Functions of Multiple Instances for Learning Target Signatures

Changzhe Jiao, *Student Member, IEEE*, and Alina Zare, *Senior Member, IEEE*

Abstract—The functions of multiple instances (FUMI) approach for learning target and nontarget signatures is introduced. FUMI is a generalization of the multiple-instance learning (MIL) approach for supervised learning. FUMI differs significantly from standard MIL and supervised learning approaches because only data points which are functions of class concepts/signatures are available. In particular, this paper addresses the problem in which data points are convex combinations of target and nontarget signatures. Two algorithms, convex FUMI (cFUMI) and extended cFUMI (eFUMI), are presented and applied to the problem of hyperspectral unmixing and target detection. cFUMI learns target and nontarget signatures (i.e., target and nontarget endmembers), the number of nontarget signatures, and the proportion of each signature for every data point. The eFUMI algorithm extends the cFUMI to allow for additional "bag level" uncertainty in training labels. For these methods, training data need only binary labels indicating whether a data point (or some spatial area in the case of eFUMI) contains or does not contain some proportion of the target; the specific target proportions for the training data are not needed. After learning the target signature using the binary-labeled training data, target detection can be performed on test data. Results for subpixel target detection on simulated and real airborne hyperspectral data are shown.

Index Terms—Background, detection, endmember, hyperspectral, multiple-instance learning (MIL), subpixel, target, unmixing.

I. INTRODUCTION

HYPERSPECTRAL target detection is the problem of searching for a known signature within a hyperspectral scene. Generally, the target spectral signature has been fully characterized prior to the application of a target detection algorithm. However, obtaining an effective target signature is often a challenging problem. Spectral signatures collected from handheld sensors on the ground over a target of interest or obtained from a spectral library and collected using laboratory instruments do not account for variation due to environmental or atmospheric conditions. Methods to modify spectra for a given environment are often prohibitively computationally complex and require knowledge of a large number of environmental

and atmospheric parameters [1]. Target signatures collected in advance or manually extracted from a scene may also not be the most discriminating signatures for a material given some background. In this paper, two algorithms for target characterization (i.e., estimation of target spectral signatures) from training imagery are presented. The goal of these algorithms is to estimate the target spectral signature from mixed training data that are effective for a follow-on target detection task. Since these algorithms extract the spectral signature from training imagery, then background materials, environmental and atmospheric conditions, and other such variables are addressed during target characterization.

This task of target characterization from training data can be viewed as a supervised learning problem. The applicability and effectiveness of a supervised learning algorithm for a particular problem hinge on the availability and accuracy of training data in the format assumed by the learning algorithm. A huge number of approaches have been developed in the machine learning literature to accommodate problems with varying levels of training data accuracy and availability. For the problem of discriminating between two classes (e.g., target versus nontarget), an enormous number of approaches have been developed including support vector machines [2], relevance vector machines [3], K-nearest neighbor approaches [4]–[7], naive Bayes classifiers [8], likelihood ratio tests, linear discriminant approaches [9], and others. These approaches generally require each input training point $\mathbf{x}_n$ to be paired with the desired classification output y_n and assume that each data point is a *pure* example of one class. In many of these approaches, training data samples are selected as prototypes for each class of interest. However, many problems are plagued with *mixed* data in which each data point is a function of many classes and *pure* unmixed training data cannot be easily obtained. Thus, the prototypes estimated by many of these methods would be ineffective for hyperspectral target detection. In this paper, two algorithms for target and nontarget signature estimation given *mixed* training data with *unspecific* labels are presented. These algorithms are then applied to mixed hyperspectral data for target and background endmember estimation and used for subpixel target detection.

A. Hyperspectral Unmixing

Hyperspectral unmixing is composed of two major tasks: *endmember estimation* and *abundance estimation*. To perform spectral unmixing, a mixing model must be assumed. In the literature, both linear and nonlinear models have been developed and found to be accurate in various physical scenarios [10].

The linear mixing model assumes that each pixel is a convex combination of the endmembers

$$\mathbf{x}_n = \sum_{k=1}^{K} p_{nk}\mathbf{e}_k + \boldsymbol{\varepsilon}_n, \quad n = 1, \ldots, N \tag{1}$$

where N is the number of data points, K is the number of endmembers (or materials), $\mathbf{x}_n$ is the spectral signature of the nth pixel, $\boldsymbol{\varepsilon}_n$ is an error/noise term, $\mathbf{e}_k$ is the spectral signature of the kth endmember, and p_{nk} is the abundance of the kth endmember in the nth pixel. The abundances in this model are constrained to $\sum_{k=1}^{K} p_{nk} = 1$ and $p_{nk} \geq 0, \forall n, k$. Generally, only the N data points are known, and the remaining parameters, including the number of endmembers, endmember signatures, and all abundance values, need to be estimated. Solving for these unknown parameters is an ill-posed inverse problem.

To solve these ill-posed problems, many unsupervised methods have been developed to estimate solutions by enforcing a number of broad assumptions about hyperspectral imagery [10]–[15]. These broad assumptions include constraining the solution space by requiring endmember signatures to be found within the input data [16]–[21], minimum volume constraints [22]–[25], enforcing sparsity assumptions [26]–[31], or incorporating spatial information to enforce smoothly varying abundance values across neighboring pixels [32]–[36]. Previous hyperspectral unmixing methods that estimate both endmembers and proportions are unsupervised algorithms. However, if prior information about the approximate location of particular materials of interest is available, then this would provide scene-specific information. This scene-specific information is likely more accurate and applicable to a particular scene than the broad, and potentially inaccurate, assumptions that are often used in unsupervised spectral unmixing algorithms. However, leveraging this sort of scene-specific information requires the development of supervised unmixing methods. The *c*FUMI and *e*FUMI algorithms presented here are supervised unmixing algorithms that can leverage scene-specific information.

B. Hyperspectral Target Detection

The wealth of spectral information in hyperspectral imagery provides for the ability to perform subpixel target detection. Subpixel detection techniques include *statistical* methods, *subspace* approaches, detectors based on the linear mixing model, and others [1], [37]–[40]. In statistical methods, the target and background signals are modeled as random variables distributed according to some respective underlying probability distribution. The detection problem can then be posed as a binary hypothesis test with two competing hypotheses [target absent ($\mathbf{H}_0$) or target present ($\mathbf{H}_1$)], and a detector can be designed using the generalized likelihood ratio test (GLRT) approach [41]. In [42]–[44], the adaptive coherence/cosine estimator (ACE) is derived following the GLRT approach. The structured-background ACE detector (which is applied in the experiments presented in this paper) is defined in

$$\mathbf{H}_0 : \mathbf{x} = \mathbf{n}, \text{ target absent}$$

$$\mathbf{H}_1 : \mathbf{x} = \sigma_s\mathbf{s} + \sigma_b\mathbf{n}, \text{ target present} \tag{2}$$

where $\mathbf{n}$ represents the Gaussian random noise distributed as $\boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_b)$ where $\boldsymbol{\Sigma}_b$ is the covariance of the background noise obtained from the training data, $\boldsymbol{s}$ is the known target signature which is scaled by an unknown target abundance σ_s, and σ_b is an unknown scalar value associated with the target data. The resulting GLRT for (2) is shown in

$$D_{\eta_{ACE}}(\mathbf{x}) = \frac{(\mathbf{s}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}_b^{-1}(\mathbf{x}-\boldsymbol{\mu})}{\sqrt{(\mathbf{s}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}_b^{-1}(\mathbf{s}-\boldsymbol{\mu})}\sqrt{(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}_b^{-1}(\mathbf{x}-\boldsymbol{\mu})}}$$

$$= \cos(\theta) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ACE} \tag{3}$$

where η is a threshold parameter and θ is the angle between the test point and target signature.

In addition to ACE, a large number of hyperspectral target detection methods have been developed in the literature [45]. These target detection algorithms rely on an accurate spectral signature for the target material to be known in advance. However, in a number of scenarios, target signatures may not be known in advance. For instance, an analyst or agent on the ground may identify a target of interest with approximate GPS coordinates and would like to determine if that material can be found elsewhere in the scene or in another imagery (i.e., laboratory spectra for the target are not available but an approximate location of a *possibly subpixel* example has been identified in an image). Also, spectra found in a spectral library for a target material may not adequately apply to a particular scene collected under different environmental and atmospheric conditions. In this paper, FUMI algorithms are proposed as a method to estimate target signatures from training data given only inaccurate and unspecific labels.

C. MIL

Multiple-instance learning (MIL) is a variation on supervised learning for problems with incomplete labels for the training data. Instead of receiving a set of training data with accurate labels, the learner receives a set of labeled "bags." A bag is defined to be a multiset of data points. In the case of binary labels, a bag is defined as positive if at least one of the data points in the bag is an instance of the target class. The accurate number of target instances in each positive bag is unknown. Negatively labeled bags are entirely nontarget data.

Thus, in MIL, there are no labels on the individual data points. MIL methods are effective for developing classifiers for cases where accurate instance-level labeled training data are unavailable. MIL was originally motivated and proposed from the goal to predict the drug molecule activity [46]. Since then, many MIL methods have been proposed and developed, such as learning axis-parallel concepts [46] which find an axis-parallel rectangle constructed by the conjunction of the features to approximate the target concept, diverse density [47] which tries to find a concept prototype that lies close to at least one instance in each positive bag while being far away from all instances that belong to negatively labeled bags, expectation–maximization (EM) version of diverse density [48] which identifies the target concept using an EM approach which assumes that the bag label

$$F = \frac{(1-\mu)}{2} \sum_{i=1}^{N} w_{l(\mathbf{x}_i)} \left\| \left(\mathbf{x}_i - l(\mathbf{x}_i)p_{iT}\mathbf{e}_T - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k \right) \right\|_2^2 + \frac{u}{2} \sum_{k=1}^{M} \|(\mathbf{e}_k - \boldsymbol{\mu}_0)\|_2^2 + \frac{u}{2} \|(\mathbf{e}_T - \boldsymbol{\mu}_0)\|_2^2 + \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N} p_{ik} \quad (7)$$

is a set of hidden variables determined by the most positive instance, and others [49]–[51]. Here, *concepts* refer to generalized class prototypes in the feature space.

