Multiple instance dictionary learning for subsurface object detection using handheld EMI

Alina Zare[a], Matthew Cook[a], Brendan Alvey[a], Dominic K. Ho[a]

[a]University of Missouri, Columbia, MO 65211

ABSTRACT

A dictionary learning approach for subsurface object detection using handheld electromagnetic induction (EMI) data is presented. A large number of unsupervised and supervised dictionary learning methods have been developed in the literature. However, the majority of these methods require data point-specific labels during training. In the application to subsurface object detection, often the specific training data samples that correspond to target and non-target are not known and difficult to determine manually. In this paper, a dictionary learning method that addresses this issue using the multiple instance learning techniques is presented. Results are shown on real EMI data sets.

Keywords: Dictionary Learning, Landmine Detection, Electromagnetic Induction, Extended Functions of Multiple Instances

1. INTRODUCTION

Electromagnetic induction (EMI) sensors have been widely investigated for use in the detection of buried landmines. In this investigation, the data used was collected using an EMI sensor that was alongside a ground-penetrating radar on a handheld system built to detect landmines. The two sensors work individually to detect different types of mines and then the results can be fused to boost detection performance. Previous work has explored using a fixed dictionary to model the EMI data from this sensor.[1] In this paper, we propose a method to estimate a dictionary from the data instead of modeling the data with a fixed parametrically-defined dictionary.

The rest of this paper is organized as follows: In Section 2, the Discrete Spectrum of Relaxation Frequencies (DSRF) dictionary is introduced. In Section 3 the Extended Functions of Multiple Instances algorithm is detailed. In Section 4 results on EMI data are presented and discussed. Lastly in Section 5 summarizes the paper and discusses potential future work.

2. DISCRETE SPECTRUM OF RELAXATION FREQUENCIES DICTIONARY

A common model used to represent EMI data in algorithms for the detection of buried landmines is the Discrete Spectrum of Relaxation Frequencies (DSRF),[1]

$$H(w) = c_0 + \sum_{k=1}^{K} \frac{c_k}{1 + j\omega/\zeta_k}, \tag{1}$$

where c_0 is the DC shift, c_k is the real spectral amplitudes, and ζ_k are the relaxation frequencies. This model has been previously used for landmine detection as a feature generator by computing the parameters for each point.[2,3]

Given this model, a fixed dictionary can be created by varying over a large set of ζ_k values given a few fixed values of ω. This results in a dictionary where each column takes the form,

$$\boldsymbol{d}_i = \left[\frac{1}{1 + j\omega_1/\zeta_i}, \frac{1}{1 + j\omega_2/\zeta_i}, \ldots, \frac{1}{1 + j\omega_N/\zeta_i} \right]^T, \tag{2}$$

which is graphically shown in Figure 1.

Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX,
edited by Steven S. Bishop, Jason C. Isaacs, Proc. of SPIE Vol. 9454, 94540G
© 2015 SPIE · CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2179177

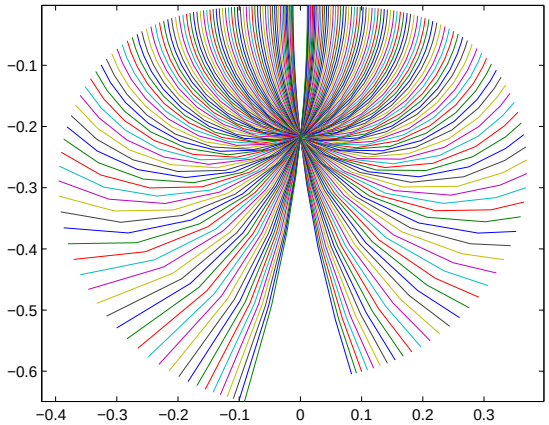


Figure 1: Argand diagrams of the DSRF dictionary

3. EXTENDED FUNCTIONS OF MULTIPLE INSTANCES

Functions of Multiple Instances (FUMI) is a method for dictionary learning that models each data point as a linear combination of a group of dictionary elements.[4] When the current point is labeled as a target point (i.e., representing a target of interest) the linear combination can be written as

$$\boldsymbol{x}_n^+ = w_{nT}\boldsymbol{d}_T + \sum_{m=1}^{M} w_{nm}\boldsymbol{d}_m, \tag{3}$$

then, for points labeled as non-target the linear combination can be written as

$$\boldsymbol{x}_n^- = \sum_{m=1}^{M} w_{nm}\boldsymbol{d}_m. \tag{4}$$

In both equations w_{nT} is the weight of the target element, w_{nm} is the weight of the m^{th} non-target dictionary element, d_T is the target dictionary element, and d_m is the m^{th} element in the non-target dictionary. In these equations all of the weights are constrained such that they are positive and sum to one.

