened by a linear transformation yielding data with zero mean and unit covariance matrix. Then, density estimation is applied to the transformed data. This method was proposed in [32], and it helps provide a better approximation of optimal bandwidth. Given an observed data x , the whitened data z is written as

$$z = S^{-\frac{1}{2}} (x - E(x)) \quad (10)$$

where $E(x)$ is the sample mean, and S is sample covariance

$$S = E\left[(x - E(x))(x - E(x))^T\right]. \quad (11)$$

The resulting KDE for the original data can be written as

$$\hat{f}(x) = \frac{\det(S)^{-\frac{1}{2}}}{Nh^d} \sum_{n=1}^N \phi\left(\frac{1}{h}S^{-\frac{1}{2}}(x - x_n)\right). \quad (12)$$

For our estimation, the Gaussian kernel with bandwidth

$$h = AN^{-\frac{1}{d+4}}, \quad \text{where} \quad A = \left[\frac{4}{2d+1}\right]^{\frac{1}{d+4}} \quad (13)$$

is used, which provides a good approximation for the whitened data [30]. As an example, we applied KDE to the pole cloud in Fig. 1. In order to visualize the density estimate, slice cuts along the x -axis were made corresponding to $p_x = \{3500, 3720, 3940, 4160, 4380, 4600\}$ (Hertz). The contour plots shown in Fig. 2 demonstrate the estimated pdf on each slice. From this figure, it is obvious that kernel density methods are, in fact, of use in capturing the markedly non-Gaussian nature of these pole distributions.

IV. ENHANCED ESTIMATION AND CLASSIFICATION

In this paper, we make use of the nonparametrically constructed pdfs discussed in the previous section to develop new parameter estimation algorithms as well as a collection of classification schemes. While the details of all methods are discussed at length in this section, the reader is referred to Table I for a summary of the different techniques.

A. Estimation

For estimation, the ML estimation used in (5) is replaced by the MAP estimation for features using our nonparametric priors

$$\hat{p}_i, \hat{\nu}_i = \arg \max_{p, \nu} f_i(y/p, \nu) f_i(p) \quad (14)$$

where $f_i(p)$ is the pdf of poles corresponding to the i th object. Given the assumption of Gaussian noise, (14) is equivalent to

$$\hat{p}_i, \hat{\nu}_i = \arg \min_{p, \nu} \left\| C^{-1/2} (y - m(p, \nu)) \right\|_2^2 - \ln f_i(p). \quad (15)$$

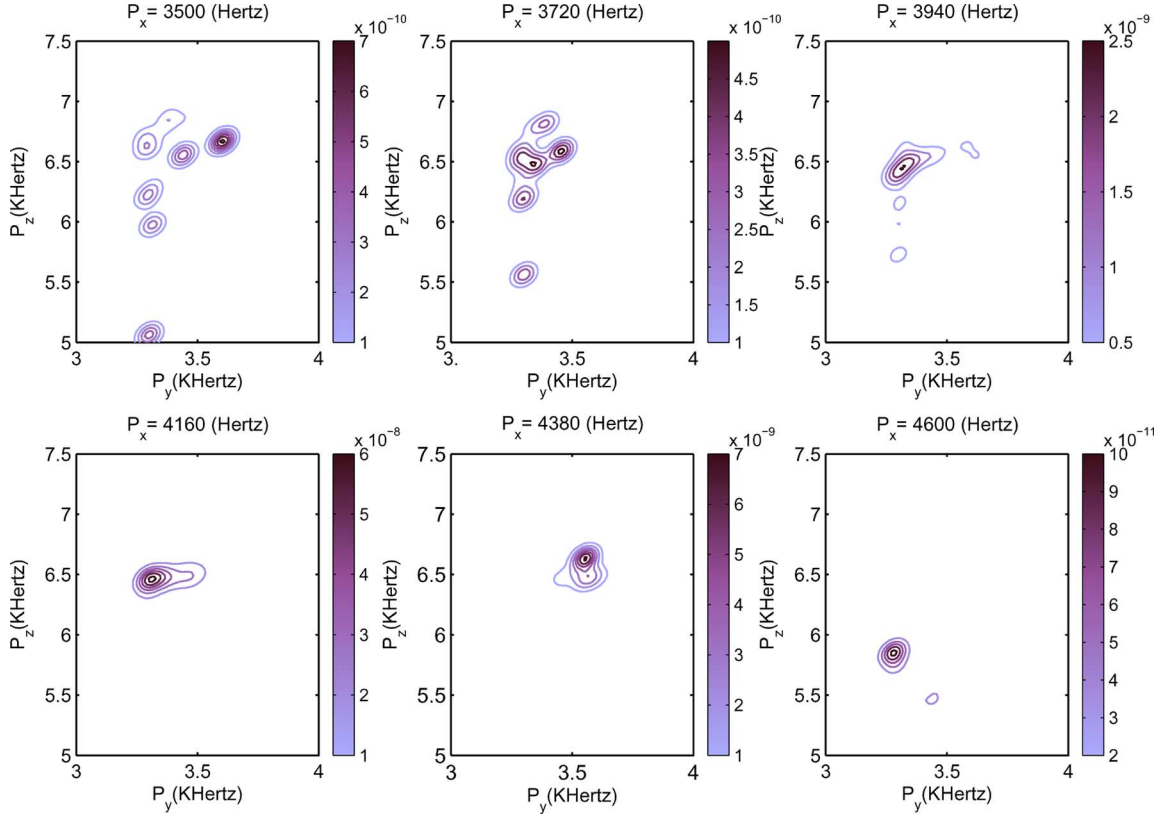


Fig. 2. Contours corresponding to slice cuts across the x -axis poles for pole cloud in Fig. 1.