The functions of multiple instances (FUMI) approaches [52]–[54] are a generalization of MIL. FUMI treats each data point as a function of positive or negative concepts. Given some functional form, FUMI learns the target and nontarget concepts and the function parameters defining the relationship between each data point and the concepts. In the case of hyperspectral image unmixing, a FUMI concept is equivalent to the spectral signature of an endmember. In this paper, two FUMI algorithms, convex FUMI (*c*FUMI) and extended *c*FUMI (*e*FUMI), are introduced. The *c*FUMI algorithm assumes the exact knowledge of target *locations* in a training image (i.e., *c*FUMI assumes that you can exactly identify several mixed pixels that contain some portion of the target signature). *e*FUMI, on the other hand, needs only an approximate knowledge of target locations in training data (or, in other words, only sets of signatures where at least one spectrum in the set contains some subpixel proportion of the target). This paper extends previous work [52]–[54] on *c*FUMI and *e*FUMI by providing a unified and complete derivation of the *c*FUMI and *e*FUMI algorithms, effective initialization strategies, improved optimization strategies for *c*FUMI, and a normalization strategy for input data. Furthermore, this paper provides significant experimental results on simulated and real measured hyperspectral data to illustrate the effectiveness and range of behaviors of this approach beyond what is published in previous preliminary conference papers.

II. FUMI

A. cFUMI

*c*FUMI treats each data point as a function of positive and/or negative concepts. Given some functional form, *c*FUMI learns the target and nontarget concepts and the function parameters defining the relationship between each data point and the concepts. Specifically, suppose that there is a given data set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ where each data point is some unknown function of concepts, $\mathbf{x}_i = f(\mathbf{E}_i, \mathbf{p}_i)$, where $\mathbf{p}_i$ denotes the set of functional parameters for $\mathbf{x}_i$ and $\mathbf{E}_i$ is the "bag" of concepts that contribute in a nonnegligible way to the data point $\mathbf{x}_i$. Each training point $\mathbf{x}_i$ is given a binary label $l(\mathbf{x}_i)$ where $l(\mathbf{x}_i) = 1$ if $\mathbf{e}_T \in \mathbf{E}_i$ and $l(\mathbf{x}_i) = 0$ if $\mathbf{e}_T \notin \mathbf{E}_i$, where $\mathbf{e}_T$ is the target concept. Using these binary labels, *c*FUMI estimates the functional parameters $\mathbf{p}_i$ for each data point and the target and nontarget concepts for the data set $\mathbf{E} = \cup_i \mathbf{E}_i$, where $\mathbf{E}_i$ is a subset of $\mathbf{E}$ for data point $\mathbf{x}_i$, and $\mathbf{E}_i$ and $\mathbf{E}_j$ are not mutually exclusive.

In this paper, the specific functional form considered is that each data point is assumed to be a convex combination of target and nontarget concepts, as shown in

$$\mathbf{x}_i = p_{iT}\mathbf{e}_T + \sum_{k=1}^{M} p_{ik}\mathbf{e}_k \quad (4)$$

where $\mathbf{x}_i$ is a data point, $\mathbf{e}_T$ is the target signature, $\mathbf{e}_k$ is a nontarget signature for $k = 1, \ldots, M$, and p_{ik} is the proportion value of the kth signature in data point i. The proportions are constrained to sum-to-one and be nonnegative

$$p_{iT} + \sum_{k=1}^{M} p_{ik} = 1, \quad p_{iT} \geq 0, \quad p_{ik} \geq 0. \quad (5)$$

When applied to hyperspectral data, the target and nontarget concepts are equivalent to the spectral signatures of target and nontarget endmembers. If $l(\mathbf{x}_i) = 1$, then $\mathbf{x}_i = p_{iT}\mathbf{e}_T + \sum_{k=1}^{M} p_{ik}\mathbf{e}_k$ with $p_{iT} > 0$. If $l(\mathbf{x}_i) = 0$, then $\mathbf{x}_i = \sum_{k=1}^{M} p_{ik}\mathbf{e}_k$. The exact proportion values for the training data are not known; the proportion of each pixel associated with a particular target is often unknown as it depends on the relationship between spatial resolution and the (unknown) field of view associated with each pixel. Hence, the binary labels are considered *unspecific* as they do not specify the degree to which a positively labeled data point is representative of the target concept.

*c*FUMI estimates the desired parameter values (i.e., $\mathbf{e}_T$, $\mathbf{E}$, $\mathbf{P}$, and M) by minimizing the objective function in (7), shown at the top of the page. There are four terms in this objective function. The first term computes the squared error between the input data and its estimate found using the current target and nontarget signatures and proportions where u is a parameter constant controlling the relative importance of the first, second, and third terms. The value for $w_{l(\mathbf{x}_i)}$ is shown in

$$w_{l(\mathbf{x}_i)} = \begin{cases} 1 & \text{if } l(\mathbf{x}_i) = 0 \\ \frac{\alpha N_n}{N_t} & \text{if } l(\mathbf{x}_i) = 1 \end{cases} \quad (6)$$

where N_n is the number of negatively labeled samples and N_t is the number of positively labeled samples. Therefore, if the parameter α is set to 1, then the weight on the target points is scaled such that the collection of target points has the same influence on the first term as the collection of nontarget training points. Furthermore, α can be set to larger than 1 to emphasize the importance of target training data over background data.

The second and third terms of the objective encourage target and nontarget signatures that provide a tight fit around the data by minimizing the squared difference between each signature and the global data mean $\boldsymbol{\mu}_0$. These terms were motivated by the volume-related term in the SPICE [25] algorithm. The fourth term is a sparsity-promoting term used to determine M, the number of nontarget signatures needed to describe the input data where $\gamma_k = \Gamma/(\sum_{n=1}^{N} p_{nk}^{(t-1)})$, where Γ is a parameter constant that controls the degree that sparsity is promoted. Higher values of Γ generally result is a smaller estimate M value. The $p_{nk}^{(t-1)}$ values are the proportion values estimated in the previous iteration of the algorithm. Thus, as the proportions for a particular endmember decrease, the weight of its associated sparsity-promoting term increases. This approach for estimating the number of background endmembers follows the approach presented by the SPICE algorithm [25].

$$F = \frac{(1-u)}{2}\sum_{i=1}^{N} w_i \left\| \left(\mathbf{x}_i - z_i p_{iT}\mathbf{e}_T - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k \right) \right\|_2^2 + \frac{u}{2}\sum_{k=1}^{M} \|(\mathbf{e}_k - \boldsymbol{\mu}_0)\|_2^2 + \frac{u}{2}\|(\mathbf{e}_T - \boldsymbol{\mu}_0)\|_2^2 + \sum_{k=1}^{M}\gamma_k \sum_{i=1}^{N} p_{ik} \tag{8}$$

$$E[F] = \sum_{z_i \in \{0,1\}} \left[\frac{(1-u)}{2}\sum_{i=1}^{N} w_i P\left(z_i|\mathbf{x}_i,\boldsymbol{\theta}^{(t-1)}\right)\left\|\mathbf{x}_i - z_i p_{iT}\mathbf{e}_T - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k\right\|_2^2 \right] + \frac{u}{2}\sum_{k=1}^{M}\|\mathbf{e}_k - \boldsymbol{\mu}_0\|_2^2 + \frac{u}{2}\|\mathbf{e}_T - \boldsymbol{\mu}_0\|_2^2 + \sum_{k=1}^{M}\gamma_k \sum_{i=1}^{N} p_{ik} \tag{9}$$

cFUMI estimates the target and nontarget signatures, proportions, and number of nontarget signatures using alternating optimization to minimize the objective function in (7). When updating one parameter, all other parameters are held constant. The objective function is iteratively minimized until some stopping criterion is reached such as convergence or a maximum number of iterations. The method is summarized in Algorithm 1.

Algorithm 1 cFUMI algorithm

1: Initialize $\theta^0 = \{\mathbf{e}_T, \mathbf{E}, \mathbf{P}\}$, $t = 1$
2: **repeat**
3:        $t \leftarrow t + 1$
4:        Update $\mathbf{e}_T$ and $\mathbf{E}$ by minimizing (7) wrt. $\mathbf{e}_T$, $\mathbf{E}$
5:        Update $\mathbf{P}$ by minimizing (7) wrt. $\mathbf{P}$ s.t. (5)
6:        Prune each $\mathbf{e}_k, k = 1, \ldots, M$ if $\max_n(p_{nk}) \leq \tau$
          where τ is a fixed threshold (e.g., $\tau = 10^{-6}$)
7: **until** Convergence
8: **return** $\mathbf{e}_T, \mathbf{E}, \mathbf{P}$
   ∗The update equations used in lines (4)–(5) are derived in Appendix A.

After learning target and nontarget signatures, target detection on test data can be carried out. In the results shown in this paper, target detection is carried out using the ACE detector using the estimated target signature $\mathbf{e}_T$.

B. eFUMI

cFUMI was able to address the unspecificity associated with unknown target and nontarget proportion values. However, cFUMI requires data point-specific binary labels and thus cannot address location uncertainty. In other words, the specific proportion values associated with each training data point are unneeded; however, each training point requires a binary target/nontarget label for application of the cFUMI algorithm. However, in many applications even unspecific binary labels for each data point are difficult or impossible to obtain, whereas bag-level or approximate spatial/ground information may be available (with respect to UTM or pixel coordinates). For example, UTM coordinates of several targets of interest may have been collected with an accuracy ranging across several pixels. Thus, accurate pixel-level ground truth is unknown; only approximate locations are available. The proposed eFUMI algorithm addresses both unspecificity due to unknown proportion values and location uncertainty in training data.

eFUMI extends cFUMI to be able to learn target and nontarget concepts without the need of instance-level labels. Specifically, the goal of eFUMI is to estimate a target concept $\mathbf{e}_T$, nontarget concepts $\mathbf{e}_k$, $\forall k = 1, \ldots, M$, the number of needed nontarget concepts M, and the function parameters $\mathbf{p}_i$, which define the

relationship between each data point $\mathbf{x}_i$ and concepts. These are estimated given a set of input training data, $\{\mathbf{x}_i\}_{i=1}^{N} \in \mathbb{R}^D$, which have been partitioned into K "bags," $\mathbf{B} = \{B_1, \ldots, B_K\}$, with associated bag-level labels, $L = \{L_1, \ldots, L_K\}$. Bag-level labels are *unspecific* since, if for bag B_j with $L_j = 1$ (thus, B_j is a positive bag), this indicates that there is at least one data point in B_j with a positive p_{iT} indicating some presence of target, as in (10), where $\boldsymbol{\varepsilon}_i$ is an error term

$$\text{if } L_j = 1, \exists \mathbf{x}_i \in B_j \text{ s.t. } \mathbf{x}_i = p_{iT}\mathbf{e}_T + \sum_{k=1}^{M} p_{ik}\mathbf{e}_k + \boldsymbol{\varepsilon}_i, p_{iT} > 0 \tag{10}$$

However, the exact number of data points in a positive bag with a target contribution (i.e., $p_{iT} > 0$), is unknown; also, the target proportions are unknown. Furthermore, if B_j is a negative bag (i.e., $L_j = 0$), then this indicates that none of the data in B_j contains any target, as in

$$\text{if } L_j = 0, \forall \mathbf{x}_i \in B_j, \mathbf{x}_i = \sum_{k=1}^{M} p_{ik}\mathbf{e}_k + \boldsymbol{\varepsilon}_i. \tag{11}$$

Given the training data of this form, eFUMI addresses this problem using an EM approach in which the instance-level labels are the hidden latent variables in the EM model.