The FUMI model and associated algorithms work well with the accurate data point labeling of training points. However, the above model does not effectively address the case when there are label inaccuracies. The Extended Functions of Multiple Instances (eFUMI)[4,5] aims to fix this problem. To do this the eFUMI no longer considers point-wise labeling and, instead, uses "bag" labels (i.e., labels for sets of data points). In this way, a group of points is labeled as positive if it contains at least one target point, and a bag is labeled as non-target only if it contains exclusively non-target points.

With the new method of assigning labels to training points the objective function for the eFUMI model takes the following form,

$$F = \frac{-(1-u)}{2}\sum_{n=1}^{N}\left\|\boldsymbol{x}_n - z_n w_{nT}\boldsymbol{d}_T - \sum_{m=1}^{M} w_{nm}\boldsymbol{d}_m\right\|_2^2 - \frac{u}{2}\left(\sum_{m=1}^{M}\|\boldsymbol{d}_m - \boldsymbol{\mu}_0\|_2^2 + \|\boldsymbol{d}_T - \boldsymbol{\mu}_0\|_2^2\right) - \sum_{m=1}^{M}\gamma_m\sum_{n=1}^{N} w_{nm}, \tag{5}$$

where the first term is a measure of the error between the reconstruction of the actual signal x_i given the current target element d_T, non-target elements d_m, and weight values w_{nT} and w_{nm}. The parameter z_n is a hidden latent variable indicating whether the nth point is or is not a target point. The second term encourages the dictionary elements to be close to the global data mean μ_0, and the last term is a sparsity promoting term for the proportions w_{nm} where $\gamma_k = \frac{\Gamma}{\sum_{n=1}^{N} w_{nm}^{n-1}}$ to determine the number of needed background dictionary elements.

In order to minimize Equation 5, an Expectation Maximization (EM) algorithm is used. When taking the expectation with respect to the two possible values of z_n the objective function becomes

$$E[F] = \sum_{z_n \in \{0,1\}} \left[-\frac{1}{2}(1-u) \sum_{n=1}^{N} P(z_n|\mathbf{x}_n, \boldsymbol{\theta}^{(t-1)}) \left\| \mathbf{x}_n - z_n w_{nT} \mathbf{d}_T - \sum_{m=1}^{M} w_{nm} \mathbf{d}_m \right\|_2^2 \right] \tag{6}$$

$$-\frac{u}{2} \sum_{m=1}^{M} \|\mathbf{d}_m - \boldsymbol{\mu}_0\|_2^2 - \frac{u}{2} \|\mathbf{d}_T - \boldsymbol{\mu}_0\|_2^2 - \sum_{m=1}^{M} \gamma_m \sum_{n=1}^{N} w_{nm}, \tag{7}$$

where

$$P(z_n|\mathbf{x}_n, \theta^{(t-1)}) = \begin{cases} p(z_n = 0|\mathbf{x}_n \in B_j^+, \boldsymbol{\theta}^{(t-1)}) = e^{-\beta \left\| \mathbf{x}_n - \sum_{m=1}^{M} w_{nm} \mathbf{d}_m \right\|_2^2} \\ p(z_n = 1|\mathbf{x}_n \in B_j^+, \boldsymbol{\theta}^{(t-1)}) = 1 - e^{-\beta \left\| \mathbf{x}_n - \sum_{m=1}^{M} w_{nm} \mathbf{d}_m \right\|_2^2} \\ p(z_n = 0|\mathbf{x}_n \in B_j^-, \boldsymbol{\theta}^{(t-1)}) = 1 \\ p(z_n = 1|\mathbf{x}_n \in B_j^-, \boldsymbol{\theta}^{(t-1)}) = 0 \end{cases}. \tag{8}$$

From these equations the closed-form update equations for the dictionary elements and weight values can be found. Optimization of the objective function can then be achieved using alternating optimization. The full algorithm for implementing eFUMI training is shown in Algorithm 1.

