Classification of Unexploded Ordnance Using Incomplete Multisensor Multiresolution Data

David Williams, *Member, IEEE*, Chunping Wang, Xuejun Liao, *Senior Member, IEEE*, and Lawrence Carin, *Fellow, IEEE*

***Abstract*—We address the problem of unexploded ordnance (UXO) detection in which data to be classified are available from multiple sensor modalities and multiple resolutions. Specifically, features are extracted from measured magnetometer and electromagnetic induction data; multiple-resolution data are manifested when the sensors are separated from the buried targets of interest by different distances (e.g., different sensor-platform heights). The proposed classification algorithm explicitly emphasizes features extracted from fine-resolution imagery over those extracted from less reliable coarse-resolution data. When fine-resolution features are unavailable (due to undeployed sensors), the algorithm analytically integrates out the missing features via an estimated conditional density function, which is conditioned on the observed features (from deployed sensors). This density function exploits the statistical relationships that exist among features at different resolutions, as well as those among features from different sensors (in the multisensor case). Experimental classification results are shown for real UXO data, on which the proposed algorithm consistently achieves better classification performance than common alternative approaches.**

***Index Terms*—Classification, incomplete data, missing data, multiresolution, multisensor, unexploded ordnance (UXO).**

I. Introduction

THE PROBLEM of unexploded ordnance (UXO) classification continues to receive significant attention in the scientific community [1]–[5]. The objective of a UXO detection (or classification) task is to distinguish buried UXO targets from non-UXO targets (i.e., clutter). To this end, sensors are used to measure data (e.g., magnetic fields) over a 2-D grid. These (raw data) sensor measurements can therefore be considered to be in the form of an image. The ultimate classification task is then performed using features extracted from this imagery. This paper addresses the problem of multisensor UXO detection when magnetometer and electromagnetic induction (EMI) sensors are employed. In particular, we address the realistic case in which imagery from each of these sensor modalities may be available from multiple resolutions. Throughout this paper, the term "resolution" refers to the amount of spatial detail attainable in an image, with this quantity being inversely proportional to the distance between the sensor and the targets (or ground).

Manuscript received October 28, 2006; revised March 20, 2007. This work was supported by the Strategic Environmental Research and Development Program.

The authors are with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA (e-mail: dpw@ee.duke.edu; cw36@ee.duke.edu; xjliao@ee.duke.edu; lcarin@ee.duke.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.2007.896558

Magnetometer and EMI constitute the principal sensors used in UXO detection and classification [6]–[8]. It is a time-consuming task to deploy these sensors over the large domains that must be interrogated for possible buried UXO. There has therefore been interest in deploying these sensors on helicopters, thereby accelerating the data-collection process. Although helicopter-borne sensors afford increased collection speed, they also incur the deleterious effect of a loss of signal strength. In particular, for a distance r between the sensor and target, magnetometers and EMI sensors measure field strengths proportional to $1/r^3$ and $1/r^6$, respectively. Therefore, the increased sensor height required by the helicopter manifests a significant loss of signal strength, which undermines the ability to detect small or deeply buried UXO. In practice, therefore, one may be interested in deploying a helicopter-borne sensor over as large a region as possible, with ground-based sensors applied only locally, over a coarse set of lines running through the site under test.

This paper addresses the problem of classification for data sets in which the features of different data points are extracted from sensor imagery at different resolutions. Additionally, this paper considers the more general case in which multiple sensor modalities—each of which may operate at multiple resolutions—are employed. In the multisensor scenario, incomplete data are manifested when some data points are interrogated by only a subset of the available sensors. Incomplete data also exist in the single-sensor case when not all data points have features extracted from imagery at all resolution levels. Although the classification algorithm introduced in this paper is applicable for data sets that fit the general multiresolution framework, we focus specifically on the problem of UXO detection. In summary, the novel problem we address in this paper is of multisensor multiresolution incomplete-data UXO classification.

It is important to emphasize that this paper addresses a multiresolution classification problem that has not been examined previously (see [9] for a thorough review of multiresolution work). In the most previous "multiresolution" image classification work [10], [11], the original imagery actually exists at only a single resolution; the term "multiresolution" refers simply to a wavelet or other multiresolution decomposition [12] of the original single-resolution imagery. In contrast, this paper utilizes multiple raw images, each at a unique resolution. The ultimate classification objective also distinguishes this paper from other multiresolution image classification work. Most multiresolution classification work strives for pixel-level classification via image segmentation [13]–[15]. In contrast, in this paper, a given image belongs to a single class (UXO or non-UXO).

Several approaches can be employed to handle the missing-data problem in which some data points are characterized by features extracted from only a subset of the possible sensors and/or resolution levels. One approach would build a separate classifier for each type of data. Assuming the set of possible data is relatively small, this approach would be reasonable. The major drawback with this method, however, is that the dependences between the different types of data are not exploited. In addition to ignoring the correlations between sensors, the severe fragmentation of the data set—based on the combinations of which sensors and resolutions are observed—may leave insufficient data to train each classifier.

A different method would concatenate the features from the various resolutions; incomplete data arising from missing sensors and/or resolutions would be handled in some way such as by imputation [16]. However, such an approach would treat features obtained from images at different resolutions equally. Intuitively, one should favor using features extracted from high-resolution imagery.

The algorithm proposed in this paper extends the work in [17]—in which missing data are analytically integrated out—to the case of multiresolution imagery. The algorithm, which does not suffer from any of the drawbacks that plague the aforementioned methods, requires only a single classifier, regardless of the number of sensors or the number of resolutions involved in the problem. Moreover, all data are utilized, so that correlations among sensors as well as among features at different resolutions are exploited. Additionally, features extracted from different resolutions are not treated equally; rather, fine-resolution features are given more importance. Furthermore, the missing data that exist are handled in a principled manner, avoiding explicit imputation. Specifically, the missing data are integrated out via the use of an estimated conditional density function that relates the dependences of features both of a single given sensor at different resolutions as well as of features from different sensors.

The remainder of this paper is organized as follows. Section II explains notation necessary for the proposed classification algorithm introduced in Section III. Section IV describes the UXO model inversion (and feature extraction) processes. Experimental classification results are shown in Section V. Section VI consists of a discussion, which is followed in Section VII by concluding comments and directions for future work.

II. Notation

Consider the case in which a sensor generates raw data in the form of an image, from which features are extracted subsequently. Assume we possess S such sensors, the sth of which can operate at $R_s + 1$ resolutions; the resolution is a function of the distance between the sensor and the ground under which the targets are buried. Each of the S sensors may or may not be of the same modality, and the possible resolutions of each sensor are in general unique. We define Δ_r^s to be the rth sensor-target separation distance (hereafter, simply "separation distance") of the sth sensor, for $s = 1, 2, \ldots, S$, and $r = 0, 1, 2, \ldots, R_s$. Let Δ_0^s denote the smallest separation distance of the sth sensor. The resolution of an image, which is inversely proportional to the separation distance, is written $\mathcal{R}(\cdot)$. The image that results from operating a sensor at its smallest separation distance is referred to as a fine-resolution image. Sensors operating at larger separation distances generate coarse-resolution imagery.

Assume that for a given sensor, the type of features extracted from the raw-image data are fixed, regardless of the separation distance of the sensor that generated the data. That is, for a given sensor, the specific features extracted will be identical for all separation distances, but the feature values will in general be unique for each separation distance.