The assumed *complete* data log-likelihood is proportional to (8), shown at the top of the page, where z_i is the unknown instance-level labels. The eFUMI objective function mimics the cFUMI objective function; however, instead of having the data point-specific labels $l(\mathbf{x}_i)$, these are replaced with hidden latent variables z_i. To address the fact that the z_i values are unknown, the expected values of the log likelihood with respect to z_i are taken in (9), shown at the top of the page. In (9), $\boldsymbol{\theta}^t$ is the set of parameters estimated at iteration t, and $P(z_i|\mathbf{x}_i, \boldsymbol{\theta}^{(t-1)})$ is the probability of individual points containing any proportion of target or not. $P(z_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(t-1)})$ is determined given the parameter set estimated in the previous iteration and the constraints of the bag-level labels L_j, as shown in

$$P\left(z_i|\mathbf{x}_i, \boldsymbol{\theta}^{(t-1)}\right)$$
$$= \begin{cases} e^{-\beta\left\|\mathbf{x}_i - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k\right\|_2^2} & \text{if } z_i = 0, L_i = 1 \\ 1 - e^{-\beta\left\|\mathbf{x}_i - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k\right\|_2^2} & \text{if } z_i = 1, L_i = 1 \\ 0 & \text{if } z_i = 1, L_i = 0 \\ 1 & \text{if } z_i = 0, L_i = 0 \end{cases} \tag{12}$$

where β is a scaling parameter and $r_b = \|\mathbf{x}_i - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k\|_2^2$ is the approximation residual between $\mathbf{x}_i$ and its representation

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JIAO AND ZARE: FUNCTIONS OF MULTIPLE INSTANCES FOR LEARNING TARGET SIGNATURES

5

using only background endmembers. The definition of $P(z_i|\mathbf{x}_i, \boldsymbol{\theta}^{(t-1)})$ in (12) indicates that, if a point $\mathbf{x}_i$ is a non-target point, it should be fully represented by the background endmembers with very small residual r_b; thus, $P(z_i = 0|\mathbf{x}_i, \boldsymbol{\theta}^{(t-1)}) = e^{-\beta r_b} \to 1$. Otherwise, if $\mathbf{x}_i$ is a target point, it may not be well represented by only the background endmembers, so the residual r_b may be large and $P(z_i = 1|\mathbf{x}_i, \boldsymbol{\theta}^{(t-1)}) = 1 - e^{-\beta r_b} \to 1$. Note that z_i is unknown only for the positive bags; in the negative bags, z_i is fixed to 0. This constitutes the *E-step* of the EM algorithm.

The *M-step* is performed by optimizing (9) for each of the desired parameters. The method is summarized in Algorithm 2.

Algorithm 2 *e*FUMI EM algorithm

1: Initialize $\boldsymbol{\theta}^0 = \{\mathbf{e}_T, \mathbf{E}, \mathbf{P}\}$, $t = 1$
2: **repeat**
3:        *E-step*: Compute $P(z_i|\mathbf{x}_i, \boldsymbol{\theta}^{(t-1)})$ given $\boldsymbol{\theta}^{t-1}$
4:        *M-step*:
5:          Update $\mathbf{e}_T$ and $\mathbf{E}$ by maximizing (9) wrt. $\mathbf{e}_T$, $\mathbf{E}$
6:          Update $\mathbf{P}$ by maximizing (9) wrt. $\mathbf{P}$ s.t. (5)
7:          Prune each $\mathbf{e}_k, k = 1, \ldots, M$ if $\max_n(p_{nk}) \le \tau$
         where τ is a fixed threshold (e.g., $\tau = 10^{-6}$)
8:          $t \leftarrow t + 1$
9: **until** Convergence
10: return $\mathbf{e}_T, \mathbf{E}, \mathbf{P}$
     ∗The update equations used in lines (5)–(6) are derived in Appendix B.

III. INITIALIZATION AND PARAMETER SETTINGS

*c*FUMI and *e*FUMI require several parameters to be set. For all experimental results shown, initialization for the algorithm and parameters is determined using the following.

A. Initialization

Nontarget signatures are initialized by using the VCA algorithm [55] on all data in the negatively labeled bags. Then, using these initial nontarget signatures, the data in the positively labeled bags are unmixed. The data point with the largest reconstruction error using the initial nontarget signatures is set as the initial target signature $\mathbf{e}_T$. All proportion values are initialized to $1/(M+1)$ for all data points in positive bags and to $(1/M)$ for all points in negative bags (since the proportion on the target endmember is fixed to 0).

B. Parameter Settings

There are a few parameters that must be set in *c*FUMI and *e*FUMI. In both algorithms, the u parameter trades off between the residual error term and the volume-related terms in the objective function. In all of our results, $u \in [0.01, 0.1]$. Valid values of u are in the set (0,1). Values that tend to 1 are appropriate for data with large noise levels. In contrast, values of u that tend toward 0 are appropriate for data with low noise levels and/or simulated data.

The initial value of M does not play a large role in the accuracy of the algorithm, provided that it is initialized to a value
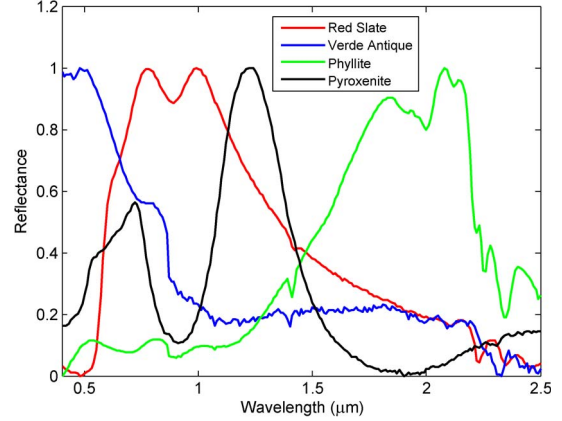

Fig. 1. Signatures from ASTER library used to generate simulated data.

TABLE I
ERROR AND STANDARD DEVIATION (OVER TEN RUNS),
RANDOM DATA, NO NORMALIZATION

Algorithms	NMSE($\times 10^{-5}$)	MSAD($\times 10^{-5}$)
*c*FUMI	2.13 ± 0.15	1.95 ± 0.12
*e*FUMI	4.05 ± 0.26	3.97 ± 0.25

larger than the needed number of nontarget signatures. However, initializing with a very large M value will require a larger number of iterations in which to prune the unnecessary nontarget signatures, resulting in a longer running time. The sparsity-promoting parameter Γ controls the degree of sparsity and the resulting number of nontarget signatures. A larger Γ value tends to result in a smaller number of nontarget signatures.

The scaling parameter β used in the calculation of $P(z_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(t-1)}) = 1 - e^{-\beta\|\mathbf{x}_i - \sum_{k=1}^M p_{ik}\mathbf{e}_k\|_2^2}$ aids in separating target and nontarget points in positively labeled bags. As can be seen in the definition of $P(z_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(t-1)})$, setting β is related to the magnitude of the input data, the number of background endmembers, and the spectral similarity between the target and background endmembers. For example, larger data magnitude corresponds to a larger reconstruction error in general, and thus, a smaller β value is needed. Similarly, more background endmembers result in general with smaller reconstruction error, and thus, a larger β value is needed. In our experience, normalizing the input data such that each data point has unit norm greatly helps in the setting of this parameter. In our experiments with normalized data, $\beta \in [30, 50]$ has been found to work well. In all experimental results in this paper, the parameters were set manually. Each experiment lists the parameters used to generate the results shown.

IV. EXPERIMENTAL RESULTS

The proposed *c*FUMI and *e*FUMI were tested both on synthetic data and real hyperspectral data.[1]

A. FUMI on Synthetic Data

Both *c*FUMI and *e*FUMI were run on simulated data generated from four spectra selected from the ASTER spectral library

---

[1]Data and code used in this paper can be found at: http://engineers.missouri.edu/zarea/tigersense/code/.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

TABLE II
ERROR AND STANDARD DEVIATION (OVER TEN RUNS), HIGHLY MIXED DATA, NO NORMALIZATION

p_{t_mean}	cFUMI		eFUMI	
	NMSE ($\times 10^{-2}$)	MSAD ($\times 10^{-2}$)	NMSE($\times 10^{-2}$)	MSAD($\times 10^{-2}$)
0.3	$0.11 \pm 1.59 \times 10^{-2}$	$0.10 \pm 1.52 \times 10^{-2}$	$0.18 \pm 2.88 \times 10^{-2}$	$0.17 \pm 2.65 \times 10^{-2}$
0.5	$4.09 \times 10^{-2} \pm 5.94 \times 10^{-3}$	$3.74 \times 10^{-2} \pm 5.61 \times 10^{-3}$	$6.28 \times 10^{-2} \pm 6.39 \times 10^{-3}$	$6.02 \times 10^{-2} \pm 5.95 \times 10^{-3}$
0.7	$1.08 \times 10^{-2} \pm 1.11 \times 10^{-3}$	$9.79 \times 10^{-3} \pm 9.65 \times 10^{-4}$	$1.57 \times 10^{-2} \pm 1.42 \times 10^{-3}$	$1.49 \times 10^{-2} \pm 1.33 \times 10^{-3}$

TABLE III
ERROR AND STANDARD DEVIATION (OVER TEN RUNS), NOISY DATA, NO NORMALIZATION

SNR	cFUMI		eFUMI	
	NMSE($\times 10^{-2}$)	MSAD($\times 10^{-2}$)	NMSE($\times 10^{-2}$)	MSAD($\times 10^{-2}$)
10	7.06 ± 0.77	6.89 ± 0.81	8.35 ± 0.85	8.13 ± 0.82
20	2.24 ± 0.29	2.13 ± 0.28	2.88 ± 0.31	2.75 ± 0.32
30	$0.68 \pm 4.29 \times 10^{-2}$	$0.63 \pm 4.36 \times 10^{-2}$	$0.95 \pm 6.55 \times 10^{-2}$	$0.86 \pm 6.39 \times 10^{-2}$
40	$0.23 \pm 1.58 \times 10^{-2}$	$0.22 \pm 1.55 \times 10^{-2}$	$0.34 \pm 2.26 \times 10^{-2}$	$0.35 \pm 2.41 \times 10^{-2}$

[56]. Specifically, the Red Slate, Verde Antique, Phyllite, and Pyroxenite spectra from the rock class with 211 bands and wavelengths ranging from 0.4 to 2.5 μm (as shown in Fig. 1) were used as endmembers to generate hyperspectral data. Red Slate was labeled as the target endmember. Five bags were generated with 1000 points in each bag. The first two bags were positive bags, and each had 250 points with nonzero target proportion values. The constituent background endmembers and the number of background endmembers for each data point were drawn randomly. The corresponding background proportion values were generated by drawing from a Dirichlet distribution. The α parameter in the Dirichlet distribution was varied to generate a range of simulated data sets from those containing some pure target pixels to highly mixed data sets. For a more precise description of how the simulated data was generated, a pseudocode describing the generation method is shown in Algorithm 3 and 4 in Appendix C.