Algorithm 1 *e*FUMI EM algorithm

Intialize $\theta^0 = \{\mathbf{d}_T, \mathbf{D}, \mathbf{W}\}$, $t = 1$
repeat
  E-step: Compute $p(z_n|\mathbf{x}_n, \theta^{(t-1)})$
  M-Step:
    Update $\mathbf{d}_T$ by minimizing the expectation of the log likelihood
    Update $\mathbf{D}$ by minimizing the expectation of the log likelihood
    Update $\mathbf{W}$ by minimizing the expectation of the log likelihood
  $\theta^t \leftarrow \{\mathbf{d}_T, \mathbf{D}, \mathbf{W}\}$
  $t \leftarrow t + 1$
until Convergence
return $\theta^t = \{\mathbf{d}_T, \mathbf{D}, \mathbf{W}\}$

3.1 Multi-Target eFUMI

If given a data set with a diverse set of target signatures, then the above eFUMI algorithm must be trained several times (once for each unique target signature) to build a large enough dictionary to account for all the different variations in target. This is due to the fact that the above the eFUMI model does not learn multiple target dictionary elements. In order to learn several target elements simulatenously, the model can be modified slightly as shown below

$$F = -\frac{1}{2}(1-u) \sum_{i=1}^{N} w_i \left\| (\boldsymbol{x}_i - z_i \sum_{t=1}^{T} p_{it} \boldsymbol{e}_t - \sum_{k=1}^{M} p_{ik} \boldsymbol{e}_k) \right\|_2^2 - \frac{u}{2} \sum_{k=1}^{M} \|(\boldsymbol{e}_k - \boldsymbol{\mu}_0)\|_2^2 - \frac{u}{2} \|(\boldsymbol{e}_T - \boldsymbol{\mu}_0)\|_2^2 - \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} p_{ik}. \tag{9}$$

Given this model, an EM algorithm can be derived. The expectation with respect to z_n mimics Equation 7 and is shown below,

$$E[F] = \sum_{z_n \in \{0,1\}} \left[-\frac{1}{2}(1-u) \sum_{n=1}^{N} P(z_n|\mathbf{x}_n, \boldsymbol{\theta}^{(t-1)}) \left\| \mathbf{x}_n - z_n \sum_{t=1}^{T} p_{it} \boldsymbol{e}_t - \sum_{m=1}^{M} w_{nm} \mathbf{d}_m \right\|_2^2 \right] \tag{10}$$

$$-\frac{u}{2} \sum_{m=1}^{M} \|\mathbf{d}_m - \boldsymbol{\mu}_0\|_2^2 - \frac{u}{2} \|\mathbf{d}_T - \boldsymbol{\mu}_0\|_2^2 - \sum_{m=1}^{M} \gamma_m \sum_{n=1}^{N} w_{nm}, \tag{11}$$

where

$$P(z_n|\mathbf{x}_n, \theta^{(t-1)}) = \begin{cases} p(z_n = 0|\mathbf{x}_n \in B_j^+, \boldsymbol{\theta}^{(t-1)}) = e^{-\beta \left\| \mathbf{x}_n - \sum_{m=1}^{M} w_{nm} \mathbf{d}_m \right\|_2^2} \\ p(z_n = 1|\mathbf{x}_n \in B_j^+, \boldsymbol{\theta}^{(t-1)}) = 1 - e^{-\beta \left\| \mathbf{x}_n - \sum_{m=1}^{M} w_{nm} \mathbf{d}_m \right\|_2^2} \\ p(z_n = 0|\mathbf{x}_n \in B_j^-, \boldsymbol{\theta}^{(t-1)}) = 1 \\ p(z_n = 1|\mathbf{x}_n \in B_j^-, \boldsymbol{\theta}^{(t-1)}) = 0 \end{cases}. \tag{12}$$

The algorithm with multiple target elements follows the same alternating optimization procedure as the algorithm outlined in Alg. 1.