Let $\boldsymbol{x}_i^{(s)} \in \mathbb{R}^{F_s}$ be the F_s features of the sth sensor for the ith item (i.e., object, which may be UXO or non-UXO), which is extracted from the data corresponding to the highest resolution image of the sth sensor. For all larger separation distances, let $\boldsymbol{z}_i^{(s,r)} \in \mathbb{R}^{F_s}$ be the F_s features of the sth sensor for the ith item, which is extracted from the image obtained with the sth sensor operating at the rth separation distance. Define $\boldsymbol{x}_i = [\boldsymbol{x}_i^{(1)}, \boldsymbol{x}_i^{(2)}, \ldots, \boldsymbol{x}_i^{(S)}]$ to be the concatenated feature vectors extracted from imagery at each sensor's respective smallest separation distance. Similarly, define $\boldsymbol{z}_i = [\boldsymbol{z}_i^{(1)}, \boldsymbol{z}_i^{(2)}, \ldots, \boldsymbol{z}_i^{(S)}]$ to be the concatenated feature vectors extracted from each sensor's coarse-resolution imagery, where $\boldsymbol{z}_i^{(s)} = [\boldsymbol{z}_i^{(s,1)}, \boldsymbol{z}_i^{(s,2)}, \ldots, \boldsymbol{z}_i^{(s,R_s)}]$. Hereafter, we shall refer to $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ as primary and auxiliary features (or data), respectively.

The data can alternatively be partitioned in terms of its observed and missing components. Let o_i^x be the set of sensors for which the ith item's primary features are observed. Let m_i^x be the (complementary) set of sensors for which the primary features are missing for the ith item. Similarly, let o_i^z be the set of sensor and separation-distance pairs for which the auxiliary features for the ith item are observed. Let m_i^z be the (complementary) set of sensor and separation-distance pairs for which the auxiliary features for the ith item are missing. To simplify notation, we shall suppress the superscripts when doing so will not cause confusion (e.g., $\boldsymbol{x}_i^{o_i^x}$, the primary features (from all sensors) that are observed for the ith item, will be written as $\boldsymbol{x}_i^{o_i}$). The primary and auxiliary data of the ith item can thus be written as $\boldsymbol{x}_i = [\boldsymbol{x}_i^{o_i}; \boldsymbol{x}_i^{m_i}]$ and $\boldsymbol{z}_i = [\boldsymbol{z}_i^{o_i}; \boldsymbol{z}_i^{m_i}]$, respectively.

Data for a given item are deemed to be complete if we possess all primary features, for all sensors, for that data point (i.e., $m_i^x = \emptyset$). A data point is otherwise deemed incomplete. It should be noted that there exist two different types of incomplete data. First, data would be incomplete if some subset of sensors were never deployed (at any separation distance) for the corresponding item (UXO or non-UXO). Data could also be incomplete even when all sensors were deployed for the item; specifically, the data would still be considered incomplete in this case if the data had not been interrogated at the smallest separation distance of every sensor.

III. Classification With Incomplete Data

Assume we have a set of labeled (incomplete) data

$$\mathcal{D}_{\mathrm{L}} = \{\boldsymbol{x}_i, \boldsymbol{z}_i, y_i, \epsilon_i, o_i^x, o_i^z, m_i^x, m_i^z\}_{i=1}^{N_L} \tag{1}$$

where $y_i \in \{-1, 1\}$ is the label (indicating non-UXO or UXO, respectively) of the ith item, and $\epsilon_i \in [0, 0.5)$ is the corresponding labeling error rate. The labeling error rate is simply the probability that a true label was flipped (corrupted) to the wrong label (e.g., $\{y_i^{\text{true}} = 1\} \rightarrow \{y_i = -1\}$). Such imperfect labels can be manifested when a human analyst performs the labeling without excavating the buried object.

Let $\boldsymbol{w}_s = [w_s^{(1)}, w_s^{(2)}, \ldots, w_s^{(F_s)}]$ represent the F_s weights of a classifier on the primary features of the sth sensor. Let $\boldsymbol{w} = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_S]$ be the classifier weights on the primary features of each sensor (i.e., $\boldsymbol{x}_i$). Note that the number of features from each sensor need not be identical. It must be emphasized that the weights—and, hence, the resulting classifier—are on the features extracted from only the fine-resolution imagery. We emphasize this caveat by using different notation for primary features extracted from the fine-resolution imagery ($\boldsymbol{x}_i$) and auxiliary features extracted from coarse-resolution imagery ($\boldsymbol{z}_i$).

In logistic regression (with a hyperplane classifier), the probability of label y_i given feature vector $\boldsymbol{x}_i$ is $p(y_i|\boldsymbol{x}_i, \boldsymbol{w}) = \sigma(y_i \boldsymbol{w}^T \boldsymbol{x}_i)$, where $\sigma(\eta) = (1 + \exp(\eta))^{-1}$ is the sigmoid link function, and $\boldsymbol{w}$ constitutes a classifier. Accounting for imperfections in the labeling process arising from a known labeling error rate ϵ_i, the probability of label y_i given $\boldsymbol{x}_i$ and ϵ_i is [18]

$$p(y_i|\boldsymbol{x}_i, \epsilon_i, \boldsymbol{w}) = \epsilon_i + (1 - 2\epsilon_i)\sigma(y_i \boldsymbol{w}^T \boldsymbol{x}_i). \tag{2}$$

Note that the standard case of perfect labels is recovered when $\epsilon_i = 0$. If all features are extracted from complete data, the weights of the classifier can be learned easily by maximizing the likelihood of the data. Here, we consider the case in which the data are in general incomplete in the sense described previously.

Recall that the classifier is to be designed for only the primary data—the features extracted from the finest resolution imagery. We first partition $\boldsymbol{x}_i$ into its observed and missing parts $\boldsymbol{x}_i = [\boldsymbol{x}_i^{o_i}; \boldsymbol{x}_i^{m_i}]$ and then apply the same partition to $\boldsymbol{w}$ to obtain $\boldsymbol{w} = [\boldsymbol{w}_{o_i}; \boldsymbol{w}_{m_i}]$. With $\eta_i = \boldsymbol{w}_{m_i}^T \boldsymbol{x}_i^{m_i}$, (2) can be written as

$$p(y_i|\boldsymbol{x}_i^{o_i}, \epsilon_i, \boldsymbol{w}) = \epsilon_i + (1 - 2\epsilon_i)\sigma\left(y_i(\boldsymbol{w}_{o_i}^T \boldsymbol{x}_i^{o_i} + \eta_i)\right). \tag{3}$$

If the missing data $\boldsymbol{x}_i^{m_i}$ is integrated out, the needed probability of y_i given all observed features can be written as

$$\begin{aligned} &p\left(y_i|\boldsymbol{x}_i^{o_i}, \boldsymbol{z}_i^{o_i}, \epsilon_i, \boldsymbol{w}\right) \\ &= \int p\left(y_i|\boldsymbol{x}_i^{o_i}, \epsilon_i, \boldsymbol{w}\right) p\left(\boldsymbol{x}_i^{m_i}|\boldsymbol{x}_i^{o_i}, \boldsymbol{z}_i^{o_i}\right) d\boldsymbol{x}_i^{m_i} \qquad (4) \\ &= \epsilon_i + (1-2\epsilon_i)\int \sigma\left(y_i\left(\boldsymbol{w}_{o_i}^T \boldsymbol{x}_i^{o_i} + \eta_i\right)\right) p\left(\eta_i|\boldsymbol{x}_i^{o_i}, \boldsymbol{z}_i^{o_i}\right) d\eta_i. \qquad (5) \end{aligned}$$

Although the classifier uses only the primary data, the auxiliary data are exploited when primary data are missing, via the conditional density function $p(\boldsymbol{x}_i^{m_i}|\boldsymbol{x}_i^{o_i}, \boldsymbol{z}_i^{o_i})$. That is, when primary data are available, they are utilized; when primary data are unavailable, the auxiliary data become relevant and are exploited.