Three synthetic data sets were generated: 1) random data; 2) highly mixed data; and 3) noisy data. The random data set contains mixed data with some pure target points. This data set was generated using Algorithm 3 with parameters $n_{tar} = 250$, $N_b = 0$, $p_{t\_mean} = (1/m)$ where m is the number of background endmembers for the point under consideration, and $\sigma = 1$. The highly mixed data set contains all highly mixed data generated using Algorithm 3 with parameters $n_{tar} = 250$ and $N_b = 1$. p_{t_mean} was varied to be 0.3, 0.5, and 0.7. The noisy data set was generated using the same process as the random data set but with added Gaussian white noise. The signal-to-noise ratio of the Gaussian noise was varied from 10 to 40 dB.

In our experiments, algorithms were considered to have converged if the change in the objective function value is below a predefined threshold of $\tau = 10^{-6}$. Accurate data point-specific labels are used for running cFUMI. For eFUMI, only bag-level labels were used. In the results in this section, the normalized mean square error (NMSE) and mean spectral angle distance (MSAD) between the estimated Red Slate spectrum and the true Red Slate spectrum were calculated to evaluate the results, where $\text{NMSE} = \mathbf{E}[\|\mathbf{e}_{tru} - \mathbf{e}_{est}\|/\|\mathbf{e}_{tru}\|]$, $\text{MSAD} = \mathbf{E}[\cos^{-1}((\mathbf{e}_{tru}^T \cdot \mathbf{e}_{est})/(\|\mathbf{e}_{tru}\| \cdot \|\mathbf{e}_{est}\|))]$, and $\mathbf{e}_{tru}$ and $\mathbf{e}_{est}$ are the truth and estimated target spectra, respectively.
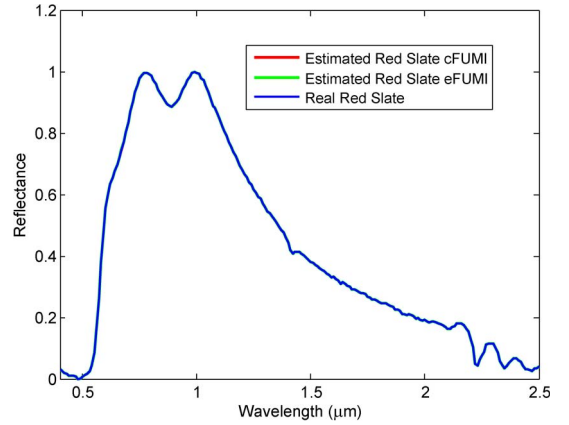

Fig. 2.   Estimated target spectra from random data without normalization.

1) cFUMI and eFUMI Synthetic Data Results: In this section, the parameter settings used for cFUMI on all three synthetic data sets were $u = 0.05$, $M = 4$, and $\Gamma = 10$, the parameter settings for eFUMI on random and noisy data were $u = 0.05$, $M = 4, \Gamma = 10$, and $\beta = 20$, and those for highly mixed data were $u = 0.05$, $M = 4, \Gamma = 10$, and $\beta = 45$. The results are shown in Tables I–III, where p_{t_mean} and SNR denote the mean target (Red Slate) proportion value of the highly mixed target points and signal-to-noise ratio of synthetic data with Gaussian white noise, respectively. Figs. 2 and 3 are example plots of the estimated target (Red Slate) signature. Experimental results show that both cFUMI and eFUMI algorithms are capable of detecting highly mixed subpixel target spectrum and are robust to noise.

2) Normalized Data and Constrained Endmembers: cFUMI and eFUMI performance was also examined after normalization of the input data. Specifically, the algorithms were tested on the highly mixed synthetic data in two scenarios: 1) cFUMI and eFUMI were run on input data that were normalized to have unit norm, and 2) cFUMI and eFUMI were run on normalized input data, and a constraint was placed on the estimated endmembers to also have unit norm. This constraint was enforced by renormalizing the estimated endmembers after each iteration. In this section, the parameter settings used for cFUMI were

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JIAO AND ZARE: FUNCTIONS OF MULTIPLE INSTANCES FOR LEARNING TARGET SIGNATURES
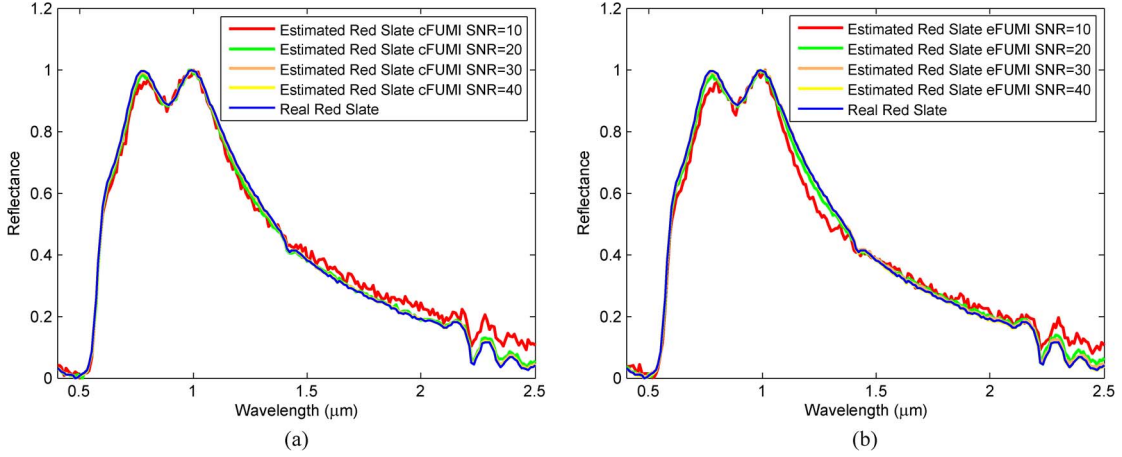
7

Fig. 3.   Estimated target spectrum, noisy data, no normalization. (a) cFUMI. (b) eFUMI.

TABLE IV
ERROR AND STANDARD DEVIATION (OVER TEN RUNS), HIGHLY MIXED DATA, NORMALIZED, NO CONSTRAINT

Pt_mean	cFUMI		eFUMI	
	NMSE($\times 10^{-2}$)	MSAD($\times 10^{-2}$)	NMSE($\times 10^{-2}$)	MSAD($\times 10^{-2}$)
0.3	4.05 ± 0.44	3.97 ± 0.43	4.66 ± 0.55	4.57 ± 0.56
0.5	3.18 ± 0.30	3.02 ± 0.28	3.51 ± 0.35	3.46 ± 0.32
0.7	2.98 ± 0.19	2.98 ± 0.22	3.21 ± 0.26	3.15 ± 0.25

TABLE V
ERROR AND STANDARD DEVIATION (OVER TEN RUNS), HIGHLY MIXED DATA, NORMALIZED WITH CONSTRAINT

Pt_mean	cFUMI		eFUMI	
	NMSE ($\times 10^{-2}$)	MSAD ($\times 10^{-2}$)	NMSE ($\times 10^{-2}$)	MSAD ($\times 10^{-2}$)
0.3	3.46 ± 0.23	3.44 ± 0.23	4.05 ± 0.31	3.98 ± 0.28
0.5	2.12 ± 0.13	2.03 ± 0.11	3.18 ± 0.20	3.14 ± 0.19
0.7	$0.92 \pm 6.83 \times 10^{-2}$	$0.89 \pm 7.13 \times 10^{-2}$	2.15 ± 0.15	2.02 ± 0.14

$u = 0.05, M = 4$, and $\Gamma = 10$, and the parameter settings for eFUMI were $u = 0.05, M = 4, \Gamma = 10$, and $\beta = 40$.

The experimental results are summarized in Tables IV and V and are compared with the results in Section IV-A1. These results show that normalization of synthetic data decreases performance in comparison to the unnormalized results from Section IV-A1. This is likely due to the fact that additional discriminating information between the endmembers based on the magnitude is lost after normalization. However, after adding the constraint on estimated endmembers, performance improves greatly in comparison to the normalized result with no constraint enforced. This is due to the fact that adding the constraint reduces the search space.

B.   eFUMI on Real Hyperspectral Data

eFUMI was also run on several real hyperspectral data sets, including the data collected over Pavia, Italy [57], a portion of the RIT SHARE 2012 data set [58], and the MUUFL Gulfport data set [59]. In the following experiments, all the results shown are on unnormalized data. In these experiments, algorithms were considered to have converged if the change in the objective function value is below a threshold of $\tau = 10^{-6}$ or if 500 iterations was reached.

1) eFUMI on Pavia University Data: A subset of the ROSIS Pavia University data set collected over an urban area of Pavia in northern Italy by the ROSIS spectrometer on July 8, 2002 was used for the first real hyperspectral data experiment. The image contains 610×340 pixels with 103 bands corresponding to the 430–850-nm wavelength range at a 4-nm spectral sampling interval. The geometric resolution is 1.3 m. Holzwarth *et al.* provide more information on the sensor and preprocessing steps [57].

The RGB image of this scene and the regions labeled forming the positive and negative bags are shown in Fig. 4. Painted Metal Sheets were chosen as the target material in this experiment. The set of points in a rectangular area surrounding the painted metal sheets in the scene was grouped together to form a positive bag. All others (outside of the rectangular area) were grouped together as one negative bag.

The detected corresponding target proportion map is shown in Fig. 5. Fig. 6 plots the estimated target endmember signature and manually extracted spectra of painted metal sheets according to the ground truth of the Pavia University hyperspectral image. This shows that the detected target spectrum is a good fit of the target points. The parameter settings of eFUMI for this experiment are $u = 0.01, M = 4, \Gamma = 2$, and $\beta = 0.01$.

2) eFUMI on RIT SHARE 2012 Data: eFUMI was also run on the RIT SHARE 2012 data set collected near Rochester, NY

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

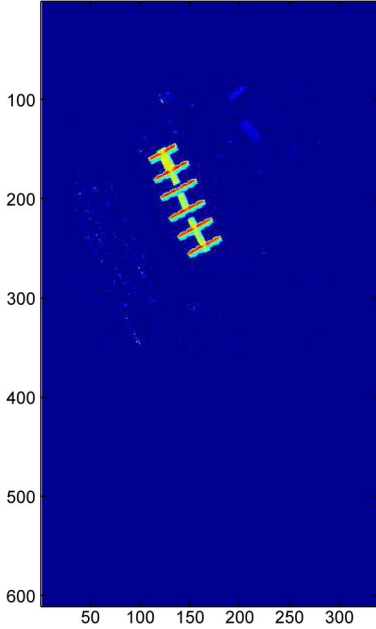Fig. 4.   Pavia University data set and labeled region for detection.