4. EXPERIMENTS AND RESULTS

4.1 Data Set Description

Data was collected using a handheld Electromagnetic Induction (EMI) sensor. Buried targets were categorized by their metallic content and purpose. The metallic content of a target was defined as either metal, low-metal or non-metal. The purpose of a target was assigned as either anti-personnel (AP), anti-tank (AT) or clutter (CL). In total there were 428 buried objects in the data set at various depths. A full summary of the data set is shown in Table 1.

Table 1: Data set summary

	AP	**AT**	**CL**	**Totals**
Metal	43	74	18	**135**
Low	35	118	0	**153**
Non	0	76	64	**140**
Totals	**78**	**268**	**82**	**428**

4.2 Adaptive Coherence Estimator

In the following results, the Adaptive Coherence Estimator (ACE)[6] was applied as the method for target detection given the target signatures estimated using the above eFUMI algorithm. The detector computes a detection statistics based on the angle between a reference target signature and a data point given an estimated background distribution. The ACE detection statistic is computed as,

$$ACE(\boldsymbol{x}) = \frac{[(\boldsymbol{s} - \boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_b^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_b)]^2}{[(\boldsymbol{s} - \boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_b^{-1} (\boldsymbol{s} - \boldsymbol{\mu}_b)][(\boldsymbol{x} - \boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_b^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_b)]}, \tag{13}$$

where x is the current data point, s is the target dictionary element, μ_b is the background mean, and Σ_b^{-1} is the background covariance. In our implementation, the mean of the background is estimated locally using an RX-style computation with a sliding window and a guard band. The background covariance is estimated globally. Confidence maps are computed using each of the dictionary elements individually. The maps are then aggregated together using the max across each target dictionary element.

4.3 Alarm-based EFUMI Results

The eFUMI algorithm was first applied to estimate the target signature given several different instances of the same target type. Target and background alarms were found using the JOMP algorithm.[3] The eFUMI algorithm estimated a target signtures given several encounters of each target type. The target signatures for each target type was trained several times using a 10-fold cross-validation scheme. The results of four of these folds for a few target types are shown in Figures 2 through 6.

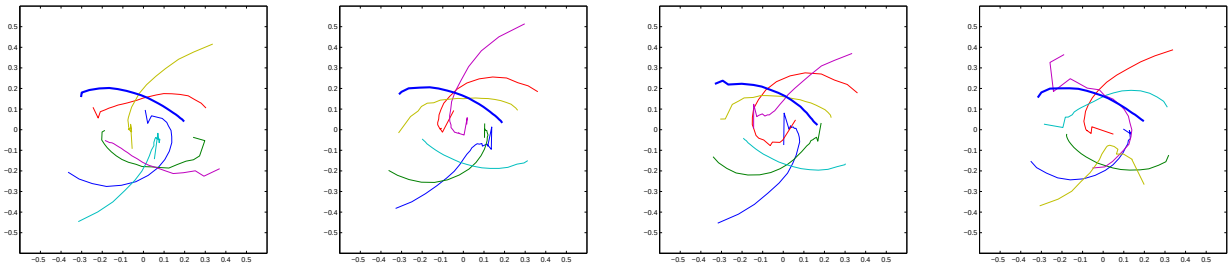


Figure 2: eFUMI estimated target dictionary element for Target Type 2, target dictionary element is shown as the thicker blue line.

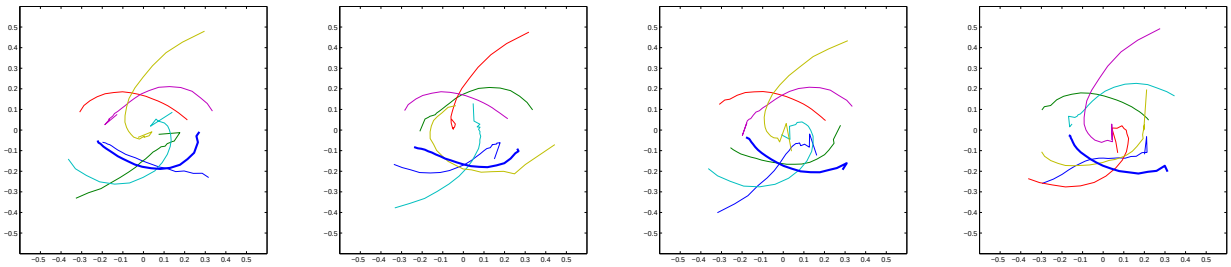


Figure 3: eFUMI estimated target dictionary element for Target Type 11, target dictionary element is shown as the thicker blue line.