The integration in (5) can be performed analytically by making two mild assumptions. First, we assume that $p(\boldsymbol{x}_i, \boldsymbol{z}_i)$ is a Gaussian mixture model (GMM), which can accurately model many reasonably well-behaved distributions. This density function describes the relationships among the same features obtained from different resolutions of a given sensor; it also describes the relationships among features from different sensor modalities. It then follows that

$$p(\boldsymbol{x}_i, \boldsymbol{z}_i) = p\left(\boldsymbol{x}_i^{m_i}, \boldsymbol{x}_i^{o_i}, \boldsymbol{z}_i^{o_i}\right) p\left(\boldsymbol{z}_i^{m_i}|\boldsymbol{x}_i^{m_i}, \boldsymbol{x}_i^{o_i}, \boldsymbol{z}_i^{o_i}\right) \tag{6}$$

where $p(\boldsymbol{x}_i^{m_i}, \boldsymbol{x}_i^{o_i}, \boldsymbol{z}_i^{o_i})$ is also necessarily a GMM. Introducing the notation $\chi_i^{o_i} = [\boldsymbol{x}_i^{o_i}, \boldsymbol{z}_i^{o_i}]$, this ($K$-component) GMM is

$$p\left(\boldsymbol{x}_i^{m_i}, \chi_i^{o_i}\right) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x}_i^{m_i} \\ \chi_i^{o_i} \end{bmatrix}; \begin{bmatrix} \mu_k^{m_i} \\ \mu_k^{o_i} \end{bmatrix}, \begin{bmatrix} \Sigma_k^{m_i m_i} & \Sigma_k^{m_i o_i} \\ \Sigma_k^{o_i m_i} & \Sigma_k^{o_i o_i} \end{bmatrix}\right) \tag{7}$$

where π_k is the nonnegative mixing proportion that sum to unity. Moreover, $p(\boldsymbol{x}_i^{m_i}|\chi_i^{o_i})$ is a GMM as well. Because of the linear relation $\eta_i = \boldsymbol{w}_{m_i}^T \boldsymbol{x}_i^{m_i}$, $p(\eta_i|\chi_i^{o_i})$ is also a GMM

$$p\left(\eta_i|\chi_i^{o_i}\right) = \sum_{k=1}^{K} \delta_k^i \mathcal{G}\left(\frac{\eta_i - \boldsymbol{w}_{m_i}^T \xi_k^{m_i}}{\sqrt{\boldsymbol{w}_{m_i}^T \Omega_k^{m_i} \boldsymbol{w}_{m_i}}}\right) \tag{8}$$

with the parameters

$$\delta_k^i = \frac{\pi_k \mathcal{N}\left(\chi_i^{o_i}; \mu_k^{o_i}, \Sigma_k^{o_i o_i}\right)}{\sum_{\ell=1}^{K} \pi_\ell \mathcal{N}\left(\chi_i^{o_i}; \mu_\ell^{o_i}, \Sigma_\ell^{o_i o_i}\right)} \tag{9}$$

$$\xi_k^{m_i} = \mu_k^{m_i} + \Sigma_k^{m_i o_i}\left(\Sigma_k^{o_i o_i}\right)^{-1}\left(\chi_i^{o_i} - \mu_k^{o_i}\right) \tag{10}$$

$$\Omega_k^{m_i} = \Sigma_k^{m_i m_i} - \Sigma_k^{m_i o_i}\left(\Sigma_k^{o_i o_i}\right)^{-1} \Sigma_k^{o_i m_i} \tag{11}$$

where $\mathcal{G}(\eta_i) = (2\pi)^{-1/2} \exp\{-\eta_i^2/2\}$ is the standard univariate Gaussian density function with zero mean and unit variance. The requisite GMM density function estimation can be accurately performed using all available data, via the variational Bayesian expectation-maximization algorithm presented in [17].

The second (very accurate) assumption is that the sigmoid function can be approximated as a probit function (i.e., a Gaussian cumulative distribution function)

$$\sigma(\alpha) \approx \int_{-\infty}^{\alpha} \mathcal{G}\left(\frac{u}{\beta}\right) du \tag{12}$$

where $\beta = (\pi/\sqrt{3})$. The accuracy of this approximation is shown in Fig. 1.

Mirroring the derivation in [17], it can then be shown that the integral in (5) can be computed analytically. The result of this integration is that the probability of y_i given only the observed portions of $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ can be expressed as a mixture of sigmoids:

$$\begin{aligned} &p\left(y_i|\boldsymbol{x}_i^{o_i}, \boldsymbol{z}_i^{o_i}, \epsilon_i, \boldsymbol{w}\right) \\ &\approx \epsilon_i + (1 - 2\epsilon_i) \sum_{k=1}^{K} \delta_k^i \sigma\left(\frac{y_i \beta(\boldsymbol{w}_{m_i}^T \xi_k^{m_i} + \boldsymbol{w}_{o_i}^T \boldsymbol{x}_i^{o_i})}{\sqrt{\boldsymbol{w}_{m_i}^T \Omega_k^{m_i} \boldsymbol{w}_{m_i} + \beta^2}}\right). \qquad (13) \end{aligned}$$

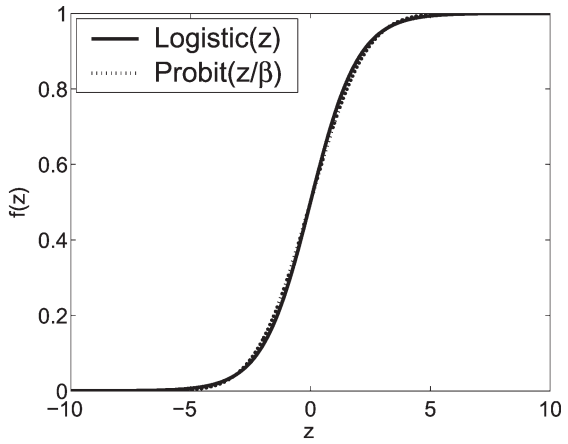

Fig. 1. Illustration of the accuracy of the approximation made between the logistic function and the (scaled) probit function.

The log-likelihood function of the incomplete data in (1) is then

$$\ell(\boldsymbol{w}) = \log p\Big(\{y_i\}_{i=1}^{N_L} | \{\boldsymbol{x}_i^{o_i}\}_{i=1}^{N_L}, \{\boldsymbol{z}_i^{o_i}\}_{i=1}^{N_L}, \{\epsilon_i\}_{i=1}^{N_L}, \boldsymbol{w}\Big) \tag{14}$$

$$\approx \sum_{i=1}^{N_L} \log \left[\epsilon_i + (1-2\epsilon_i) \times \sum_{k=1}^{K} \delta_k^i \sigma \left(\frac{y_i \beta \left(\boldsymbol{w}_{m_i}^T \xi_k^{m_i} + \boldsymbol{w}_{o_i}^T \boldsymbol{x}_i^{o_i} \right)}{\sqrt{\boldsymbol{w}_{m_i}^T \Omega_k^{m_i} \boldsymbol{w}_{m_i} + \beta^2}} \right) \right]. \tag{15}$$

The objective function (15) to be maximized is no longer concave for two reasons. First, the concavity is destroyed by the imperfect labels resulting from ϵ_i. Even in the case of perfect labels, (15) is not concave because of the particular form of the argument of the sigmoid function, arising from the incomplete data. Since (15) is not concave, an intelligent initialization of $\boldsymbol{w}$ is valuable for avoiding local maxima. We therefore initialize $\boldsymbol{w}$ as follows. We "complete" the data set by replacing the missing features $\boldsymbol{x}_i^{m_i}$ with the conditional mean $\mathbb{E}[\boldsymbol{x}_i^{m_i} | \boldsymbol{x}_i^{o_i}] = \sum_{k=1}^{K} \delta_k^i \xi_k^{m_i}$, where δ_k^i and $\xi_k^{m_i}$ are defined in (9) and (10), respectively. For the initialization, we also treat all labels as perfect, artificially setting all $\epsilon_i = 0$. This "completed" and "perfectly" labeled data set is submitted to the standard logistic regression [19] to obtain $\boldsymbol{w}_0$, which is then used as the initialization of $\boldsymbol{w}$ in maximizing (15) by a modified form of gradient ascent (additional details are shown in the Appendix). Empirical evidence [20] suggests that this initialization successfully avoids most local maxima.

Thus, the maximum-likelihood logistic regression classifier $\boldsymbol{w}$ can then be obtained, in spite of the missing data (and imperfect labels). Thereafter, the class predictions of an unlabeled testing data point with incomplete (missing) features is computed trivially using (13) (with $\epsilon_i = 0$ since no actual labeling has transpired).