TABLE I
SUMMARY OF CLASSIFICATION AND ESTIMATION METHODS

	Classifier	Classification method	Estimation method
Feature Space	Mahalanobis	$\hat{i} = \arg \min_i (\hat{p}_i - \bar{p}_i)^T C_i^{-1} (\hat{p}_i - \bar{p}_i)$	$\hat{p}_i, \hat{\nu}_i = \arg \min_{p, \nu} \ C^{-\frac{1}{2}} (y - m(p, \nu))\ _2^2$
	KDE	$\hat{i} = \arg \max_i f_i(\hat{p}_i)$	$\hat{p}_i, \hat{\nu}_i = \arg \min_{p, \nu} \ C^{-\frac{1}{2}} (y - m(p, \nu))\ _2^2 - \ln f_i(p)$
	SVM	$\text{sgn}(\sum_{s_i \in \text{support vectors}} \alpha_i y_i s_i \hat{p}_i + b)$	$\hat{p}_i, \hat{\nu}_i = \arg \min_{p, \nu} \ C^{-\frac{1}{2}} (y - m(p, \nu))\ _2^2$
Signal Space	ML-GLRT	$\hat{i} = \arg \max_i f(y/H_i, \hat{\theta})$	$\hat{\theta} = \arg \min_{\theta} \ C^{-\frac{1}{2}} (y - m(\theta))\ _2^2$
	MAP-GLRT	$\hat{i} = \arg \max_i f(y/H_i, \hat{p}_i, \hat{\nu}_i) f_i(\hat{p}_i)$	$\hat{p}_i, \hat{\nu}_i = \arg \min_{p, \nu} \ C^{-\frac{1}{2}} (y - m(p, \nu))\ _2^2 - \ln f_i(p)$
	LA-GLRT	$\hat{i} = \arg \max_i \left\{ -\frac{1}{2} \ln(\det(J)) + \ln f_i(\hat{p}_i) + \ln f(y/\hat{p}_i, \hat{\nu}, H_i) \right\}$	$\hat{p}_i, \hat{\nu}_i = \arg \min_{p, \nu} \ C^{-\frac{1}{2}} (y - m(p, \nu))\ _2^2 - \ln f_i(p)$

Since $-\ln f_i(p) \geq 0$, we can write (15) as

$$\hat{p}_i, \hat{\nu}_i = \arg \min_{p, \nu} \left\| C^{-1/2} (y - m(p, \nu)) \right\|_2^2 + \left(\sqrt{-\ln f_i(p)} \right)^2. \quad (16)$$

Thus, the optimization problem can be implemented using the nonlinear least square estimation. In this paper, we have used the Levenberg–Marquardt (LM) algorithm [33]. Gradients were computed using a finite difference approximation, and a simple 1-D search was used to find the regularization parameter in the LM method [33]. A coarse grid search in parameter space was employed to find good initial values for the optimization methods. Although more sophisticated initialization methods and exact gradient calculations could be used (particularly, if one were to transition these ideas into an operational system), for purposes of demonstrating the effectiveness of the methods that we propose in this paper, these simple methods prove to be sufficient.

B. Optimal Classification

It is known that the classification rule which minimizes the probability of error is the Bayes decision rule [28]

$$\hat{i} = \arg \max_i f(H_i/y) \quad (17)$$

where $f(H_i/y)$ is the posterior probability distribution of i th hypothesis. Using the Bayes rule, (17) can be written as

$$\hat{i} = \arg \max_i \frac{f(y/H_i) f(H_i)}{f(y)} \quad (18)$$

where $f(H_i)$ is the prior distribution of each hypothesis and is assumed to be uniform in this paper. Expanding $f(y/H_i)$ in terms of priors using the Bayesian rule and ignoring the constant denominator, the optimal Bayes classifier can be written as

$$\hat{i} = \arg \max_i \int f(y/H_i, \theta) f_i(\theta) d\theta. \quad (19)$$

where $f_i(\theta) = f_i(p)f(\nu)$ is the prior probability distribution of parameters, assuming a uniform distribution for the nuisance parameters. The vector $\theta = [p^T, \alpha^T, r^T, c^T, \delta^T]^T = [p^T, \nu^T]^T$ has 13 parameters which makes the integration in (19) almost impossible numerically. Although the Monte Carlo methods can be used to evaluate (19), they are usually quite time consuming. Therefore, we investigate other solutions in this paper.

As shown in (19), the optimal classifier operates on the entire signal y . It is known that if the model parameters (such as poles in our case) are sufficient statistics [34], then the optimal classification can be done in a lower dimensional feature space. Unfortunately, for the UXO, problem poles are not sufficient statistics [16] and, therefore, do not lead to optimal classifiers. Despite this theoretical shortcoming, these parameters are still widely used as the basis for solving the UXO classification problem due to convenience and the success of such methods in practice where other issues (such as the inaccuracies of the underlying dipole model) become important.

Based on the aforementioned discussion, we see that classifiers are divided into two groups: signal-based or feature-based, where the former employs the entire received signal for classification and the latter only makes use of the feature parameters. Both approaches will be considered in this section.

C. Signal-Space Classifiers

1) *Generalized Likelihood Ratio Test (GLRT)*: The GLRT classifier [35] approximates the optimal classifier in (19) using the estimated values for the parameters. We have implemented two forms of the GLRT. The first is based on an ML approach both for the estimation of the parameters as well as the classification. We refer to this method as the ML-GLRT, and it takes the form

$$\hat{i} = \arg \max_i f(y/H_i, \hat{\theta}) \quad (20)$$

$$\hat{\theta} = \arg \min_{\theta} \left\| C^{-\frac{1}{2}} (y - m(\theta)) \right\|_2^2. \quad (21)$$

The second approach uses the prior models developed in Section III both for purposes of estimation as well as classification. As such, it is a MAP technique and is referred to as the MAP-GLRT

$$\hat{i} = \arg \max_i f(y/H_i, \hat{p}_i, \hat{\nu}_i) f_i(\hat{p}_i) \quad (22)$$

where

$$\hat{p}_i, \hat{\nu}_i = \arg \min_{p, \nu} \left\| C^{-\frac{1}{2}} (y - m(p, \nu)) \right\|_2^2 - \ln f_i(p).$$

Taking the logarithm of (22) and assuming the noise is Gaussian, we have

$$\hat{i} = \arg \min_i \left\{ \left\| C^{-\frac{1}{2}} (y - m(\hat{p}_i, \hat{\nu}_i)) \right\|_2^2 - \ln f_i(\hat{p}_i) \right\}. \quad (23)$$

We note that the first term can be interpreted as the residual and extends the feature-based kernel classifier (see Section IV-D1) to a signal-space classifier.

2) *Laplacian Approximated Classifier (LA-GLRT)*: The Laplacian approximation [36] is used to approximate the posterior distribution in (19). It is based on the asymptotic Gaussian approximation of the likelihood in (19) and can be written as

$$f(H_i/y) \propto \int f(y/H_i, \theta) f_i(\theta) d\theta \approx \frac{(2\pi)^{\frac{q}{2}}}{\sqrt{\det(J)}} f_i(\hat{\theta}) f(y/\hat{\theta}, H_i) \quad (24)$$

where q is the number of parameters, $\hat{\theta}$ is the ML or MAP estimation of parameter θ , and J is the observed Fisher information matrix (OFIM) [37] defined as

$$J_{ij} = -\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(y/\theta) |_{\theta=\hat{\theta}}. \quad (25)$$

Taking the logarithm of (24), we will have a classifier based on the Laplacian approximation which can be written as

$$\hat{i} = \arg \max_i \left\{ -\frac{1}{2} \ln (\det(J)) + \ln f_i(\hat{p}) + \ln f(y/\hat{\theta}, H_i) \right\}. \quad (26)$$

This is the same as the MAP-GLRT classifier with the exception of the first term on the right-hand side of (26), which is related to the accuracy of the underlying parameter estimates. More specifically, for the case where the noise is taken as additive and Gaussian, the OFIM can be interpreted as an approximation to the Hessian of the residual function $\|C^{-1/2}(y - m(\hat{\theta}))\|_2^2$, which is evaluated at the MAP estimate of the parameters. Thus, for example, when the parameter estimates are quite accurate, the Hessian would be “small” and the first term on the right-hand side of (26) would be large, thereby adding to the evidence that the corresponding hypothesis is correct.