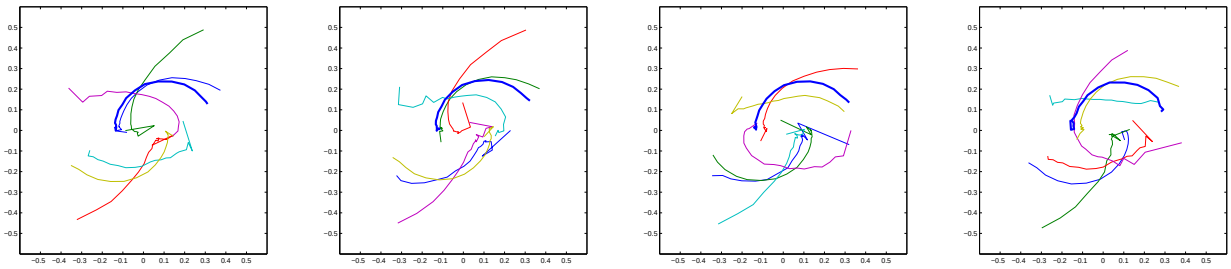


Figure 4: eFUMI estimated target dictionary element for Target Type 9, target dictionary element is shown as the thicker blue line.

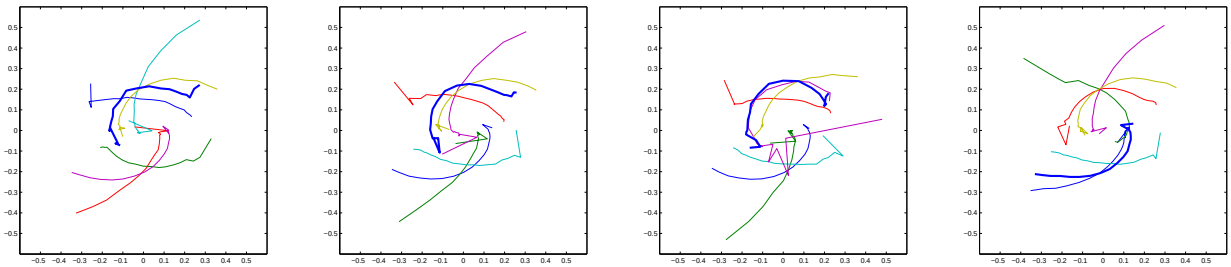


Figure 5: eFUMI estimated target dictionary element for Target Type 18, target dictionary element is shown as the thicker blue line.

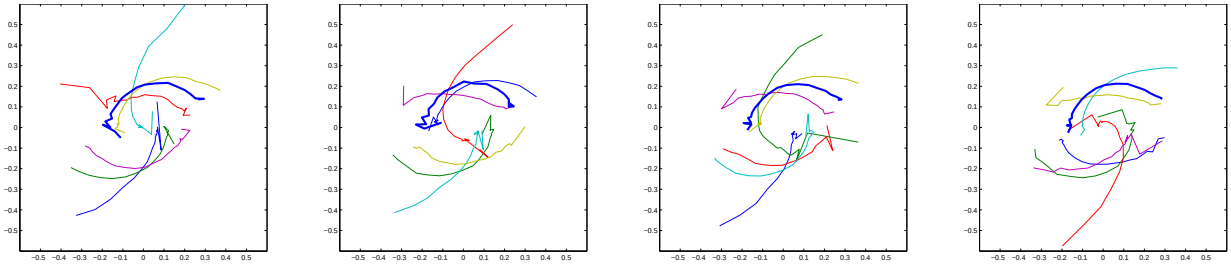


Figure 6: eFUMI estimated target dictionary element for Target Type 18, target dictionary element is shown as the thicker blue line.

As can be seen in each of figure, the estimated target signature is consistent over several folds of the cross-validation scheme.