IV. Multisensor Multiresolution UXO Data

The proposed classification algorithm is designed for data sets consisting of imagery that exists at multiple resolutions for multiple sensors, which is a realistic scenario for UXO detection tasks. Here, we consider the case in which we possess two modalities, which are the magnetometer and EMI. Moreover, it is assumed that each sensor can operate at two different image resolution levels. The image resolution is a function of the distance between the sensor and the (buried) target. Therefore, these unique image resolutions are manifested by deploying a given sensor on different platforms (at different heights).

Coarse-resolution imagery is generated by a given sensor when a relatively large distance separates the sensor from the targets; here, this situation corresponds to the case in which the sensor collects data while located on a low-altitude airborne platform (e.g., a helicopter) that flies above the area of interest. In contrast, fine-resolution imagery is generated by a given sensor when a smaller distance separates the sensor from the targets; here, this situation corresponds to the case in which the sensor collects data while located on a ground-based platform. Because the magnetometer and EMI sensor may not be located on the same platform, each sensor may interrogate unique areas of land that overlap only partially. As a result, some targets may be characterized by imagery from only one of the sensors, which is a case of incomplete data.

A. Feature Extraction Models

The magnetometer and EMI sensor data used in this study are magnetic field measurements as a function of spatial position. The data from each sensor can therefore be considered to be in the format of an image. Features are extracted from this imagery and then subsequently used in the classification stage discussed in Section III. The features we use here are the parameters of UXO models developed in [8] that are fit via a model inversion process. Specifically, the measured (image) data are the input to the inversion, and the model parameters (features) are the output of the inversion. We subsequently employ these fitted model parameters as the features of the classification stage. As a result, regardless of the resolution of the image, the same features (parameters) are extracted. Although the features are identical, the actual values of these features extracted from images at different resolutions (for any given data point) will be unique.

1) Magnetometer Model: Ferrous objects cause changes in the observed background magnetic field of the Earth; magnetometers sense these changes. It has been shown that the spatially dependent magnetometer signal is well-modeled by a simple magnetic dipole [21]. The success of this magnetic-dipole model for sensing buried UXO [6], [7], [22], [23] motivates us to employ the model for the measured spatially dependent magnetometer data here.

In the $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ coordinate system of the sensor, let the z-direction be normal to the air–soil interface. Let the position vectors of the sensor (i.e., observation point) and target-dipole be $\boldsymbol{r}_s = [x_s \ y_s \ z_s]$ and $\boldsymbol{r}_t = [x_t \ y_t \ z_t]$, respectively. Define $\boldsymbol{r}_{ts} = \boldsymbol{r}_s - \boldsymbol{r}_t$ to be the vector directed from the dipole to the sensor, with $\boldsymbol{r} = (\boldsymbol{r}_{ts}/|\boldsymbol{r}_{ts}|)$ of the corresponding unit vector. It should also be noted that the orientation of the magnetic-dipole—completely summarized by the angles θ and ϕ—is different from the direction of the ordnance itself.

When the sensor is sufficiently distant from the buried target relative to the target dimensions, the (vector) magnetic field may be represented approximately as [8], [24]

$$\mathbf{H} = \frac{1}{2\pi}\frac{\boldsymbol{m}\cdot\boldsymbol{r}}{|\boldsymbol{r}_{ts}|^3} \tag{16}$$

where $\boldsymbol{m}$ is the magnetic-dipole moment. The magnetometer employed to collect the data used in this paper measures the z-component of the magnetic field as a function of position on the surface. This measurement is subsequently fit to the model in (16) via a simple gradient search. Specifically, the parameters that the model inversion fits are the target position (x, y, and depth z), the magnetic-dipole strength, ($m = |\boldsymbol{m}|$), and the magnetic-dipole orientation (θ and ϕ). We retain the last four parameters (z, m, θ, and ϕ) of the model as features for the classification stage.

2) EMI Model: A model for the EMI response of targets that generalizes the magnetometer model via a frequency-dependent magnetic dipole has been developed in [8]. Specifically, the magnetic-dipole moment $\boldsymbol{m}$ of a target is represented as

$$\boldsymbol{m} = \mathbf{M}\mathbf{H}^{\text{inc}} \tag{17}$$

where $\mathbf{H}^{\text{inc}}$ denotes the incident (excitation) magnetic field, and $\mathbf{M}$ is the magnetization tensor that relates the magnetic field to the magnetic-dipole moment. For a UXO assumed to be rotationally symmetric with the axis of rotation along the z direction, the (frequency-dependent) magnetization tensor can be expressed as a diagonal matrix [22]

$$\mathbf{M} = \text{diag}\left[m_{p0} + \sum_i \frac{\omega m_{pi}}{\omega - j\omega_{pi}}, m_{p0} + \sum_i \frac{\omega m_{pi}}{\omega - j\omega_{pi}}, m_{z0} + \sum_k \frac{\omega m_{zk}}{\omega - j\omega_{zk}}\right]. \tag{18}$$

The terms m_{z0} and m_{p0} correspond to the zero-frequency magnetic-dipole moments of the target, which are directed perpendicular to and along the target's axis of rotation, respectively. The terms m_{zk} and m_{pi} in (18) account for the frequency-dependent character of the response, while ω_{zk} and ω_{pi} correspond to EMI resonant frequencies. Because higher order dipole moments in the summations in (18) typically lack significant strength [25], here, we use only the first term in each summation, which is representative of the principal dipole mode along each of the principal axes.

If it is assumed that the EMI source responsible for the excitation magnetic field $\mathbf{H}^{\text{inc}}$ can be represented—as seen from the target—as a magnetic dipole with moment $\boldsymbol{m}_s$, then [8]

$$\mathbf{H}^{\text{inc}} = \boldsymbol{r}\frac{1}{2\pi}\frac{\boldsymbol{m}_s\cdot\boldsymbol{r}}{|\boldsymbol{r}_{st}|^3} \tag{19}$$

where $\boldsymbol{r}_{st}$ is the vector directed from the source to the target center, with ($\boldsymbol{r} = \boldsymbol{r}_{st}/|\boldsymbol{r}_{st}|$) the corresponding unit vector. Assuming sufficient proximity of the sensor's source and receiver coils, the total (frequency-dependent) magnetic field observed at the sensor will be [8]

$$\mathbf{H}^{\text{rec}} \propto \frac{\boldsymbol{r}}{|\boldsymbol{r}_{st}|^6}\boldsymbol{r}^T\mathbf{U}^T\mathbf{M}\mathbf{U}\boldsymbol{r} \tag{20}$$

where the proportionality constant depends on the strength of the dipole source $\boldsymbol{m}_s$ and the characteristics of the receiver.

The 3×3 unitary rotation matrix $\mathbf{U}$ rotates the fields from the coordinate system of the sensor to the coordinate system of the target, and $\mathbf{U}^T$ transforms the dipole fields of the target (in the $\mathbf{M}$ coordinate system) back to the coordinate system of the sensor. Explicitly, the target orientation, in terms of the angles of the target θ and ϕ with respect to the sensor coordinate system, is accounted for by

$$\mathbf{U} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{21}$$

As with the magnetometer, the EMI sensor employed in this paper measures the z-component of the magnetic field as a function of position on the surface. This measurement is subsequently fit to the model in (20) via a form of the Levenberg–Marquardt method [26]. Specifically, the parameters that the model inversion fits are the target position (x, y, and depth z), the target orientation (θ and ϕ), the magnetic dipole strengths (m_{z0}, m_{p0}, m_{zk}, and m_{pi}), and the EMI resonant frequencies (ω_{zk} and ω_{pi}). We retain five parameters (z, θ, ϕ, m_{z0}, and m_{p0}) of the model as features for the classification stage.