D. Feature-Space Classifiers

1) *Kernel Classifier (KDE)*: As an approximation of the optimal feature-space classifier, the kernel classifier can be written as

$$\hat{i} = \arg \max_i f_i(\hat{p}_i) \quad (27)$$

where f_i is the pdf of the poles of the i th object, and \hat{p}_i is the estimated vector of poles, assuming that the i th object is the true object. In [24], a feature-space classifier, which was based on the Gaussian assumption of pole distribution, was used where the poles were estimated using the ML estimation. The method reduces to the Mahalanobis classifier in (6). By using the kernel density estimated prior (7), we can estimate the poles using the MAP estimator and implement (27) directly

$$\hat{i} = \arg \max_i f_i(\hat{p}_i) \quad (28)$$

where

$$\hat{p}_i, \hat{\nu}_i = \arg \min_{p, \nu} \left\| C^{-\frac{1}{2}} (y - m(p, \nu)) \right\|_2^2 - \ln f_i(p).$$

Compared to the GLRT data-space classifier, this classifier can be considered the GLRT “approximation” to the optimal feature-based classifier.

2) *Support Vector Machine*: SVM is based on structured risk minimization, which was introduced in [38], and has been extensively considered recently. It is a linear two category classifier which maximizes the margin of a two hypothesis space [39]. In [11], the authors used SVM for a UXO detection problem which is a binary detection problem. Here, we have applied SVM for multiclass UXO discrimination, employing algorithms for the extension of SVM to multiclass case. Given a set of data points which belong to either of the two classes, SVM finds the optimal separating hyperplane that minimizes the risk of misclassifying the training samples and unseen test samples. Given a set of test points x_i and corresponding target class y_i and assuming a linear classifier of the form $x \cdot w + b$, we will have

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad \forall i \quad (29)$$

where w is the norm of hyperplane that divides the hypothesis space. To maximize the margin of classifier, we need to minimize this norm, leading to the following problem

$$\min \|w\|_2^2 \quad \text{subject to} \quad y_i(x_i \cdot w + b) - 1 \geq 0 \quad \forall i. \quad (30)$$

The equivalent quadratic programming problem is to determine the minimum number of the Lagrangian function

$$L_p = \frac{1}{2} w \cdot w - \sum_{i=1}^N \alpha_i (y_i(w \cdot x_i + b) - 1) \quad (31)$$

with $L_p = L(w, b, \alpha)$. The final classifier then can be explained in terms of support vectors

$$f(x) = \text{sgn} \left(\sum_{s_i \in \text{support vectors}} \alpha_i y_i s_i x + b \right). \quad (32)$$

For nonseparable test data, the solution is a generalized optimal separating hyperplane

$$\min \frac{1}{2} w \cdot w + x \sum_{i=1}^N \xi_i, \quad i = 1, \dots, N, \quad \xi \geq 0. \quad (33)$$

The data, which are not linearly separable, are mapped to a higher dimensional space, and kernel functions are used to lower the complexity of computation in this high-dimensional space. Kernels operate on low-dimensional space and give the same result for the inner product in high-dimensional space. Thus, if the mapping is $\phi(x)$ for kernel k , we will have

$$k(x_i, x_j) = \phi(x_i) \phi(x_j). \quad (34)$$

There are different kernels that can be used, and the most common are the following:

- 1) simple dot product kernel

$$k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad (35)$$

- 2) radial basis function

$$k(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2) \quad (36)$$

- 3) two-layer neural network

$$k(x, y) = \tanh(kx \cdot y - \delta). \quad (37)$$

Although SVM is defined for two class case, it can be used for multiclass case. Several methods were proposed in [40] and [41]. The most commonly used methods are one-against-all and one-against-one. In the former method, the i th class is trained by labeling all the samples in the i th class as positive and the rest as negative. For an N -class problem, we will have N classifiers. The one-against-one method constructs $N(N-1)/2$ classifiers consisting of all pairwise combination of classes. To combine the classifiers, different algorithms can be used; the most common is the “max wins” algorithm, where the final decided class is the one with maximum votes [40]. There is another method proposed in [41], where graph theory is used to combine the classifiers. We have used the one-against-one algorithm, in which the poles were estimated using the ML estimation and Gaussian kernels were used. We note that we cannot use the MAP estimation for this case as we have to have one pole estimate in each pairwise comparison. Table I summarizes the classification methods and the corresponding estimation algorithms.

V. CLASSIFICATION RESULTS

Classification performance was evaluated using the simulated and real data. For simulation, the responses of the four targets are simulated for a GEM3 sensor [42] that is sampled on a 1-m² area using an equally spaced 5×5 grid of measurement points. The objects include the following: a 3-in long by 1-in diameter stainless steel cylinder (S1), a 6-in long by 1-in diameter stainless steel cylinder (S2), a 3-in long by 1-in diameter aluminum cylinder (A1), and a 6-in long by 1-in diameter aluminum steel cylinder (A2). The target responses were generated with a four-pole per axis dipole model in which the sums in (3) are terminated after the four terms [21], [24]. The poles used for this simulation were generated using the methods in [21]. The objects were chosen such that the pole distribution covers most of the feature space with similar features for the pair (S1, S2) and (A1, A2) to make the classification more challenging. Fig. 3 shows the pole distributions for these objects. Classification performance was evaluated at different SNR values using 100 runs of a Monte Carlo for each case. The SNR is computed as

$$\text{SNR} = 10 \log_{10} \left(\frac{\|y\|^2}{N\sigma^2} \right) \quad (38)$$

where y is the received signal, N is the length of y , and σ^2 is the noise variance. The probability of error is calculated using

$$p_e = \frac{1}{4} \sum_{i=1}^4 P(\text{error}/H_i) \quad (39)$$

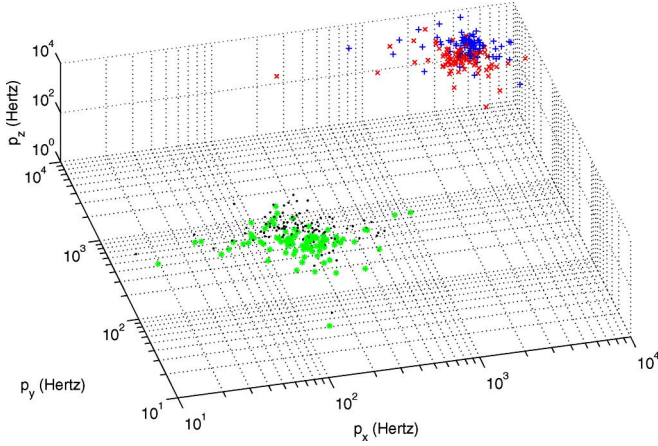


Fig. 3. Pole distributions for S1 (+), S2 (x), A1 (·), and A2 (*) objects (see text for a physical description of targets).