Using the 10-fold cross-validation scheme, the ACE detector, Equation 13, was used to compare the estimate eFUMI dictionary to the fixed DSRF dictionary shown in Section 2. The resulting ROC curves are shown in Figure 7.

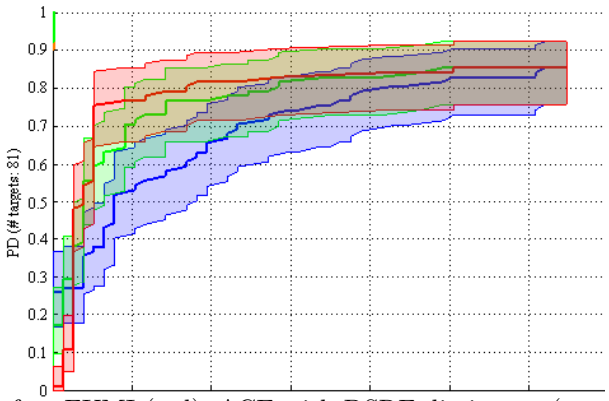


Figure 7: ROC curves for eFUMI (red), ACE with DSRF dictionary (green), and JOMP (blue)

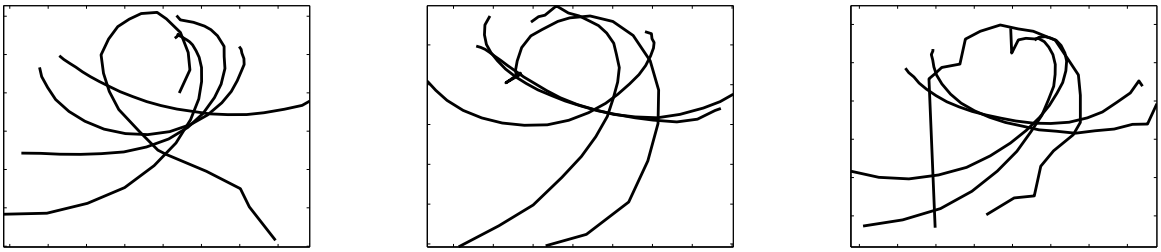
The ROC curve shows three different curves, two of them use the fixed DSRF dictionary with the ACE detector and the with the JOMP[3] detection algorithm. From the ROC curve, it can be seen that the estimated dictionary out-performs the other two methods tested.

4.4 MT-EFUMI Learned Dictionaries with ACE Confidence Maps

In this experiment three different lanes from the data set described above were considered individually for training and testing the MT-eFUMI algorithm. Bags of labeled data were generated by running the ACE detector using

the DSRF dictionary and labeling all data points with a ACE confidence above 0.4 as being a target and all other points were labeled as non-target points.

In figure 8 the target dictionaries that are learned by MT-eFUMI are shown.



(a) (b) (c)

Figure 8: Argand diagrams of the target dictionaries learned by Multi-Target eFUMI. Each subfigure is matched to its corresponding subfigure in Figure 9.

The ACE detector described in Equation 13 was then applied using the estimated target signatures. The ACE confidence maps are shown in Figure 9