In fitting the more sophisticated EMI model, parameters from the magnetometer inversion are used to constrain the search of some of the EMI model parameters. Specifically, the depth (z) and cross-sectional position (x and y) of the target found by the magnetometer inversion are used to initialize the target location in the EMI inversion. This initialization helps avoid some local maxima in the inversion process. To overcome other local maxima, several (model-fitting) solutions are obtained, with each solution resulting from randomly initializing the remaining parameters of the model. The final parameters of the model are taken to be those of the solution that minimizes the mean-square error between the measured and model-fit data.

B. Simulation of Multiresolution Imagery

We possess measured magnetometer and EMI (image) data measured by ground-based sensors; we simulate multiresolution imagery from the available single-resolution imagery in the following manner. The image-simulation process for the two sensors is identical, so here, we explain the process in terms of the magnetometer sensor. During the explanation, we shall refer to Fig. 2, which illustrates the various stages of the process for one example target.

We begin with ground-based magnetometer data measurements for a target [Fig. 2(a)]. The magnetometer model inversion explained in Section IV is performed, which provides model parameters. These model parameters are then assumed to

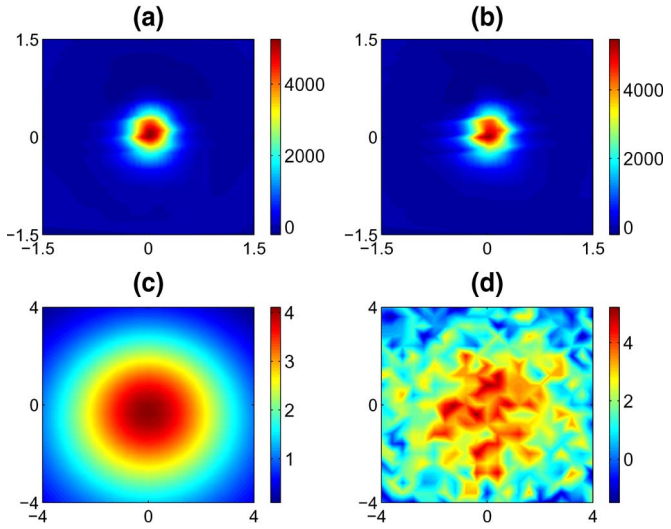

Fig. 2. Process of simulating coarse-resolution (helicopter-based) magnetometer data. (a) Measured fine-resolution (ground-based) magnetometer data. (b) Synthesized fine-resolution (ground-based) magnetometer data using the model parameters obtained from the model inversion with the data in (a). (c) Synthesized noise-free coarse-resolution (helicopter-based) magnetometer data using the model parameters obtained from the model inversion and the data in (a). (d) Same as (c) except with (sensor) noise added.

be the true model parameters of the target. Using these model parameters, ground-based data can be synthesized [Fig. 2(b)]. If the model fitting was successful, the measured and synthesized data should be nearly identical. Using these same model parameters, one can instead synthesize coarse-resolution (helicopter-based) data [Fig. 2(c)] by increasing the value representing the distance between the sensor and the ground. The distance from the ground-based sensors to the ground surface is 0.3 m, while the distance from the helicopter-based sensors to the ground surface is assumed to be 5.0 m.

As stated until now, this synthesis procedure would be unrealistic because the sensor noise of the helicopter-based sensor should be higher than that of the ground-based sensor. To reflect this fact, white Gaussian noise [$\mathcal{N}(0, \sigma^2)$, where here, $\sigma = 1$]—representing sensor noise—is added to this synthesized data to produce noisy data [Fig. 2(d)]. This noisy helicopter-based sensor data are then taken to be the "raw" coarse-resolution sensor measurements, which are analogous to the raw ground-based sensor data from which the original model parameters were obtained. These noisy data are subsequently used to obtain coarse-resolution features via the magnetometer model inversion. It is important to reiterate that the particular features (but not the feature values) extracted from any image from a given sensor will be identical, regardless of the image's resolution.

It should be noted that the amplitude of the response in Fig. 2(a) and (b) is much larger than that in Fig. 2(c) and (d) because the response is proportional to $1/r_{ts}^3$, where r_{ts} is the distance between the target and the sensor [see (16)]. Also, note that the (physical) area shown in Fig. 2(a) and (b) is 3 m $\times$ 3 m, while the area in Fig. 2(c) and (d) is 8 m $\times$ 8 m; this larger area must be considered in order to ensure that the full response is captured, for the response expands spatially as the sensor-target distance increases.

V. Experimental Results

To evaluate the proposed incomplete-data classification algorithm, we applied it to a UXO data set consisting of 166 items, 41 of which are UXO. This data set was collected by the multisensor towed array detection system [27]. This system is composed of arrays of full-field cesium vapor magnetometers and time-domain electromagnetic pulsed induction sensors. The magnetometers were Geometrics Model 822ROV, while the EMI sensors were highly modified Geonics EM-61 sensors. The data were collected at a bombing target on the Badlands Bombing Range, Ogala Sioux Reservation, Pine Ridge, South Dakota. The UXO items present at the site included M 38 (100 lb) sand-filled practice bombs, M 57 (250 lb) practice bombs, 2.25- and 2.75-in rocket bodies and rocket warheads, and ordnance scrap (such as tail fins and casing parts).

For every item, we possess both (measured) fine-resolution and (synthesized) coarse-resolution features from each of the two sensors (a magnetometer and an EMI sensor). As mentioned earlier, four magnetometer features are used to characterize the magnetometer data at each resolution level, while five EMI features are used to characterize the EMI data at each resolution level.

In all experiments, it is assumed that coarse-resolution (helicopter-based sensor) data are available from both sensors for all data points. This choice is motivated by the fact that it would be relatively quick, easy, and inexpensive to acquire such data vis-á-vis ground-based sensor data. In contrast, it is assumed that fine-resolution sensor data will be missing for some data points, with these amounts made specific later. It should be noted, however, that the proposed algorithm can function successfully even when data points are completely missing data from a given sensor (i.e., at all resolutions).

Because data are available from two different sensors, many different combinations of missing data are possible. To conduct an extensive investigation of the proposed algorithm, 36 different combinations of missing data are considered (explained in more detail below). The binary Cartesian product of a set S is the set of ordered pairs

$$S \times S = \{(\alpha, \beta) |\ \alpha \in S \text{ and } \beta \in S\}. \tag{22}$$

In (22), let α and β be the fraction of data points that are missing fine-resolution magnetometer and fine-resolution EMI features, respectively. We conduct experiments using the elements of the binary Cartesian product of the set $S = \{0, 0.1, 0.25, 0.5, 0.75, 0.9\}$ as the pairs of amounts of missing primary (fine-resolution) data. For each of the 36 combinations considered, 100 independent trials are run. Each trial has a random partition of the data set into training and testing data, and randomly selected data points that are assumed to be missing the primary data. Note that primary data will be missing for both training and testing data.

This experimental setup was employed for three different amounts of training data: when 25%, 50%, and 75% of the data were labeled training data. All classification results shown are for the remaining unlabeled testing data. In all experiments, it was assumed that there was no labeling error ($\epsilon = 0$). In each experiment, four algorithms are applied; each of which handles

the multiresolution data in a different manner. However, a logistic regression classifier is used in all four methods, which are explained as follows.

The proposed approach builds a classifier for only the primary data; it handles missing primary data by integrating out the missing data, using the estimated density function relating both the primary and auxiliary data. This density function—a GMM—is accurately estimated using all available data, via the variational Bayesian expectation-maximization algorithm presented in [17]. Because class labels are not used in the estimation, both labeled and unlabeled data can be utilized. This fact ensures that the density function can be accurately estimated even when limited (labeled) training data are available.

The second method builds a separate classifier for data from each resolution. Building separate classifiers for data from each resolution in the case of a single sensor with two resolutions would entail that one classifier be built for features extracted from fine-resolution imagery, and a second classifier be built for features extracted from coarse-resolution imagery. The generalization of this case to multiple sensors with multiple resolutions is employed here as the second method. Specifically, four separate classifiers are constructed, one to handle each sensor-resolution pair combination. A more detailed explanation of this method is provided in the Appendix.