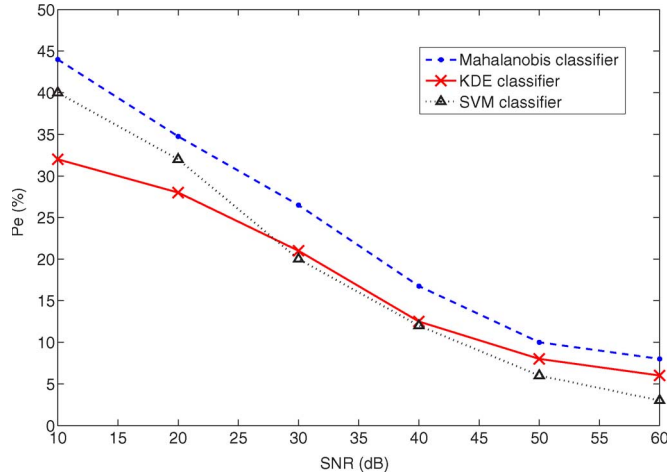


Fig. 4. Probability of error as a function of SNR for feature-based classifiers.

which corresponds to the average sum of the off-diagonal elements of confusion matrix. Figs. 4 and 5 show the probability of error as a function of SNR for the feature and signal classifiers, respectively. Comparing the plots, we see that the signal classifiers (GLRT and Laplacian) have considerably better performance compared to the feature-space classifiers (Mahalanobis, kernel density classifier (KDE), and SVM). In addition, the probability of error decreases rapidly as SNR increases for the signal-space classifiers comparing to the feature-space classifiers.

Table II shows the confusion matrices for the Mahalanobis, KDE, SVM, MAP-GLRT, and LA-GLRT classifiers. It is shown that only the Mahalanobis classifier confuses a steel target with an aluminum target. The Laplacian approximated classifier has the best performance and shows an 80% improvement over simple mean covariance at classifier used in [24] at moderate and high SNRs and about 40% improvement at low SNRs. As we expected, the performance of Laplacian classifier gets better as the SNR increases. This can be explained in a couple of ways. Most directly, as the SNR decreases, the Gaussian approximation to the integrand of (24) becomes less valid. Alternatively, as the SNR decreases, we are seeing a corresponding decrease in accuracy in the estimation of the model parameters. Roughly

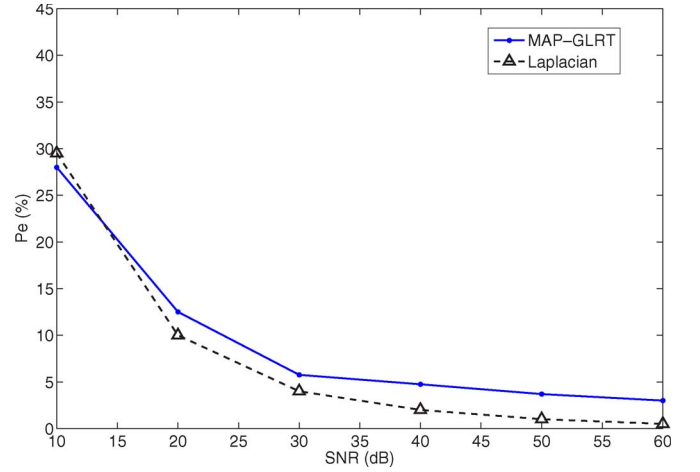


Fig. 5. Probability of error as a function of SNR for signal-based classifiers.

TABLE II
CONFUSION MATRICES FOR THE SIMULATED TARGETS AT SNR OF 30 dB

Mahalanobis classifier					KDE				
	S1	S2	A1	A2		S1	S2	A1	A2
S1	66	34	0	0	S1	72	28	0	0
S2	44	56	0	0	S2	28	72	0	0
A1	1	6	77	16	A1	0	0	72	28
A2	1	4	0	95	A2	0	0	0	100

SVM					MAP-GLRT				
	S1	S2	A1	A2		S1	S2	A1	A2
S1	78	21	0	0	S1	100	0	0	0
S2	50	45	5	0	S2	8	92	0	0
A1	0	0	98	2	A1	0	0	100	0
A2	0	0	0	100	A2	0	0	15	85

LA-GLRT				
	S1	S2	A1	A2
S1	100	0	0	0
S2	2	98	0	0
A1	0	0	93	7
A2	0	0	7	93

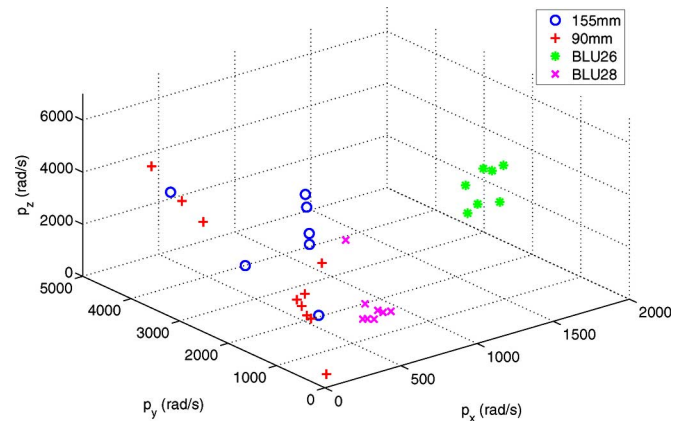


Fig. 6. Pole distributions for 155-mm (o), 90-mm (+), BDU-26 (*), and BDU-28 (x) UXO items.

speaking, as the parameter estimates worsen, the OFIM gets larger, indicating a worse fit of the model to the data, and the $-\ln(\det(J))$ term in (26) becomes increasingly negative, thereby making the corresponding hypothesis less likely.