Fig. 5.   Proportion map of target endmember (painted metal sheets).

[58]. A subset of the AVON hyperspectral imagery from this collection containing 300×320 pixels with 360 bands corresponding to wavelengths from 400.3 to 2452.8 nm was used. The spatial resolution was 1 m. Two sets of these data were provided: 1) AVON AM collected in the morning of September 20, 2012 and 2) AVON PM collected in the afternoon of the same day. There were two target types with 12 examples each type: man-made blue and brown tarps. As shown in Fig. 7, a 5×5 rectangular region centered around each approximate target location was labeled as positive bags and the large background regions (shown as large yellow blocks in Fig. 7) were labeled as negative bags. The parameter settings for eFUMI for both AVON AM and AVON PM data were $u = 0.05, M = 6, \Gamma = 2$, and $\beta = 0.1$.
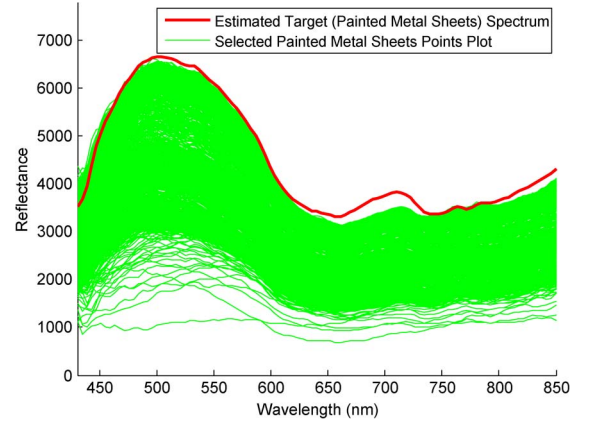

Fig. 6.   Detected target endmember and ground-truth target points.

Fig. 7.   RGB image of a subset of the RIT SHARE 2012 AVON AM data containing 24 scattered target points.

In order to investigate the quality of the estimated target spectra, ACE [42]–[44] was applied for target detection for both target types. Fig. 8(a) shows the detection map for the Blue targets. Fig. 8(b) displays a zoomed-in RGB image of the target area. As can be seen, the ACE detection results match the RGB image displaying the locations of the blue targets very well. This illustrates that eFUMI is capable of detecting a quality target signature from inaccurately labeled points.

For quantitative evaluation, receiver operating characteristic (ROC) curve analysis was conducted using cross-validation between AVON AM and AVON PM both for Blue and Brown targets. The ROC curves were generated using the Bullwinkle Scoring algorithm [60]. The detailed value of probability of detection (PD) versus false alarm rate (FAR) is shown in Table VI, where the FAR is in the unit of false alarms/m². ACE detection results obtained using ASD-measured ground spectra from the scene over the targets and results obtained using manually selected target spectra extracted over the targets from the scene were used as comparison. As shown in Table VI, eFUMI cross-validation results outperform the ASD-measured ground spectra as well as the manually extracted target spectra from the scene.

3) eFUMI on Mississippi Gulfport Data: Then, eFUMI was tested on the MUUFL Gulfport data set collected over the University of Southern Mississippi-Gulfpark Campus. This data set contains 325×337 pixels with 72 bands corresponding to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JIAO AND ZARE: FUNCTIONS OF MULTIPLE INSTANCES FOR LEARNING TARGET SIGNATURES 9
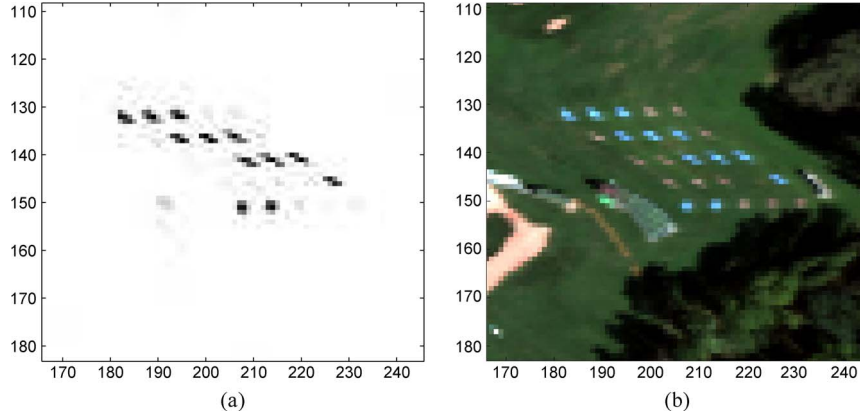

Fig. 8. ACE with *e*FUMI target spectrum detection results of blue targets on AVON AM. (a) ACE output map zoomed. (b) Zoomed target region of Fig. 7.

TABLE VI
PD VERSUS FAR BY ACE DETECTOR USING ESTIMATED AND GROUND-TRUTH TARGET SPECTRA, RIT SHARE2012 AVON DATA

FAR	1×10^{-5}	1×10^{-4}	5×10^{-3}	FAR	1×10^{-5}	1×10^{-4}	5×10^{-3}
*e*FUMI Blue on Avon PM	100%	100%	100%	*e*FUMI Brown on Avon PM	100%	100%	100%
ASD Blue on Avon PM	0%	0%	16.67%	ASD Brown on Avon PM	0%	0%	16.67%
Manually-extracted Blue on Avon PM	100%	100%	100%	Manually-extracted Brown on Avon PM	91.67%	100%	100%
*e*FUMI Blue on Avon AM	100%	100%	100%	*e*FUMI Brown on Avon AM	83.33%	100%	100%
ASD Blue on Avon AM	0%	8.33%	25%	ASD Brown on Avon AM	0%	0%	8.33%
Manually-extracted Blue on Avon AM	100%	100%	100%	Manually-extracted Brown on Avon AM	75%	100%	100%

wavelengths from 367.7 to 1043.4 nm at a 9.5-9.6-nm spectral sampling interval [59]. The spatial resolution is 1 m. Two sets of these data (Gulfport Campus Flight 1 and Gulfport Campus Flight 2) were selected as cross-validated training and testing data. Throughout the scene, there are 64 manmade targets. The spatial location of the targets is shown as scattered points over an RGB image of the scene in Fig. 9. Target types considered in this experiment were cloth panels of four different colors: Brown (15 examples), Dark Green (15 examples), Faux Vineyard Green (FVG) (12 examples), and Pea Green (15 examples). Some of the targets are in the open, and some are occluded. The targets also vary in size; for each target type, there are targets that are 0.25 m^2, 1 m^2, and 9 m^2 in area. A 5×5 rectangular region around each ground-truth point for each target was labeled as positive bags; this size was chosen since the accuracy of the GPS device used to record the ground-truth locations had a 5-m accuracy. The parameter settings of *e*FUMI for Flight 1 data are $u = 0.05, M = 7, \Gamma = 5$, and $\beta = 60$; for Flight 2 data, they are $u = 0.1, M = 7, \Gamma = 5$, and $\beta = 65$.

The estimated Brown, Dark Green, FVG, and Pea Green spectra by *e*FUMI are shown in Fig. 10(a)–(d), respectively. For comparison, the VCA algorithm was also run on the input data. From the endmembers found by VCA, the endmember that was most spectrally similar to the target was manually labeled as the target spectral signature. As can be seen from these figures, the VCA algorithm (since it does not have the benefit of the ability to use the inaccurate labels like *e*FUMI) is less accurate than *e*FUMI when compared to target spectral signatures manually extracted from the scene.

For quantitative evaluation, ROC curve analysis was conducted on the ACE detection results using the estimated target signatures for each target type. The analysis was done using cross-validation on two flights over the scene. The detection results are shown in Fig. 11 for Flight 1 (target signatures estimated using Flight 2) and in Fig. 12 for Flight 2 (target signatures
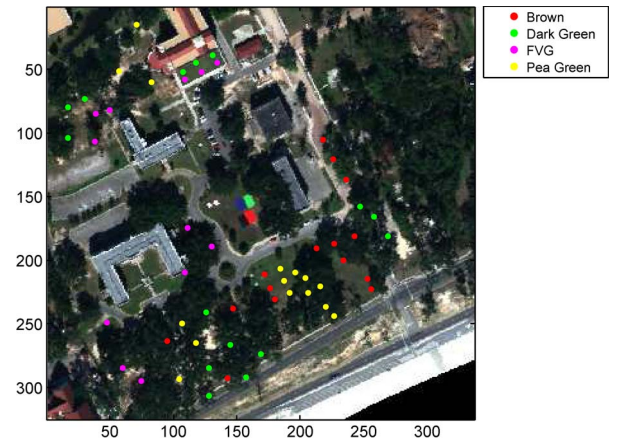

Fig. 9. MUUFL Gulfport data set RGB image and the 64 target locations.

estimated using Flight 1). As can be seen in Figs. 11 and 12, the scoring results of Brown, Dark Green, FVG, and Pea Green targets are very close to or even better than that of the manually extracted target spectra. Table VII shows the detailed PD value at three fixed FARs from Figs. 11 and 12.