The third method builds a classifier for the concatenated primary and auxiliary data; it handles missing primary data by integrating out the missing data, via the approach used in [17]. The difference between this method and the proposed method is that this method builds a classifier for both auxiliary and primary data, whereas the proposed method does so only on the latter. The fourth method also builds a classifier for the concatenated primary and auxiliary data; however, this method imputes (i.e., "fills in") missing primary data with the unconditional mean of the observed data.

The area under a receiver operating characteristic (ROC) curve (AUC) is given by the Wilcoxon statistic [28]

$$\text{AUC} = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} 1_{x_m > y_n} \tag{23}$$

where $x_1, \ldots, x_M$ are the classifier outputs of data belonging to class 1, $y_1, \ldots, y_N$ are the classifier outputs of data belonging to class -1, and 1 is an indicator function. As a measure of classification performance, the AUC is a more useful quantity than accuracy (i.e., the fraction of classifications that are correct) when significant class imbalance exists, as it does in this data set. Moreover, the AUC can summarize performance more compactly than an ROC curve. For these reasons, we present the results of the classification experiments in terms of the AUC.

The results of all of the experiments are compactly summarized in Figs. 3–5. The results are displayed in these figures as images, which are interpolated from the results of the finite set of 36 pairs of missing-data conditions explained previously. Specifically, the images display the AUC values as a function of the amounts of missing fine-resolution magnetometer and missing fine-resolution EMI sensor data. The results from which the resulting images were interpolated were the mean AUC values over the 100 independent trials of the 36 pairs of conditions.

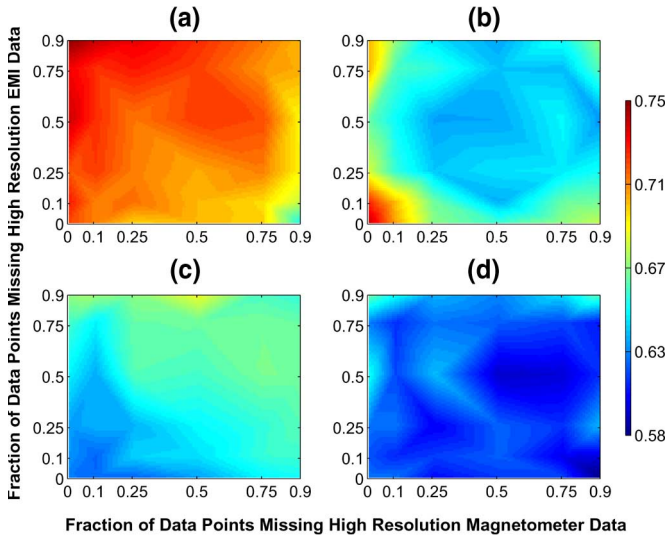

Fig. 3. Experimental results in terms of AUC when 25% of the data set is (labeled) training data. (a) Proposed method. (b) Four separate classifiers are built, one for each possible combination of missing data. (c) One classifier is built on all features, with missing data integrated out analytically. (d) One classifier is built on all features, with missing data handled via unconditional mean imputation.

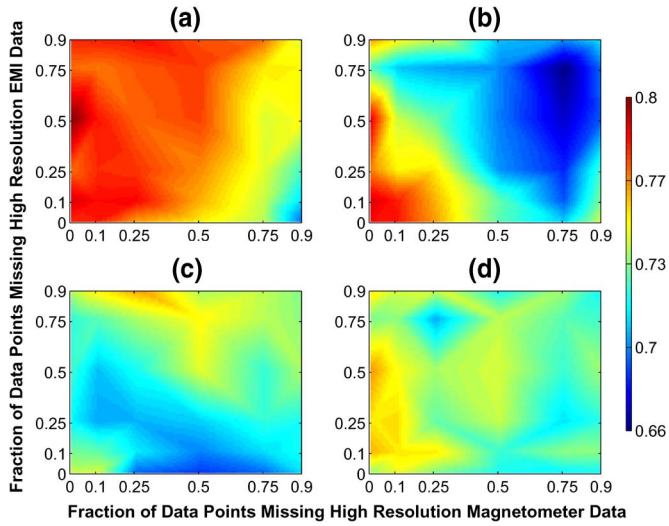

Fig. 4. Experimental results in terms of AUC when 50% of the data set is (labeled) training data. Refer to the caption of Fig. 3 for additional details.

The color scales are identical in the four panels within each figure, so visual comparisons among the methods' results can be made easily.

As can be seen from the three figures, the proposed method consistently performs better than the other three competing methods, regardless of the amounts of missing high-resolution data.

VI. Discussion

It should be emphasized that in the proposed method, the classifier weights are on the primary features, which are extracted from fine-resolution imagery. However, the auxiliary features extracted from coarse-resolution imagery are still utilized in the algorithm when primary data are missing.

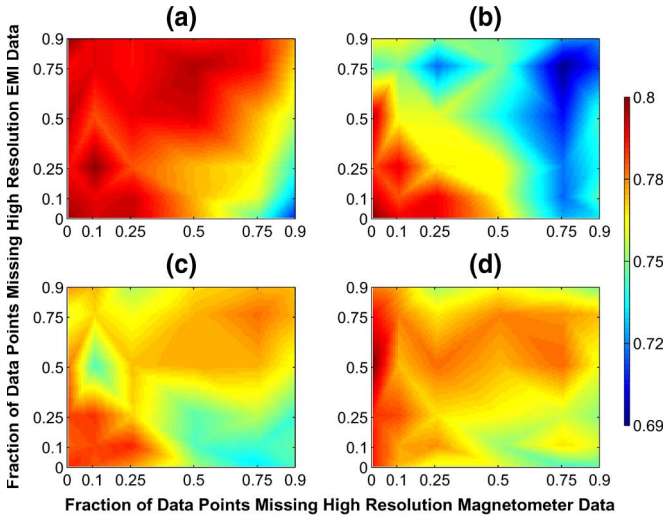

Fig. 5. Experimental results in terms of AUC when 75% of the data set is (labeled) training data. Refer to the caption of Fig. 3 for additional details.

Specifically, missing primary data are analytically integrated out via the estimated density function, which models the relationship between the features from coarse-resolution imagery and those same features from fine-resolution imagery. The experimental results consistently show that the proposed method outperforms the alternative methods. Here, we explain the reasons underlying this result.

Fine-resolution imagery contains salient aspects that are absent in coarse-resolution imagery. Therefore, features extracted from fine-resolution imagery should be preferred to features extracted from coarse-resolution imagery. Our proposed approach emphasizes the importance of the finer resolution data by building a classifier with only that data. The coarser resolution data are still exploited (via the estimated density function), albeit in an auxiliary role.

If a classifier is instead built on the conglomerated features extracted from different resolutions—as it was in two of the alternative methods—the information about the relative "quality" of the features is ignored. Concatenating features extracted from different resolution imagery also causes the feature dimension to grow quickly, which can in turn lead to overfitting of the training data. In theory, a prior could be incorporated to combat overfitting, but additional complications arise as a result of having incomplete data. In contrast, the proposed method has no such overfitting issues.

The proposed method also consistently outperforms the method that builds a separate classifier for data from each sensor-resolution combination. This result is possible because the proposed method utilizes side information in the form of the estimated density function. By exploiting the statistical relationship that exists among features at different resolutions (as well as among features from different sensors), better performance can be achieved. This result can perhaps best be understood from the viewpoint of superresolution techniques. Knowledge about a problem (e.g., that noise in an image is Gaussian) can be exploited to resolve a superresolution image from several blurry images [29]. Similarly, in this problem, knowledge of the statistical relationship between features at different resolutions can be exploited. Importantly, the proposed approach avoids the unnecessary intermediate step of forming an entire superresolution image; instead, the ultimate goal is addressed directly: obtaining the equivalent of "superresolution features."