TABLE III
CONFUSION MATRICES SUMMARIZING THE CLASSIFICATION OF SEVERAL UXO TARGETS FROM THE ERDC
TEST STAND DATA SET. CLASSIFICATION WAS PERFORMED USING A LEAVE-ONE-OUT METHOD

Mahalanobis classifier				
	26mm	28mm	90mm	155mm
26mm	6	0	1	0
28mm	2	6	0	0
90mm	0	1	3	6
155mm	0	0	0	7
Percent Error = 31%				
SVM				
	26mm	28mm	90mm	155mm
26mm	5	0	2	0
28mm	0	5	3	0
90mm	0	0	10	0
155mm	0	3	4	0
Percent Error = 38%				
ML-GLRT				
	26mm	28mm	90mm	155mm
26mm	0	1	6	0
28mm	0	8	0	0
90mm	0	0	5	5
155mm	0	0	4	3
Percent Error = 50%				
KDE				
	26mm	28mm	90mm	155mm
26mm	6	0	0	0
28mm	1	7	0	0
90mm	0	1	4	5
155mm	0	0	0	7
Percent Error = 25%				
MAP-GLRT				
	26mm	28mm	90mm	155mm
26mm	6	0	0	1
28mm	1	6	1	0
90mm	1	1	3	5
155mm	1	1	0	5
Percent Error = 31%				
LA-GLRT				
	26mm	28mm	90mm	155mm
26mm	6	0	0	1
28mm	1	6	1	0
90mm	1	1	3	5
155mm	1	1	0	5
Percent Error = 31%				

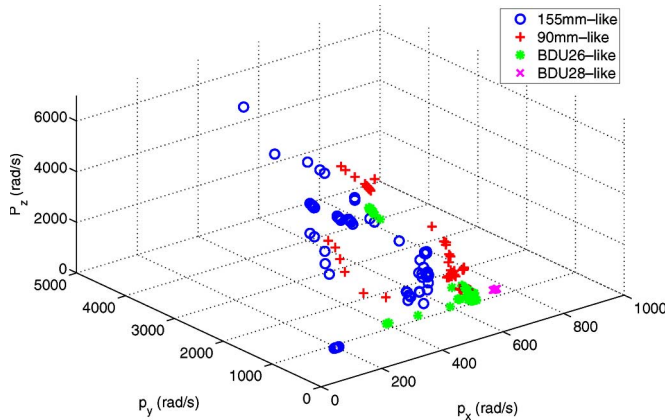


Fig. 7. Pole library estimated from the data simulated with the standardized excitation model. Targets in this library are modeled after the 155-mm (\circ), 90-mm ($+$), BDU-26 ($*$), and BDU-28 (\times) UXO items.

Classification of the field data was undertaken with the EMI data collected using an EM63 sensor over various UXOs and metallic clutters surveyed at the U.S. Army Corps of Engineers Engineer Research and Development Center (ERDC) UXO/Countermines test stand. The test stand allows precise positioning of the target and the sensor in terms of both relative location and orientation. The data were collected over a variety of spatial grids, ranging in spacing from 5–20 cm. Often, several spacings were present in a single grid, with higher sampling density near the target where the gradient of the target field was largest. A suite of 18 standard UXO targets was examined, as well as several clutter items and spheres of assorted compositions. In this paper, submunitions of diameters 26 and 28 mm (BDU-26 and BDU-28, respectively) are examined, as well as two mortar shells with diameters of 90 and 155 mm. These four targets were chosen because of the relatively large number of separate data sets collected compared to the other targets in the suite. The data sets differ from one another in target pose, orientation, and target instance. The accuracy and statistical significance of the KDE pdf increases

as the number of parameter estimates increases. It should be noted, however, that even for these four targets, the number of samples used to estimate the pdf was small. In order to minimize the contribution of the background response to the variability of the pole cloud, and therefore to the pdf estimation, the background response was subtracted from the raw data set. Subtraction was accomplished by assuming that the measured data at a sufficient distance from the target were attributable solely to the background. Random noise was suppressed using the eigenimage decomposition method described in [43], wherein the uncorrelated noise is removed by reconstructing the data from the left and right eigenvectors associated with the largest singular values.

Fig. 6 shows the pole distribution of these objects. A leave-one-out method was employed, in which all data sets, but one for a particular target, are used to construct a target library, along with the data from all other targets. Classification is then attempted using the remaining data set. The noise covariance matrix to be used in the classification algorithms was estimated using the residual snapshots over spatial domain after estimating the parameters using the unit variance noise. The results are summarized in Table III as confusion matrices for each algorithm. Because of low training samples, the SVM classifier did not work well for this case, demonstrating the superiority of the KDE-based classifier when few training samples are available.

Comparing Tables II and III, we note that the classifiers considered in Table III, although not “optimal” under the assumptions used to construct (1), are certainly more robust in practice to these assumptions being violated. It is well-known that the dipole model used to construct the simulated data is at best an approximation. This is particularly true for the larger targets which are not uniformly illuminated by the source magnetic field and are furthermore best represented by a large number of terms in the dipole principal axis polarizability function. Thus, even when the model parameter estimates are accurate, the difference between the y and m will not be uncorrelated Gaussian noise, and $\|C^{-1/2}(y - m(p, \hat{v}))\|_2^2$ can be

TABLE IV
CONFUSION MATRICES SUMMARIZING THE CLASSIFICATION OF FOUR UXO-LIKE TARGETS SIMULATED USING A STANDARDIZED EXCITATION MODEL. CLASSIFICATION WAS PERFORMED USING A LEAVE-ONE-OUT METHOD

Mahalanobis classifier				
	26mm	28mm	90mm	155mm
26mm	19	0	3	2
28mm	0	20	4	0
90mm	1	0	5	18
155mm	0	0	0	24
Percent Error = 29%				
SVM				
	26mm	28mm	90mm	155mm
26mm	22	1	1	0
28mm	0	24	0	0
90mm	0	0	6	18
155mm	0	0	3	20
Percent Error = 24%				
ML-GLRT				
	26mm	28mm	90mm	155mm
26mm	0	24	0	0
28mm	0	24	0	0
90mm	6	15	0	3
155mm	0	17	0	7
Percent Error = 67%				
KDE				
	26mm	28mm	90mm	155mm
26mm	22	0	0	2
28mm	0	24	0	0
90mm	0	0	19	5
155mm	0	0	2	22
Percent Error = 9.4%				
MAP-GLRT				
	26mm	28mm	90mm	155mm
26mm	17	0	4	3
28mm	0	24	0	0
90mm	2	0	16	6
155mm	6	0	6	12
Percent Error = 28%				
LA-GLRT				
	26mm	28mm	90mm	155mm
26mm	17	0	4	3
28mm	0	24	0	0
90mm	2	0	16	6
155mm	6	0	6	12
Percent Error = 28%				

substantial. It is the presence of this non-Gaussian deterministic “noise” (attributable to model mismatch) that also accounts for the reduced performance of the signal-space classifiers in the real data, particularly for the larger targets. The assumption that the residual is Gaussian plays a key role in the GLRT classifiers. When that assumption is not valid, the GLRT classifiers break down. On the other hand, as the dipole model becomes less applicable, the estimated poles form a more and more diffuse cloud. If the clouds remain distinct and well-defined for each target, the feature-space classifiers will continue to function well. In the presence of an unmatched model (almost a certain constraint if rapid processing is necessary), an optimality of the classifier appears to be less important than robustness to the underlying data generation mechanism. In contrast, when model mismatch is less significant, as in the case of the data simulated using a four-term per axis dipole model (Fig. 3), higher classification performance is achieved using the data-space classifiers.