To compare *e*FUMI with supervised classification methods, classification results using SVM [2], RVM [3], and mRVM [61] adopting Flight 1 data as the training set and Flight 2 data as testing are shown. For the SVM, the entire Gulfport Flight 1 is taken as training data in which all the 5×5 target regions are labeled as positive and the background is labeled as negative. For RVM and mRVM, due to training time and memory issues, during each individual run, all 5×5 target regions (positively labeled) together with a subset of 3000 pixels randomly sampled from the background of Gulfport Flight 1 (negatively labeled) are combined as the training data. For determining the
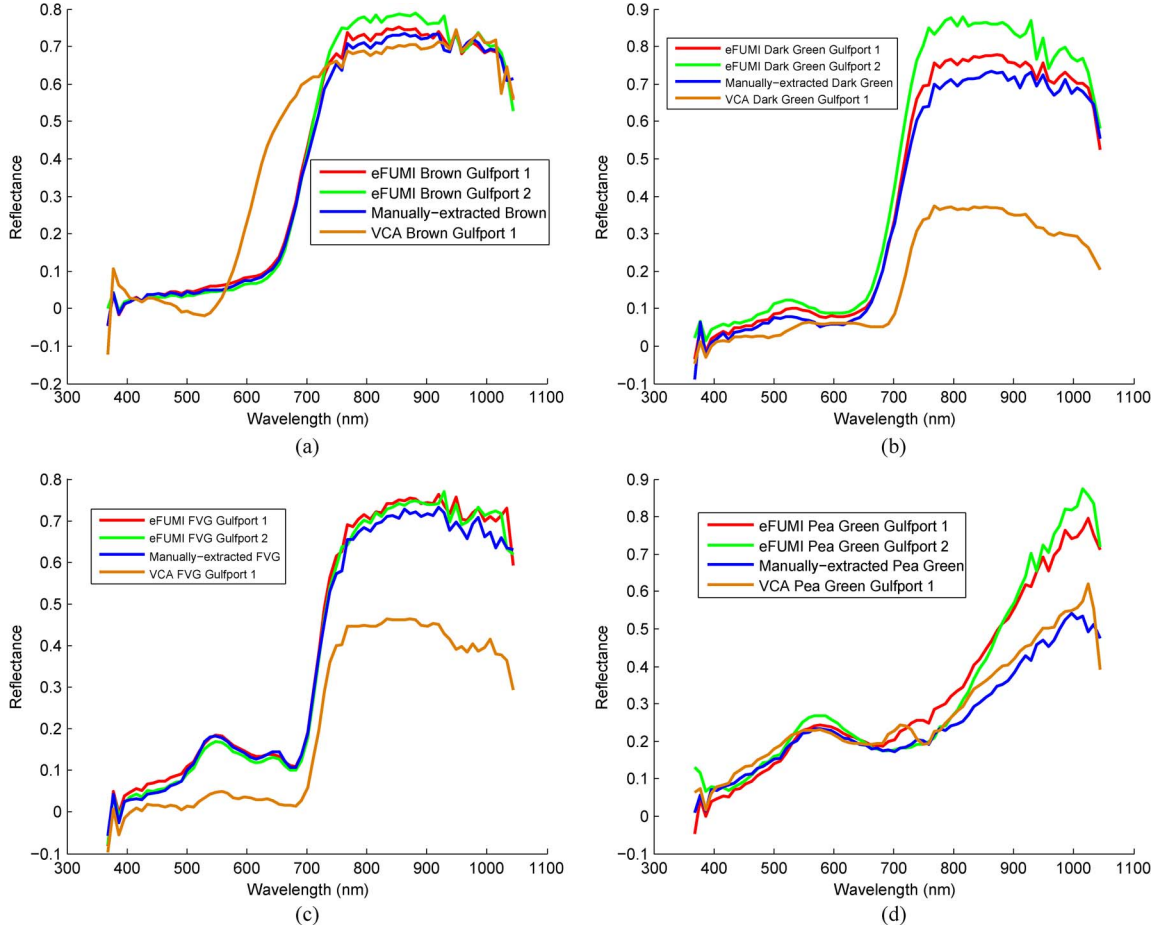
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

Fig. 10. Comparison of target spectral signatures found by *e*FUMI, VCA, and manually extracted spectra for (a) Brown, (b) Dark Green, (c) FVG, and (d) Pea Green targets.
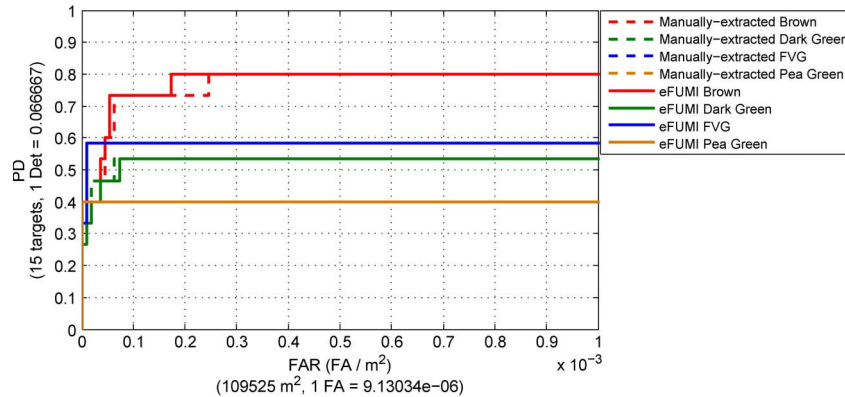

Fig. 11. Scoring results of Gulfport Data 2 estimated target spectra versus manually extracted spectra on Gulfport Data 1.

parameter settings for SVM, RVM, and mRVM, a large amount of parameter combinations were tested using the brown target. For the polynomial kernel, the polynomial order was varied from 2 to 5. For the RBF kernel, γ was varied from 0.5 to 5 at a step size of 0.5. Based on these results, we then did a fine search around the current optimal value (highest PD/FAR ratio) at a step of 0.1. Finally, the results with the highest PD/FAR ratio among all parameters except SVM with the RBF kernel are se-

lected to be shown in this paper. For SVM with the RBF kernel, an increase in γ beyond 2 will decrease both PD and FAR to eventually hit 0. Therefore, the second peak of the PD/FAR ratio for the SVM with the RBF kernel at $\gamma = 1.2$ is selected. The optimal parameters selected for the brown target were kept fixed for the other three target types. Each algorithm with each parameter setting was run ten times, and the mean of the detection results is shown in Tables VIII–X. For SVM and RVM,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JIAO AND ZARE: FUNCTIONS OF MULTIPLE INSTANCES FOR LEARNING TARGET SIGNATURES 11
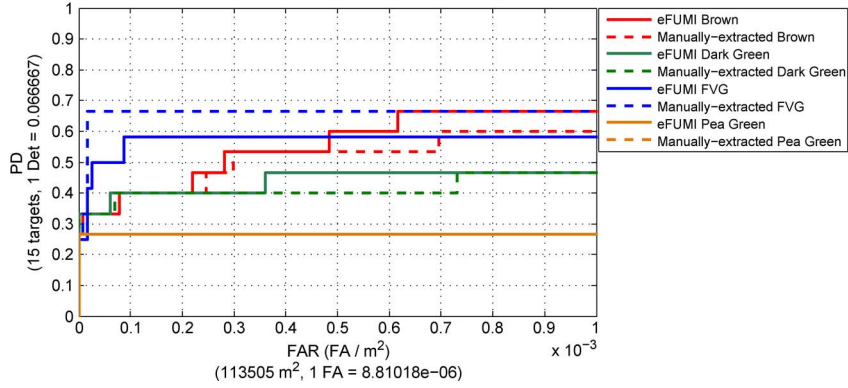

Fig. 12.   Scoring Results of Gulfport Data 1 estimated target spectra versus manually extracted spectra on Gulfport Data 2.

TABLE VII
PD VERSUS FAR BY ACE DETECTOR USING ESTIMATED AND GROUND-TRUTH TARGET SPECTRA, GULFPORT DATA

FAR	1×10^{-5}	1×10^{-4}	1×10^{-3}	FAR	1×10^{-5}	1×10^{-4}	1×10^{-3}
eFUMI Brown 2	40%	73.3%	80%	eFUMI Brown 1	33.33%	40%	66.67%
Manually-extracted Brown 2	40%	73.3%	80%	Manually-extracted Brown 1	33.33%	40%	60%
eFUMI Dark Green 2	33.33%	53.3%	53.3%	eFUMI Dark Green 1	33.33%	40%	46.67%
Manually-extracted Dark Green 2	40%	53.3%	53.3%	Manually-extracted Dark Green 1	33.33%	40%	46.67%
eFUMI FVG 2	58.33%	58.33%	58.33%	eFUMI FVG 1	25%	58.33%	58.33%
Manually-extracted FVG 2	58.33%	58.33%	58.33%	Manually-extracted FVG 1	33.33%	66.67%	66.67%
eFUMI Pea Green 2	33.3%	40%	40%	eFUMI Pea Green 1	26.67%	26.67%	26.67%
Manually-extracted Pea Green 2	40%	40%	40%	Manually-extracted Pea Green 1	26.67%	26.67%	26.67%

TABLE VIII
CLASSIFICATION RESULTS (OVER TEN RUNS) OF FOUR TYPES OF
TARGETS BY SVM ON GULFPORT DATA

Target Type	SVM with RBF kernel ($\gamma = 1.2$)		eFUMI	
	PD (%)	FAR($\times 10^{-3}$)	PD (%)	FAR($\times 10^{-3}$)
Brown	13.33	0.60	13.33	0
Dark Green	33.33	0.56	33.33	0
FVG	8.33	0.92	8.33	0
Pea Green	20.00	0.18	20.00	0

TABLE IX
CLASSIFICATION RESULTS (OVER TEN RUNS) OF FOUR TYPES OF
TARGETS BY RVM ON GULFPORT DATA

Target Type	RVM with RBF kernel ($\gamma = 0.7$)		eFUMI	
	PD (%)	FAR($\times 10^{-3}$)	PD (%)	FAR($\times 10^{-3}$)
Brown	46.67	0.81	46.67	0.22
Dark Green	40.00	0.28	40.00	0.06
FVG	50.00	9.35	50.00	0.03
Pea Green	93.33	21.35	93.33	103.081

the standard deviation is very small (in magnitude of 10^{-20}); thus, only the standard deviation of mRVM is shown. Generally, eFUMI performs much better than SVM, RVM, and mRVM. One item that must be pointed out is that the PD vs. FAR value for Pea Green by RVM is 93.33% and 21.35×10^{-3}; however, with such a high FAR, these results are essentially meaningless.

Finally, the performance of the eFUMI algorithm with respect to different label sizes was investigated. The training process was conducted on Flight 1 data, and the testing process was conducted on Flight 2 data. In this set of experiments, the size of positive bags was varied from 3×3, 5×5, 7×7, to 9×9 around ground-truth target points. Label uncertainty increases with the increase in label size as more points may or may not contain target material. The detection results are shown in Table XI. For the target type of FVG, eFUMI works well with the label in size 3×3 as compared to other sizes due to reduced uncertainty. However, we find that, for the Brown targets, eFUMI is able to detect a preferable target spectrum with label sizes 5×5 but performs worse with the label size of 3×3. This is because, for the brown target, a 3×3 region is too small to incorporate all the target information, so there are some targets that fall outside the 3×3 label due to inaccuracy in the ground truth. However, for the FVG target, eFUMI performs better with a positive bag size of 3×3 because the accuracy of the ground-truth locations is good (in comparison

to the Brown targets) and the strength of the FVG target is also weaker than that of Brown and often confused with nearby grass spectra. Thus, as expected (and desired), when the uncertainty in the labeling can be reduced (by accurately removing negative points from positive bags), the accuracy of eFUMI increases, and when the positive bags are reduced in size by removing true target points, the accuracy decreases.