VII. Conclusion

Acquiring fine-resolution imagery for all data points may be prohibitively expensive. For example, in the UXO detection problem, deploying ground-based sensors is dangerous and time-consuming. This paper presents a principled algorithm to classify imagery that is available at multiple resolutions. Because some data points may possess imagery at only a subset of resolutions, the problem can be viewed as one of incomplete-data classification. The proposed algorithm also naturally handles the case in which multiple sensor modalities—each of which may operate at multiple resolutions—are used to acquire data. In summary, the novel problem we addressed was of multisensor, multiresolution, and incomplete-data classification. Experimental results on a challenging UXO classification task employing magnetometer and EMI sensors demonstrated the advantage of the proposed algorithm over common alternatives.

Future work will focus on the development of an active data acquisition algorithm that determines which data points should receive finer resolution imagery—and at which particular resolution level—in order to most improve performance. This active sensing concept is relevant for many applications, including medical imaging, remote sensing, and video tracking.

Appendix

A. Modified Gradient Ascent

The classifier $\boldsymbol{w}$ from Section III is learned via a modified form of gradient ascent. This method uses the gradient and Hessian of the log likelihood, which we provide explicitly here. For convenience, we first rewrite the log-likelihood function (15) as

$$\ell(\boldsymbol{w}) = \sum_{i=1}^{N_L} \log \left[\epsilon_i + (1-2\epsilon_i) \sum_{k=1}^{K} \delta_k^i \sigma_k^i \right] \tag{24}$$

where

$$\sigma_k^i = \sigma\left(f_k^i\right) \tag{25}$$

$$f_k^i = y_i \beta \left(\boldsymbol{w}_{m_i}^T \xi_k^i + \boldsymbol{w}_{o_i}^T \boldsymbol{x}_i^{o_i} \right) \left(\gamma_k^i\right)^{-1} \tag{26}$$

$$\gamma_k^i = \sqrt{\boldsymbol{w}_{m_i}^T \Omega_k^i \boldsymbol{w}_{m_i} + \beta^2}. \tag{27}$$

The gradient of the log likelihood is

$$\frac{\partial \ell(\boldsymbol{w})}{\partial \boldsymbol{w}} = \sum_{i=1}^{N_L} \frac{\left[(1-2\epsilon_i) \sum_{k=1}^{K} \delta_k^i \sigma_k^i (1-\sigma_k^i) \frac{\partial f_k^i}{\partial \boldsymbol{w}} \right]}{\left[\epsilon_i + (1-2\epsilon_i) \sum_{k=1}^{K} \delta_k^i \sigma_k^i \right]} \tag{28}$$

while the Hessian of the log likelihood is shown in (29) at the top of the next page

$$\frac{\partial^2 \ell(\boldsymbol{w})}{\partial \boldsymbol{w} \partial \boldsymbol{w}^T} = \sum_{i=1}^{N_L} \left\{ \frac{\left[2(1-2\epsilon_i)\sum_{k=1}^{K} \delta_k^i \sigma_k^i \left(1-\sigma_k^i\right)^2 \frac{\partial^2 f_k^i}{\partial \boldsymbol{w} \partial \boldsymbol{w}^T}\right]}{\left[\epsilon_i + (1-2\epsilon_i)\sum_{k=1}^{K} \delta_k^i \sigma_k^i\right]} - \frac{\left[(1-2\epsilon_i)\sum_{k=1}^{K} \delta_k^i \sigma_k^i \left(1-\sigma_k^i\right) \frac{\partial f_k^i}{\partial \boldsymbol{w}}\right]\left[\sum_{k=1}^{K} \delta_k^i (1-2\epsilon_i)\, \sigma_k^i \left(1-\sigma_k^i\right) \frac{\partial f_k^i}{\partial \boldsymbol{w}}\right]^T}{\left[\epsilon_i + (1-2\epsilon_i)\sum_{k=1}^{K} \delta_k^i \sigma_k^i\right]^2} \right\} \tag{29}$$

where

$$\frac{\partial f_k^i}{\partial \boldsymbol{w}} = \begin{bmatrix} \frac{\partial f_k^i}{\partial \boldsymbol{w}_{o_i}} \\ \frac{\partial f_k^i}{\partial \boldsymbol{w}_{m_i}} \end{bmatrix} = \begin{bmatrix} \frac{y_i \beta \boldsymbol{x}_i^{o_i}}{\gamma_k^i} \\ \frac{y_i \beta \xi_k^i}{\gamma_k^i} - \frac{y_i \beta \left(\boldsymbol{w}_{m_i}^T \xi_k^i + \boldsymbol{w}_{o_i}^T \boldsymbol{x}_i^{o_i}\right) \Omega_k^i \boldsymbol{w}_{m_i}}{\left(\gamma_k^i\right)^3} \end{bmatrix} \tag{30}$$

$$\frac{\partial^2 f_k^i}{\partial \boldsymbol{w} \partial \boldsymbol{w}^T} = \begin{bmatrix} \frac{\partial^2 f_k^i}{\partial \boldsymbol{w}_{o_i} \partial \boldsymbol{w}_{o_i}^T} & \frac{\partial^2 f_k^i}{\partial \boldsymbol{w}_{o_i} \partial \boldsymbol{w}_{m_i}^T} \\ \frac{\partial^2 f_k^i}{\partial \boldsymbol{w}_{m_i} \partial \boldsymbol{w}_{o_i}^T} & \frac{\partial^2 f_k^i}{\partial \boldsymbol{w}_{m_i} \partial \boldsymbol{w}_{m_i}^T} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \frac{-y_i \beta \boldsymbol{x}_i^{o_i} \left(\Omega_k^i \boldsymbol{w}_{m_i}\right)^T}{\left(\gamma_k^i\right)^3} \\ \frac{-y_i \beta \Omega_k^i \boldsymbol{w}_{m_i} \left(\boldsymbol{x}_i^{o_i}\right)^T}{\left(\gamma_k^i\right)^3} & \frac{\partial^2 f_k^i}{\partial \boldsymbol{w}_{m_i} \partial \boldsymbol{w}_{m_i}^T} \end{bmatrix} \tag{31}$$

$$\frac{\partial^2 f_k^i}{\partial \boldsymbol{w}_{m_i} \partial \boldsymbol{w}_{m_i}^T} = \frac{-y_i \beta}{(\gamma_k^i)^3} \left[\xi_k^i (\Omega_k^i \boldsymbol{w}_{m_i})^T + (\boldsymbol{w}_{m_i}^T \xi_k^i + \boldsymbol{w}_{o_i}^T \boldsymbol{x}_i^{o_i}) \Omega_k^i + \Omega_k^i \boldsymbol{w}_{m_i} (\xi_k^i)^T\right] + \frac{3 y_i \beta (\boldsymbol{w}_{m_i}^T \xi_k^i + \boldsymbol{w}_{o_i}^T \boldsymbol{x}_i^{o_i}) \Omega_k^i \boldsymbol{w}_{m_i} (\Omega_k^i \boldsymbol{w}_{m_i})^T}{(\gamma_k^i)^5}. \tag{32}$$

B. Classification Method 2

Here, we explain in greater detail the second classification method used in the experiments. Define sensor 1 to be the magnetometer, and define sensor 2 to be the EMI sensor. Recall that data from two image resolutions are available for each of the two sensors. Let γ_x^s be the set of data points for which primary data from the sth sensor is possessed; let γ_z^s be the set of data points for which auxiliary data from the sth sensor is possessed. Context will elucidate whether the sets contain training or testing data points. Note that $\gamma_x^s \subseteq \gamma_z^s$ in all experiments in this paper because it is assumed that auxiliary data are available for all data points. Let $\gamma_z^s \backslash \gamma_x^s$ denote the set of data points in γ_z^s but not in γ_x^s. Table I compactly summarizes the manner in which the various classifiers of this method are constructed and utilized.

For example, a training data point that has both fine-resolution and coarse-resolution magnetometer data and both fine-resolution and coarse-resolution EMI sensor data would be used in the construction of all four classifiers. A testing data point that has both fine-resolution and coarse-resolution magnetometer data but only coarse-resolution EMI sensor data would be evaluated (i.e., classified) using classifier 2.