Although the performance of our classification methods when applied to real data deteriorates in comparison to our simulation study, we should point out that the results obtained here are as good as or better than most classification algorithms in the community. The fact that our approach is to be able to localize and classify the object at the same time adds more to its value, while most classification approaches assume that geometrical information of the object is available at the classification stage: an assumption which is not valid for the real field data scenarios. For example, in [16], a very simple dipole model is used as the basis of processing. Classification results using the simulations for simple targets for the case of unknown coefficient factors at an SNR of about 40 dB shows the probability of correct classification of about 50%. Results using the real field data shows the probability of detection of 50% at a false alarm of 20%. In addition, the geometry of the target (location and orientation in space) is assumed to be known. In [44], the effect of unknown geometry on the classification is considered by performing classification in separate scenarios with horizontal, vertical, or depth uncertainty

of the object. It is assumed that orientation is known. Noise is assumed to be white Gaussian noise. Using the simulations for optimal processing, the probability of correct classification is 70%, and for the suboptimum classifier, this probability is about 20%. Soil effect is proven to be negligible in the processing. Again, for small sets of shallow land mine field data, the classification performance for the optimum classifier, which involves integrations over location, is about 70% with known orientation. It should also be noted that the model used is based on the method of moments which accurately models the target, contrasted with our approximate dipole model. In order to lend credence to the conclusions drawn previously, a second simulated data set was generated. This data set was computed from the standardized excitation approach (SEA) of Schubitz *et al.* [4]. The SEA model computes the response of an arbitrary target oriented at any arbitrary orientation and positioned at any arbitrary location. Four targets were modeled with physical and electromagnetic properties similar to the aforementioned four UXO targets. A total of 217 data sets were generated for each target, with the target oriented at every 22.5° in azimuth, and every 45° in dip for each of the three burial depths typical of the target being modeled. The remaining 25 data sets were generated at a random orientation and burial depth (within a target-specific range). Given a large enough library, and thus a well-defined estimate of the target density function, the effects of model mismatch on classification should be clarified. The library poles are shown in Fig. 7. As expected, the larger targets (similar to the 155- and 90-mm UXO items), for which the dipole model is less valid, have more diffuse library pole clouds, while the smaller targets (particularly, the BDU-28-like target) have more compact distinct pole clouds. In fact, the simulated library appears quite similar in morphology to the actual UXO test stand data with the exception of the BDU-26 target. Despite this discrepancy, the conclusions drawn from the simulated data are applicable to the test stand data as well. The 100 random data sets were then classified against this library using the leave-one-out method described previously. The results are summarized in Table IV. The conclusions drawn

TABLE V
CONFUSION MATRICES SUMMARIZING THE CLASSIFICATION OF FOUR UXO-LIKE TARGETS SIMULATED USING A STANDARDIZED EXCITATION MODEL AND A SPARSE LIBRARY. CLASSIFICATION WAS PERFORMED USING A LEAVE-ONE-OUT METHOD. ABSENT VALUES REPRESENT A BREAKDOWN IN THE CLASSIFICATION ALGORITHM AND ARE INCLUDED IN THE CALCULATION OF PERCENT ERROR

Mahalanobis classifier				
	26mm	28mm	90mm	155mm
26mm	20	0	3	2
28mm	0	25	0	0
90mm	0	0	8	17
155mm	0	0	9	16
Percent Error = 31%				
SVM				
	26mm	28mm	90mm	155mm
26mm	4	19	1	0
28mm	0	24	0	0
90mm	0	0	7	17
155mm	1	0	9	14
Percent Error = 48%				
ML-GLRT				
	26mm	28mm	90mm	155mm
26mm	12	13	0	0
28mm	0	25	0	0
90mm	11	8	0	6
155mm	8	8	0	9
Percent Error = 54%				
KDE				
	26mm	28mm	90mm	155mm
26mm	20	0	3	2
28mm	0	25	0	0
90mm	0	0	5	17
155mm	0	0	2	16
Percent Error = 34%				
MAP-GLRT				
	26mm	28mm	90mm	155mm
26mm	20	0	3	2
28mm	0	25	0	0
90mm	1	4	6	14
155mm	5	2	5	13
Percent Error = 36%				
LA-GLRT				
	26mm	28mm	90mm	155mm
26mm	20	0	3	2
28mm	0	25	0	0
90mm	1	4	6	14
155mm	5	2	5	13
Percent Error = 36%				

from the test stand UXO data are borne out: The classification performance for the signal-space classifiers is not as good as the performance of the feature-space classification methods, indicating that the feature-space classification methods are more robust with respect to the model mismatch. Here, again, the poor performance of the signal-space classifier is attributable to the violation of the assumption that the residual is Gaussian. It should be noted that the best feature-based classifier using the dipole model achieves an error rate of about 10% on SEA data which is a very strong result. If we accept that the full library SEA results extrapolate to what we would see if the ERDC data set had been complete, then certainly, this performance indicates the very strong potential of the method for the field data. Finally, it should be noted that because the failure of the signal-space classifier is due to the poor representation of the target by the dipole model, the signal-space classifier performance may be improved simply by introducing a more accurate model. However, the issue of model mismatch will remain so long as the requirement for rapid data processing is imposed. In order to test the conclusion that sparseness of the library training samples will negatively impact the classification, the simulated data library was randomly decreased to 15 samples. Classification of the 100 randomly oriented and positioned data sets was attempted using the decimated library. The results are provided in Table V. Our conclusions are once again supported: The performance of the classification algorithms is strongly correlated with the training sample size. Because the pdfs that describe the target-specific parameters are estimated from the library, a smaller sample size results in a less accurate pdf description and, thus, reduced classification performance. The spread of the poles due to the model mismatch is not well captured, leading to a degradation in performance of the feature-space classifiers as well as the signal-space classifiers. Not surprisingly, the SVM classifier, which relies upon the completeness of the training data, shows a significant degradation in performance as the library is made more sparse. The library size is similar to the size of the test stand data library.

As expected, the percent error of classification is comparable to the error incurred in the classification of the test stand data.

VI. CONCLUSION

To achieve robust methods for the classification of UXO, a collection of classification methods based on the use of nonparametric prior models for the target features was proposed and compared in this paper. KDE methods formed the basis for the construction of these prior models. These priors allowed for the development of a collection of MAP-based estimation and classification schemes. The GLRT-type methods, the approximation to the statistically optimal Bayes classifier for our problem, and the SVM method were also considered.