Based on the experiment result discussed earlier, it is necessary to examine the performance of eFUMI with contaminated background information to determine the sensitivity of eFUMI given background contamination. In order to examine this, an additional synthetic data experiment was conducted. In this experiment, one positive and one negative with 100 points per bag were generated. The positive bag is composed of 50 target points (mean target proportion value of 0.551 with about 25% of these being pure target points) and 50 nontarget points. The negative bag is composed all nontarget points. Two methods of adding contamination points to the negative bag were examined: 1) One is by adding pure target points gradually from 1 to 100, and 2) the other is by adding mixed target points. These target points were generated using the same method as the target points in the positive bag. eFUMI was applied to these simulated data sets, and target detection performance given the estimated signatures was examined. Test data were generated in the same method as training data.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

TABLE X
CLASSIFICATION RESULTS (OVER TEN RUNS) OF FOUR TYPES OF TARGETS BY mRVM ON GULFPORT DATA

Target Type	mRVM with Polynomial kernel ($order = 3$)		eFUMI	
	PD (%)	FAR($\times 10^{-3}$)	PD (%)	FAR($\times 10^{-3}$)
Brown	22.00 ± 9.96	0.96 ± 0.41	22.00	0
Dark Green	16.67 ± 10.05	1.91 ± 1.24	16.67	0
FVG	19.17 ± 7.91	1.52 ± 0.04	19.17	0
Pea Green	22.67 ± 3.22	1.79 ± 0.44	22.67	0

TABLE XI
PD VERSUS FAR BY ACE DETECTOR USING ESTIMATED TARGET SPECTRA WITH DIFFERENT LABEL SIZE, GULFPORT DATA

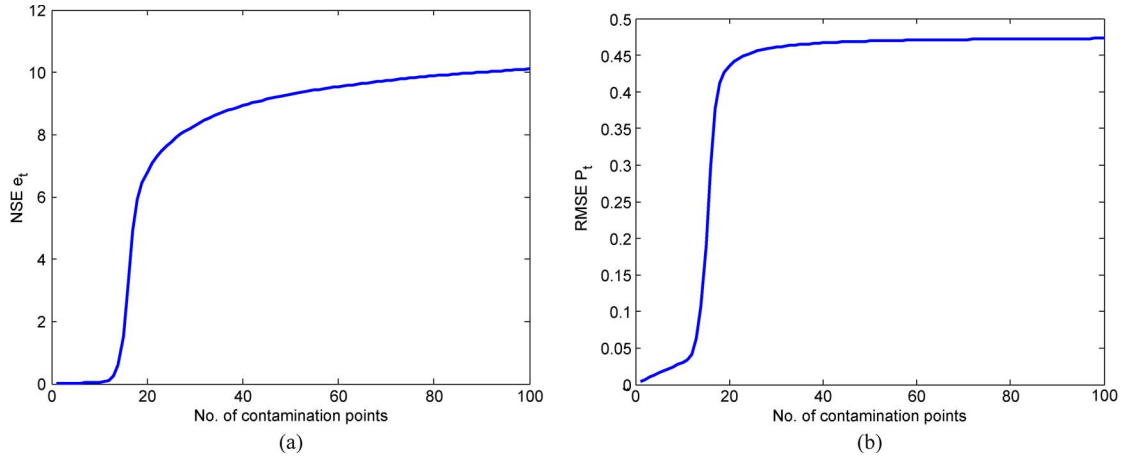| FAR	1×10^{-5}	1×10^{-4}	1×10^{-3}	FAR	1×10^{-5}	1×10^{-4}	1×10^{-3}
Manually Extracted Brown	33.33%	40%	60%	Manually Extracted FVG	33.33%	66.67%	66.67%
eFUMI Brown Label 3×3	26.67%	40%	66.67%	eFUMI FVG Label 3×3	41.67%	66.67%	66.67%
eFUMI Brown Label 5×5	33.33%	40%	66.67%	eFUMI FVG Label 5×5	25%	58.33%	58.33%
eFUMI Brown Label 7×7	26.67%	40%	66.67%	eFUMI FVG Label 7×7	16.67%	25%	25%
eFUMI Brown Label 9×9	26.67%	40%	66.67%	eFUMI FVG Label 9×9	0%	0%	16.67%

(a) (b)

Fig. 13. Error in estimated target spectra and proportions when adding pure target contamination points. (a) NSE of estimated target spectra with increasing number of pure contamination points. (b) RMSE of target proportion with increasing number of pure contamination points.
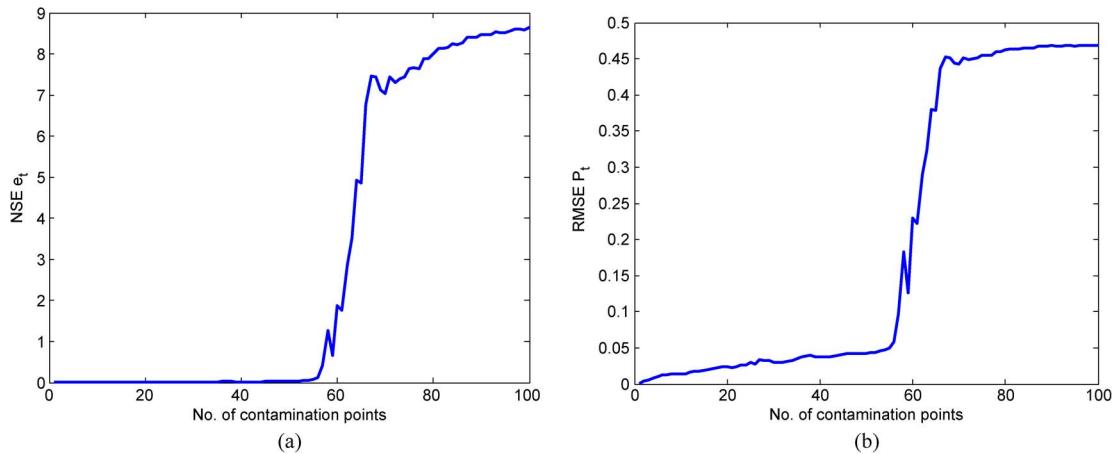

(a) (b)

Fig. 14. Error in estimated target spectra and proportions when adding mixed target contamination points. (a) NSE of estimated target spectra with increasing number of mixed contamination points. (b) RMSE of target proportion with increasing number of mixed contamination points.

Fig. 13 shows the normalized square error (NSE) and root mean square error (RMSE) between the estimated and true target signatures and the proportions as the number of pure contaminated points was increased, where $\text{NSE} = \|\mathbf{e}_{true} - \mathbf{e}_{estimate}\| / \|\mathbf{e}_{true}\|$, $\text{RMSE} = \sqrt{\mathbf{E}\left[(p_{true} - p_{estimate})^2\right]}$,

$\mathbf{e}_{true}$, $\mathbf{e}_{estimate}$ are the truth and estimated target spectra, and p_{true}, $p_{estimate}$ are the truth and estimated target proportion values, respectively. Fig. 14 shows the NSE and RMSE between the estimate and true target signatures and proportions as more and more mixed contamination points were added. As

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JIAO AND ZARE: FUNCTIONS OF MULTIPLE INSTANCES FOR LEARNING TARGET SIGNATURES 13
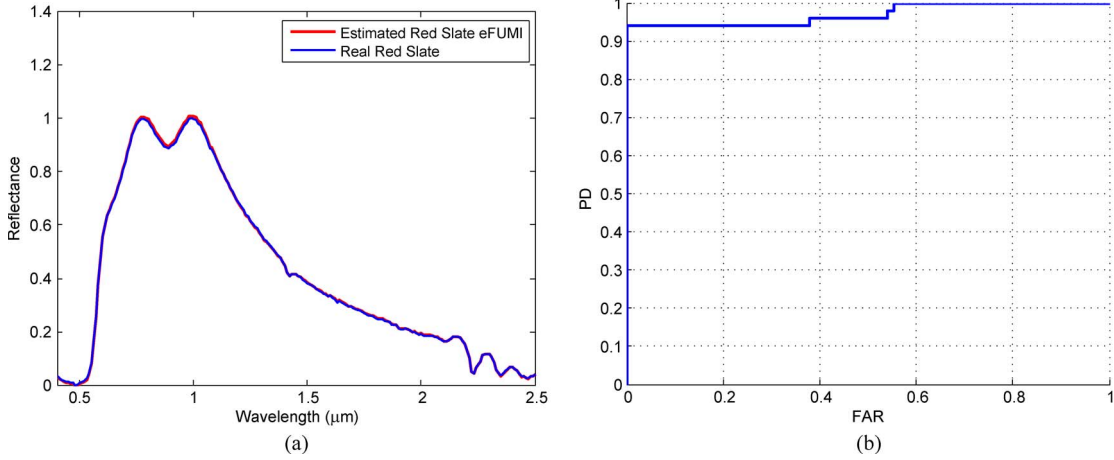

Fig. 15. Estimated target spectrum and testing ROC curve by adding ten pure target contamination points. (a) Estimated target spectra. (b) ROC curve.
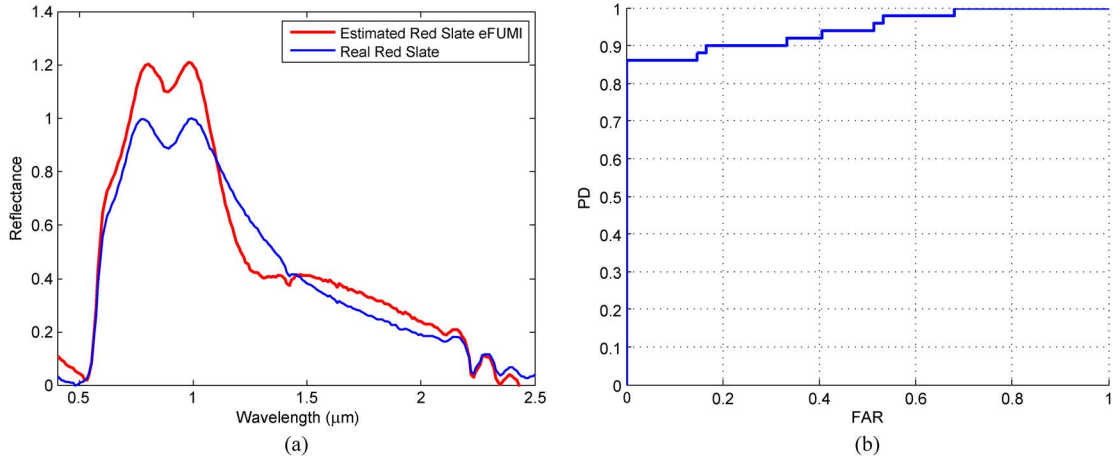

Fig. 16. Estimated target spectrum and testing ROC curve by adding 15 pure target contamination points. (a) Estimated target spectra. (b) ROC curve.