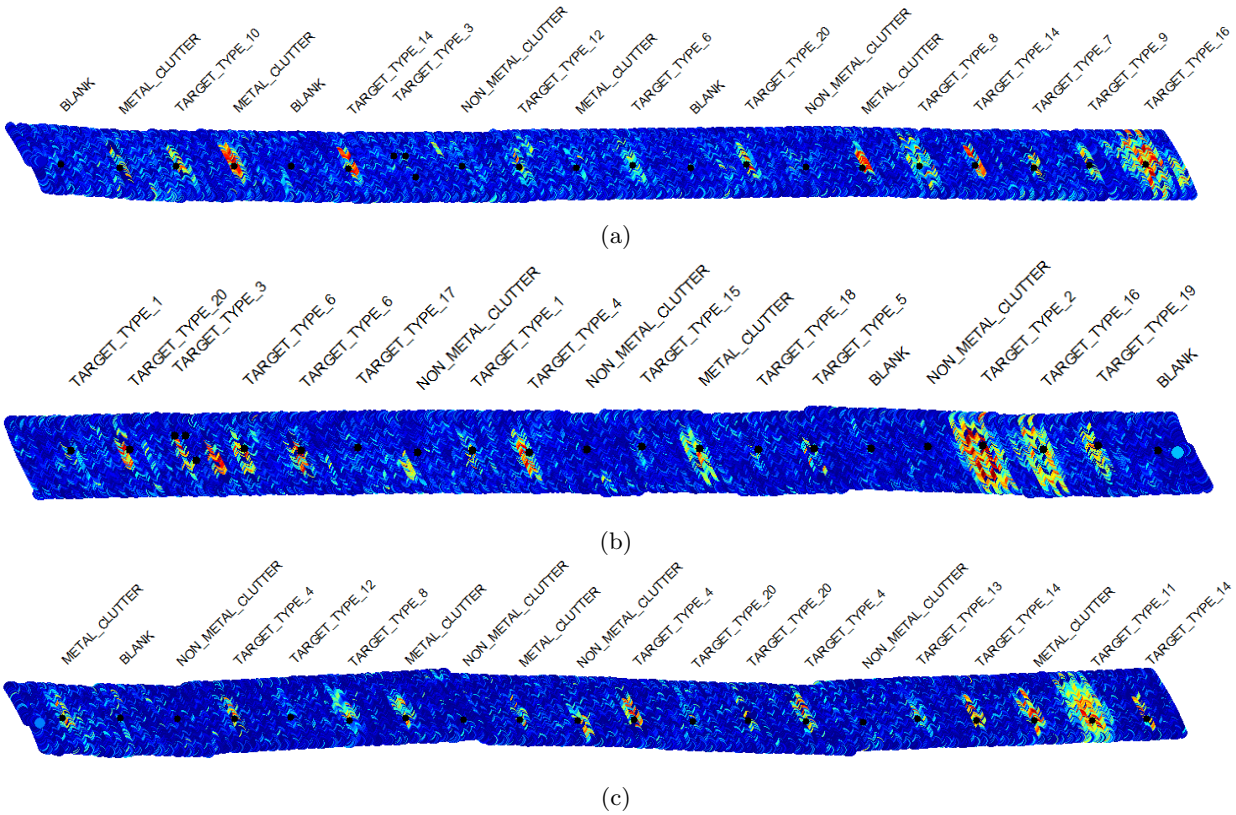


Figure 9: ACE confidence maps for three different lanes of EMI data. All maps are on a common color scale of [0 1] (0 is blue and 1 is red).

These figures show that the dictionary learned from MT-eFUMI can accurately describe most of the targets present in each lane.

5. SUMMARY

This paper described the eFUMI algorithm and its application to an EMI data set for the purpose of detecting buried targets. This method of dictionary learning showed improvement over the fixed DSRF dictionary used in previous methods and allows for uncertain labeling of the data.

ACKNOWLEDGMENTS

This work was funded by US Army Research Office grant number 66398-CS to support the US Army RDECOM CERDEC NVESD. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies either expressed or implied, of the US Army Research Office, US Army Research Laboratory, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

[1] Wei, M.-H., Scott, W., and McClellan, J., "Robust estimation of the discrete spectrum of relaxations for electromagnetic induction responses," *Geoscience and Remote Sensing, IEEE Transactions on* **48**, 1169–1179 (March 2010).

[2] Wei, M.-H., Scott, W., and McClellan, J., "Landmine detection using the discrete spectrum of relaxation frequencies," in [*Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*], 834–837 (July 2011).

[3] Goldberg, S., Glenn, T., Wilson, J. N., and Gader, P. D., "Landmine detection using two-tapped joint orthogonal matching pursuits," *Proc. SPIE* **8357**, 83570B–83570B–8 (2012).

[4] Jiao, C. and Zare, A., "Functions of multiple instances for learning target signatures," (2015). In Press.

[5] Zare, A. and Jiao, C., "Extended functions of multiple instances for target characterization," in [*6th IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*], (June 2014).

[6] Kraut, S., Scharf, L., and Butler, R., "The adaptive coherence estimator: a uniformly most-powerful-invariant adaptive detection statistic," *IEEE Transactions on Signal Processing* **53**, 427–438 (Feb 2005).