It was shown that the incorporation of prior information generally resulted in a significant performance improvement that is relative to the previous proposed techniques for both the simulated and field data. Interestingly, the classifiers which demonstrated the strongest improvements differed depending on whether the simulated or field data were being tested. In the case of simulated data, the model underlying the processing was quite similar to that used in the creation of the data. Hence, the approximately optimal signal-space classifiers worked quite well. For the field data, where the dipole model is far more of an approximation to reality, the approximately optimal methods failed to perform as well, while the techniques regarded as suboptimal were far more robust to the mismatch in model. This was particularly true for the larger targets, for which the model mismatch was more of an issue. The optimal signal-space classifiers rely on an underlying assumption that the best fit residual is Gaussian in nature. In reality, the best fit residual is not Gaussian and retains the mismatched portion of the data in addition to any additive noise. The hypothesis was successfully tested by simulating a large set of data for several targets using a model which, although also approximate, provided a greater degree of model mismatch. The decrease in performance of the classification of field data versus the

simulated data may also be attributed to the small size of the set of prior pole estimates, leading to an inaccurate definition of the prior pdf. This hypothesis was also verified using a set of unmatched simulated data.

In conclusion, the feature-space classifiers are more robust to model mismatch as mismatch manifests as additional spread in the model parameter distribution and as a non-Gaussian contribution to the best fit residual. However, if the target parameter clouds remain distinct and well defined, as demonstrated with the data simulated using the SEA model, the feature-space classification will continue to work well. Classification may, of course, be improved by the introduction of a more accurate physical model than the dipole model, which is examined here because of its simplicity and ubiquity. Alternatively, the true physics of the target may be approximated by modeling the target as a spheroid, using a high fidelity spheroidal model, such as the mean field approach developed by Weichman and Eugene [8]. The mean field technique is also advantageous because the scattered field is related to more accessible physical parameters such as axis length and conductivity rather than bulk parameters like the poles examined in this paper. Ideally, an arbitrary heterogeneous UXO-like target may be modeled by implementing an approach such as the SEA model [4] developed by Shubitize *et al.* Unfortunately, the feasibility of these approaches is limited by the required computation expense. For example, roughly 1 h on a standard desktop PC is required to compute the EM63 target response over a $3\text{ m} \times 3\text{ m}$ grid with a 10-cm spacing using the SEA approach. As greater computational resources become available, these approaches will become more accessible. However, parameter estimation will remain a sensitive problem, and techniques, such as those presented here, which accurately capture and utilize the variable nature of model parameter estimates will be important.

It is important to understand the effect of model mismatch because the constraints of the UXO problem dictate that the model must necessarily be approximate in order to tractably investigate the large areas contaminated by UXO. We have shown that reasonable violations of the assumptions of an approximate model can be handled statistically, particularly when the effect of those violations on the estimation of parameters is well understood prior to classification. To put it in another way, the difficult issue of model mismatch may be overcome, but doing so requires a significant investment in developing a rich library that describes the variation in identifying parameters well and as completely as possible. In effect, the statistical methods introduced here provide a means to extend the life of the crude yet well-understood and widespread dipole model, yielding good classification results.

In addition to the processing methods discussed here, powerful performance analysis methods incorporating these prior models can be used to bound the performance of both the estimation as well as classification methods. More specifically, Bayesian Cramer–Rao lower bounds can be developed which gives the variance in our estimates of the target parameters a lower bound. Similarly, the Chernoff bound can be formed to provide an upper bound on the probability of error for the Bayes-optimal feature-based as well as databased classifiers. These analytical bounds also enable us to see the effect of

different system parameters such as sensor dimension, sampling density, etc., on the system performance. Our future work on this project will be directed to the analytical development and use of these bounds in optimizing sensor design and deployment.

ACKNOWLEDGMENT

The authors would like to thank Dr. F. Shubitize of the Thayer School of Engineering, Dartmouth College, for making the SEA code used in Section V available to us. The authors would also like to thank the three reviewers and the Associate Editor for the very helpful comments which have substantially improved and sharpened the focus of this paper.

REFERENCES

- [1] K. Sun, K. O'Neill, F. Shubitidze, I. Shamatava, and K. Paulsen, "Theoretical analysis and range of validity of TSA formulation for application to UXO discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 9, pp. 1871–1881, Sep. 2004.
- [2] K. Sun, K. O'Neill, F. Shubitidze, I. Shamatava, and K. Paulsen, "A modified dipole model for target discrimination in EMI sensing," in *Proc. Antennas and Propag. Soc. Int. Symp.*, Jul. 2005, vol. 1A, pp. 173–176.
- [3] K. Sun, K. O'Neill, F. Shubitize, I. Shamatava, and K. D. Paulsen, "Fast data-derived fundamental spheroid excitation models with application to UXO discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp. 2573–2583, Nov. 2005.
- [4] F. Shubitize, K. O'Neill, I. Shamatava, K. Sun, and K. D. Paulsen, "A fast and accurate representation of physically complete EMI response by a heterogeneous object to enhance UXO discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1736–1750, Aug. 2005.
- [5] F. Shubitidze, K. O'Neill, K. Sun, and I. Shamatava, "Application of broadband EMI responses to infer buried object's aspect ratio," in *Proc. IEEE Int. Geosci. and Remote Sens. Symp.*, Jun. 2002, vol. 3, pp. 1542–1545.
- [6] K. Sun, K. O'Neill, L. Liu, F. Shubitidze, and I. Shamatava, "Application of Bayesian inversion for scatterer shape from EMI data," in *Proc. Antennas and Propag. Soc. Int. Symp.*, Jun. 2003, vol. 1, pp. 547–550.
- [7] F. Shubitidze, B. Barrowes, K. O'Neill, and I. Shamatava, "Advanced EMI modeling and processing approaches for UXO discrimination," in *Proc. Int. Conf. Math. Methods Electromagn. Theory*, Jun. 2006, pp. 126–131.
- [8] P. Weichman and L. Eugene, "Study of inverse problems for buried UXO discrimination based on EMI sensor data," *Proc. SPIE*, vol. 5089, pp. 1189–1200, Sep. 2003.
- [9] L. Carin, H. T. Yu, and C. E. Baum, "On the wideband electromagnetic inductance signature of conducting and permeable targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1206–1213, Jun. 2001.
- [10] S. J. Norton and I. J. Won, "Identification of buried unexploded ordnance from broadband electromagnetic induction data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2253–2261, Oct. 2001.
- [11] Y. Zhang, L. M. Collins, H. T. Yu *et al.*, "Sensing of unexploded ordnance with magnetometer and induction data: Theory and signal processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1005–1015, May 2003.
- [12] H. Huang and I. J. Won, "Characterization of UXO-like targets using broadband electromagnetic induction sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 3, pp. 652–663, Mar. 2003.
- [13] L. M. Collins, P. Gao, and L. Carin, "An improved Bayesian decision theoretic approach for land mine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 811–819, Mar. 1999.
- [14] P. Gao and L. M. Collins, "A theoretical performance analysis and simulation of time-domain EMI sensor data for land mine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 4, pp. 2042–2055, Jul. 2000.
- [15] L. M. Collins, P. Gao, and D. Schofield, "A statistical approach to landmine detection using broadband electromagnetic induction data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 950–962, Apr. 2002.
- [16] S. L. Tatum and L. M. Collins, "A comparison of algorithms for subsurface target detection and identification using time-domain electromagnetic induction data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1299–1306, Jun. 2001.