TABLE I
EXPLANATION OF CLASSIFICATION METHOD 2 OF THE EXPERIMENTS

Classifier	Features on Which the Classifier Is Built	Training Data Points Used to Train Classifier	Testing Data Points Evaluated by Classifier
1	$[\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}]$	$\gamma_x^1 \cap \gamma_x^2$	$\gamma_x^1 \cap \gamma_x^2$
2	$[\boldsymbol{x}^{(1)}, \boldsymbol{z}^{(2)}]$	$\gamma_x^1 \cap \gamma_z^2$	$\gamma_x^1 \cap (\gamma_z^2 \backslash \gamma_x^2)$
3	$[\boldsymbol{z}^{(1)}, \boldsymbol{x}^{(2)}]$	$\gamma_z^1 \cap \gamma_x^2$	$(\gamma_z^1 \backslash \gamma_x^1) \cap \gamma_x^2$
4	$[\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}]$	$\gamma_z^1 \cap \gamma_z^2$	$(\gamma_z^1 \backslash \gamma_x^1) \cap (\gamma_z^2 \backslash \gamma_x^2)$

To summarize, in the training stage, all data points that possess the requisite features are used to train the classifier. This arrangement allows more data to be used in building the classifiers and, hence, allows more accurate classifiers to be obtained. In the testing stage, a given testing data point is submitted to that classifier that fully exploits the fine-resolution features that the data point possesses.

REFERENCES

[1] S. Billings, C. Pasion, S. Walker, and L. Berans, "Magnetic models of unexploded ordnance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2115–2124, Aug. 2006.

[2] S. Chilaka, D. Faircloth, L. Riggs, and H. Nelson, "Enhanced discrimination among UXO-like targets using extremely low-frequency magnetic fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 10–21, Jan. 2006.

[3] C. Moss, T. Grzegorczyk, K. O'Neill, and J. Kong, "A hybrid time-domain model of electromagnetic induction from conducting, permeable targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2916–2926, Oct. 2006.

[4] K. Sun, K. O'Neill, F. Shubitidze, I. Shamatava, and K. Paulsen, "Fast data-derived fundamental spheroidal excitation models with application to UXO discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 11, pp. 2573–2583, Nov. 2005.

[5] J. Stalnaker, M. Everett, A. Benavides, and C. Pierce, "Mutual induction and the effect of host conductivity on the EM induction response of buried plate targets using 3-d finite-element analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 251–259, Feb. 2006.

[6] S. Hart, R. Shaffer, S. Rose-Pehrsson, and J. McDonald, "Using physics-based modeler outputs to train probabilistic neural networks for unexploded ordnance (UXO) classification in magnetometry surveys," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 4, pp. 797–804, Apr. 2001.

[7] C. Nelson, C. Cooperman, W. Schneider, D. Wenstrand, and D. Smith, "Wide bandwidth time-domain electromagnetic sensor for metal target classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1129–1138, Jun. 2001.

[8] Y. Zhang, L. Collins, H. Yu, C. Baum, and L. Carin, "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.

[9] A. Willsky, "Multiresolution Markov models for signal and image processing," *Proc. IEEE*, vol. 90, no. 8, pp. 1396–1458, Aug. 2002.
[10] H. Choi and R. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden Markov models," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1322–1331, Sep. 2001.
[11] J. Li, R. Gray, and R. Olshen, "Multiresolution image classification by hierarchical modeling with two dimensional hidden Markov models," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1826–1841, Aug. 2000.
[12] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.
[13] C. Li and R. Wilson, "Image segmentation based on a multiresolution Bayesian framework," in *Proc. IEEE Int. Conf. Image Process.*, 1998, pp. 761–765.
[14] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 2, pp. 99–113, Feb. 1991.
[15] C. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Process.*, vol. 3, no. 2, pp. 162–177, Mar. 1994.
[16] D. Rubin, *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: Wiley, 1987.
[17] D. Williams, X. Liao, Y. Xue, and L. Carin, "Incomplete-data classification using logistic regression," in *Proc. Int. Conf. Mach. Learn.*, 2005, pp. 977–984.
[18] M. Opper and O. Winther, "Gaussian processes and SVM: Mean field and leave-one-out," in *Advances in Large Margin Classifiers*, A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds. Cambridge, MA: MIT Press, 2000, pp. 311–326.
[19] P. McCullagh and J. Nelder, *Generalized Linear Models*, 2nd ed. London, U.K.: Chapman & Hall, 1989.
[20] D. Williams, "Classification and data acquisition with incomplete data," Ph.D. dissertation, Duke Univ., Durham, NC, 2006.
[21] C. Baum, Ed., *Detection and Identification of Visually Obscured Targets*. New York: Taylor & Francis, 1998.
[22] L. Carin, H. Yu, Y. Dalichaouch, A. Perry, P. Czipott, and C. Baum, "On the wideband EMI response of a rotationally symmetric permeable and conducting target," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1206–1213, Jun. 2001.
[23] N. Geng, C. 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.
[24] W. Panofsky, *Classical Electricity and Magnetism*. Reading, MA: Addison-Wesley, 1962.
[25] X. Liao and L. Carin, "Application of the theory of optimal experiments to adaptive electromagnetic-induction sensing of buried targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 961–972, Aug. 2004.
[26] W. Press, B. F. S. Teukolsky, and W. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.
[27] H. Nelson and J. McDonald, "Multisensor towed array detection system for UXO detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1139–1145, Jun. 2001.
[28] J. Hanley and B. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982.
[29] R. Tsai and T. Huang, "Multi-frame image restoration and registration," *Adv. Comput. Vis. Image Process.*, vol. 1, pp. 317–339, 1984.

David Williams (S'99–M'06) received the B.S.E. (*magna cum laude*), M.S., and Ph.D. degrees in 2002, 2003, and 2006, respectively, from Duke University, Durham, NC.

While at Duke University, he was the recipient of a James B. Duke Graduate Fellowship and a National Defense Science and Engineering Graduate Fellowship. His principal technical interests lie in the fields of machine learning and automatic target recognition.

Chunping Wang received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2001 and 2003, respectively. She is currently working toward the Ph.D. degree in electrical and computer engineering at Duke University, Durham, NC.

Her current research interests are machine learning and statistical pattern recognition.

Xuejun Liao (SM'04) was born in Qinghai, China. He received the B.S. and M.S. degrees in electrical engineering, from Hunan University, Changsha, China, in 1990 and 1993, respectively, and the Ph.D. degree in electrical engineering from Xidian University, Xi'an, China, in 1999.

From 1993 to 1995, he was with the Department of Electrical Engineering, Hunan University, working on electronic instruments. From 1995 to 2000, he was with the National Key Laboratory for Radar Signal Processing, Xidian University, working on automatic target recognition and radar imaging. Since May 2000, he has been working as a Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC. His current research interests are in planning under uncertainty, machine learning, bioinformatics, and signal and image processing.

Lawrence Carin (F'01) was born March 25, 1963 in Washington, DC, and received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1985, 1986, and 1989, respectively.

In 1989, he was with the Electrical Engineering Department, Polytechnic University, Brooklyn, as an Assistant Professor, and he became an Associate Professor in 1994. In September 1995, he was with the Electrical Engineering Department, Duke University, Durham, NC, where he is currently the William H. Younger Professor of Engineering. He was the Principal Investigator (PI) on a Multidisciplinary University Research Initiative (MURI) on demining (1996–2001), and he is currently the PI of a MURI dedicated to multimodal inversion. He was the Cofounder of Signal Innovations Group, Inc. (SIG), which is now a subsidiary of Integrian, Inc.; he serves as the Director of Technology at SIG. His current research interests include signal processing, sensing, and machine learning.

Dr. Carin was an Associate Editor of the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION from 1995 to 2004. He is a member of the Tau Beta Pi and Eta Kappa Nu honor societies.