- [17] J. T. Smith and H. F. Morrison, "Estimating equivalent dipole polarizabilities for the inductive response of isolated conductive bodies," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1208–1214, Jun. 2004.
- [18] B. Barrow and H. H. Nelson, "Model-based characterization of electromagnetic induction signatures obtained with the MTADS electromagnetic array," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1279–1285, Jun. 2001.
- [19] T. H. Bell, B. Barrow, and J. T. Miller, "Subsurface discrimination using electromagnetic induction sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1286–1293, Jun. 2001.
- [20] G. Sower, J. Endsley, and E. Christy, "Discrimination of metal land mines from metal clutter: Results of field tests," in *Proc. SPIE—Conf. Detection Mines and Minelike Targets*, Apr. 1999, pp. 78–88.
- [21] N. Geng, C. E. Baum, and L. Carin, "On the low-frequency natural response of conducting and permeable targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 347–359, Jan. 1999.
- [22] L. Pasion and D. Oldenburg, "A discrimination algorithm for UXO using time-domain electromagnetic induction," *J. Environ. Eng. Geophys.*, vol. 6, pp. 91–102, 2001.
- [23] L. S. Riggs and J. E. Mooney, "Identification of metallic mine-like objects using low frequency magnetic fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 1, pp. 56–66, Jan. 2001.
- [24] A. B. Tarokh, E. L. Miller, I. J. Won, and H. Huang, "Statistical classification of buried objects from time and frequency domain EMI data," *Radio Sci.*, vol. 39, no. RS4S05, 2004.
- [25] E. L. Miller, "On some options for statistical classification of buried objects from spatially sampled time or frequency domain EMI data," in *Proc. SPIE—Conf. Detection Mines and Minelike Targets*, Apr. 2001, vol. 4394, pp. 97–107.
- [26] H. Stark and J. W. Woods, *Probability and Random Processes With Applications to Signal Processing*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [27] M. L. Burrows, "A theory of eddy-current flaw detection," Ph.D. dissertation, Univ. Michigan, Ann Arbor, MI, 1964.
- [28] R. O. Duda, P. E. Hart, and D. G. Stark, *Pattern Classification*. Hoboken, NJ: Wiley, 2001.
- [29] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1986.
- [30] J. N. Hwang, S. R. Lay, and A. Lippman, "Nonparametric multivariate density estimation: A comparative study," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2795–2810, Oct. 1994.
- [31] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. Hoboken, NJ: Wiley, 1992.
- [32] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.
- [33] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-Verlag, 1999.
- [34] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
- [35] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I*. Hoboken, NJ: Wiley, 1968.
- [36] M. Tanner, *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions*. New York: Springer-Verlag, 1993.
- [37] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*, 3rd ed. Reading, MA: Addison-Wesley, 2002.
- [38] V. Vapnik, *Statistical Learning Theory*. New York: Springer-Verlag, 1998.
- [39] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discov.*, vol. 2, no. 2, pp. 955–974, Jun. 1998.
- [40] J. Weston and C. Watkins, "Support vector machines for multiclass pattern recognition," in *Proc. 7th Eur. Symp. Artif. Neural Netw.*, 1999.
- [41] J. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin DAGs for multiclass classification," *Adv. Neural Inf. Process. Syst.*, vol. 12, pp. 547–553, 2000.
- [42] I. J. Won, D. A. Keiswetter, D. R. Hanson, E. Novikova, and T. M. Hail, "A monostatic broadband electromagnetic induction sensor," *J. Environ. Eng. Geophys.*, vol. 2, no. 1, pp. 53–64, Mar. 1997.
- [43] T. J. Ulrych, M. D. Sacchi, and M. Graul, "Signal and noise separation: Art and science," *Geophysics*, vol. 64, no. 5, pp. 1648–1656, Sep./Oct. 1999.
- [44] P. Gao and L. Collins, "Classification of landmine-like metal targets using wideband electromagnetic induction," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1352–1361, May 2000.



Alireza Aliamiri (S'06) received the B.Sc. and M.Sc. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from Northeastern University, Boston, MA, in 2007.

From 2004 to 2007, he was a Research Assistant at the Center for Subsurface Sensing and Imaging Systems, Northeastern University. He is currently a Senior Engineer with Airvana, Inc., Chelmsford, MA, working on developing mobile broadband network infrastructures. His research interests include statistical signal processing, pattern recognition, communication, and information theory.

Jack Stalnaker received the B.S. degree in geology and anthropology from the College of Charleston, Charleston, SC, in 1998, and the Ph.D. degree in geophysics from Texas A&M University, College Station, in 2004.

He was a Postdoctoral Research Associate with the Center for Subsurface Sensing and Imaging Systems, Northeastern University, Boston, MA, from 2004 to 2007. Since 2007, he has been with the Concepts and Analysis group, Textron Systems Corporation, Wilmington, MA. His research interests are the detection and discrimination of unexploded ordnance, 3-D numerical modeling of electromagnetic induction, and the application of near-surface geophysical techniques to environmental, engineering, and anthropological (i.e., archeological) problems.



Eric L. Miller (S'90–M'95–SM'03) received the S.B., S.M., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1990, 1992, and 1994, respectively.

He is currently a Professor in the Department of Electrical and Computer Engineering, Tufts University, Medford, MA. His research interests include physics-based tomographic image formation and object characterization, inverse problems in general and inverse scattering, particularly regularization, statistical signal and imaging processing, and computational physical modeling. This work has been carried out in the context of applications, including medical imaging, nondestructive evaluation, environmental monitoring and remediation, landmine and unexploded ordnance remediation, and automatic target detection and classification.

Dr. Miller is a member of Tau Beta Pi, Phi Beta Kappa, and Eta Kappa Nu. He received the CAREER Award from the National Science Foundation in 1996 and the Outstanding Research Award from the College of Engineering at Northeastern University in 2002. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and was in the same position at the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1998 to 2002. He is the Cogeneral Chair of the 2008 IEEE International Geoscience and Remote Sensing Symposium to be held in Boston, MA.