TABLE XII
NSE AND PD AT $FAR = 0.01$ WITH DIFFERENT NUMBER OF PURE TARGET CONTAMINATION POINTS IN BACKGROUND

No. of contamination points	NSE for e_t	RMSE for p_t	PD
10	5.69×10^{-2}	3.01×10^{-2}	94%
15	1.51	0.19	86%

one might expect, the error increases much more rapidly when the background is contaminated by pure target signatures. Furthermore, for the mixed data, the error does not steadily increase with more contamination points. This is due to the fact that the added target contamination points in the background have a big change in the target proportion value, thus making a considerable change in the mean of the target proportion value contaminated in the background. For example, the first peak of error appeared when adding contamination point 58, which is a pure contamination target point ($p_t = 1$). This peak is followed by reduction in error at contamination point 59 which is a weak contamination target point with $p_t = 0.13$. This illustrates that $eFUMI$ is robust to some target contamination and the performance is dependent upon the average target proportion of the contamination points. Figs. 15 and 16 are examples of estimated target signatures with 10 and 15 contaminated pure target points. The NSE, RMSE, and PD at $FAR = 0.01$ are shown in Table XII.

V. FUTURE WORK

There is much future work to be conducted in extending the $eFUMI$ algorithm. Based on experimentation, $eFUMI$ may be sensitive in some cases to positive points in a negative bag. For example, in the MUUFL Gulfport experiment, if the positive bag size for Brown targets is reduced to 3×3, the performance decreases significantly as compared with 5×5. This is mainly because there are some target points out of the 3×3 region around each ground-truth point. This is consistent with the definition of the MIL problem. However, if extensions can be made to incorporate unlabeled data (neither positive nor negative), the algorithm may benefit from more accurate positive and additional unlabeled data by allowing smaller positive bag regions and an "unknown" guard buffer region.

The method can also be sensitive to the fixed β parameter value when there are large differences in data point magnitude (i.e., large variance in the norm of each data). To solve this problem, the data often need to be normalized to have the same energy. However, in these cases, any information that may be provided by magnitude was lost. Investigation into an adaptive β scheduling could be conducted. Optimization strategies to set parameters for a new data set will be conducted. Current work also includes extending $eFUMI$ to a more general semisupervised dictionary learning approach that can be applied to a wide range of applications and data types.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

14

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

$$\mathbf{P}^+ = \{\mathbf{E}^T\mathbf{E}\}^{-1} \cdot \left\{ \mathbf{E}^T\mathbf{X}^+ + \frac{1}{a}\mathbf{R}\left(\begin{bmatrix} 0 \\ \mathbf{V} \end{bmatrix}\right)_{1\times N^+} + \mathbf{1}_{(M+1)\times 1} \cdot \frac{1 - \mathbf{1}_{1\times(M+1)}\{\mathbf{E}^T\mathbf{E}\}^{-1} \cdot \left\{ \mathbf{E}^T\mathbf{X}^+ + \frac{1}{a}\mathbf{R}\left(\begin{bmatrix} 0 \\ \mathbf{V} \end{bmatrix}\right)_{1\times N^+} \right\}}{\mathbf{1}_{1\times(M+1)}\{\mathbf{E}^T\mathbf{E}\}^{-1}\mathbf{1}_{(M+1)\times 1}} \right\}$$

(A5)

APPENDIX A

In this section, the derivation of cFUMI update equations is provided. In order to solve for the update equation for the proportion values, $\mathbf{P}$, it should be pointed out that solving for the proportion value of one point is not dependent on any other points. Considering only points in positive bags, the cFUMI objective function becomes the form shown in (A1), and a Lagrange multiplier term for the sum-to-one constraint is added in.

$$F^+ = \sum_{i=1}^{N^+} \left[\frac{1}{2}(1-u)w_i \left\| (\mathbf{x}_i - p_{it}\mathbf{e}_t - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k) \right\|_2^2 \right]$$

$$+ \frac{u}{2}\sum_{k=1}^{M} \|\mathbf{e}_k - \boldsymbol{\mu}_0\|_2^2 + \frac{u}{2}\|\mathbf{e}_t - \boldsymbol{\mu}_0\|_2^2$$

$$+ \sum_{i} \lambda_i^+(p_{it} + \sum_{k=1}^{M} p_{ik} - 1) + \sum_{k=1}^{M} \gamma_k \sum_{i=1}^{N^+} p_{ik}. \quad \text{(A1)}$$

Then, take the partial derivative of (A1) with respect to p_{it} and p_{ik}, respectively

$$\frac{\partial F^+}{\partial p_{it}} = (1-u)w_i(-1)\mathbf{e}_t^T(\mathbf{x}_i - p_{it}\mathbf{e}_t - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k) + \lambda_i^+$$

$$\frac{\partial F^+}{\partial p_{ik}} = (1-u)w_i(-1)\mathbf{e}_k^T(\mathbf{x}_i - p_{it}\mathbf{e}_t - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k) + \lambda_i^+ + \gamma_k.$$

Let us denote $a = (1-u)w_i(-1)$. Then, rewrite the above two functions into consistent matrix form, and set the expression to 0[2,3]

$$\frac{\partial F^+}{\partial \mathbf{P}_i^+} = a\mathbf{E}^T(\mathbf{x}_i - \mathbf{E}\mathbf{P}_i^+) + \lambda_i^+\mathbf{1}_{(M+1)\times 1} + \begin{bmatrix} 0 \\ \mathbf{V} \end{bmatrix} \quad \text{(A2)}$$

where $\mathbf{P}_i^+ = \begin{bmatrix} p_{it} \\ p_{i1} \\ p_{i2} \\ \vdots \\ p_{iM} \end{bmatrix} = \begin{bmatrix} p_{it} \\ \mathbf{P}_i^- \end{bmatrix}$, $\mathbf{V} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{bmatrix}$, and $\mathbf{E} = [\mathbf{e}_t \quad \mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_M] = [\mathbf{e}_t \quad \mathbf{E}^-]$.

$\mathbf{E}$ is the endmember matrix whose column corresponds to an endmember spectrum. $\mathbf{E}^-$ is a subset of $\mathbf{E}$ which accounts for constituent background endmembers. Similarly, $\mathbf{P}_i^+$ is the proportion vector for point $\mathbf{x}_i^+$, and $\mathbf{P}_i^-$ is a subset of $\mathbf{P}_i^+$,

which accounts for the proportion values with respect to background endmembers. For points $\mathbf{x}_i^-$ from negative bags, p_{it} is constrained to 0, so $\mathbf{P}_i = \begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}$.

Then, solve for $\mathbf{P}_i^+$

$$\mathbf{P}_i^+ = \{\mathbf{E}^T\mathbf{E}\}^{-1} \cdot \left\{ \mathbf{E}^T\mathbf{x}_i + \mathbf{1}_{(M+1)\times 1}\frac{\lambda_i^+}{a} + \frac{1}{a}\begin{bmatrix} 0 \\ \mathbf{V} \end{bmatrix} \right\}. \quad \text{(A3)}$$

In order to enforce the sum-to-one constraint, multiply $\mathbf{1}_{1\times(M+1)}$ on both sides of (A3), and use the sum-to-one constraint $\mathbf{1}_{1\times(M+1)}\mathbf{P}_i = 1$ to solve λ_i^+ shown as

$$\lambda_i^+ = \frac{a\left(1 - \mathbf{1}_{1\times(M+1)}\{\mathbf{E}^T\mathbf{E}\}^{-1} \cdot \left\{\mathbf{E}^T\mathbf{x}_i + \frac{1}{a}\begin{bmatrix} 0 \\ \mathbf{V} \end{bmatrix}\right\}\right)}{\mathbf{1}_{1\times(M+1)}\{\mathbf{E}^T\mathbf{E}\}^{-1}\mathbf{1}_{(M+1)\times 1}}. \quad \text{(A4)}$$

Finally, the update equation for the proportion values is found to be in (A5), shown at the top of the page.[4]

A similar derivation can be followed for points from negative bags (by simply excluding the term for the target endmember). The resulting update equation for negative points is shown in

$$\mathbf{P}^- = (\mathbf{E}^{-T}\mathbf{E}^-)^{-1}$$

$$\times \left[\mathbf{E}^{-T}\mathbf{X}^- - \frac{1}{1-u}\mathbf{R}(\mathbf{V})_{1\times N^-} + \mathbf{1}_{M\times 1} \right.$$

$$\left. \cdot \frac{1 - \mathbf{1}_{1\times M}(\mathbf{E}^{-T}\mathbf{E}^-)^{-1}(\mathbf{E}^{-T}\mathbf{X}^- - \frac{1}{1-u}\mathbf{R}(\mathbf{V})_{1\times N^-})}{\mathbf{1}_{1\times M}(\mathbf{E}^{-T}\mathbf{E}^-)^{-1}\mathbf{1}_{M\times 1}} \right]. \quad \text{(A6)}$$

To solve for the update for the endmember matrix $\mathbf{E}$, split the objective function into two parts according to points from positive bags and points from negative bags, and drop terms that are irrelevant to $\mathbf{E}$

$$F = \sum_{i=1}^{N^+} \left[\frac{1}{2}(1-u)w_i \left\| (\mathbf{x}_i - p_{it}\mathbf{e}_t - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k) \right\|_2^2 \right]$$

$$+ \frac{1}{2}(1-u)\sum_{i=1}^{N^-} \left[\left\| (\mathbf{x}_i - \sum_{k=1}^{M} p_{ik}\mathbf{e}_k) \right\|_2^2 \right]$$

$$+ \frac{u}{2}\sum_{k=1}^{M} \|(\mathbf{e}_k - \boldsymbol{\mu}_0)\|_2^2 + \frac{u}{2}\|(\mathbf{e}_t - \boldsymbol{\mu}_0)\|_2^2. \quad \text{(A7)}$$

---

[2] $\mathbf{0}_{d\times 1}$ and $\mathbf{1}_{1\times M}$ are column and row vectors with all elements 0 or 1.

[3] $[\mathbf{A} \; \mathbf{B}]$ and $\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}$ are the concatenation of arrays $\mathbf{A}$ and $\mathbf{B}$ horizontally and vertically, respectively.

[4] $\mathbf{R}(\mathbf{a}) = [\mathbf{a}, \mathbf{a}, \ldots, \mathbf{a}]_{1\times N}$ is the matrix containing repeated entries of $\mathbf{a}$ with a repetition of N vectors.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

JIAO AND ZARE: FUNCTIONS OF MULTIPLE INSTANCES FOR LEARNING TARGET SIGNATURES 15

$$\mathbf{P}_i^+ = \left\{ P(z_i=0)[\mathbf{0}_{d\times1} \ \mathbf{E}^-]^T[\mathbf{0}_{d\times1} \ \mathbf{E}^-] + P(z_i=1)\mathbf{E}^T\mathbf{E} \right\}^{-1} \cdot \left\{ \left[P(z_i=0) \, [\mathbf{0}_{d\times1} \ \mathbf{E}^-]^T + P(z_i=1)\mathbf{E}^T \right] \mathbf{x}_i + \frac{1}{a}\begin{bmatrix} 0 \\ \mathbf{V} \end{bmatrix} + \mathbf{1}_{(M+1)\times1} \right.$$

$$\left. \cdot \frac{1 - \mathbf{1}_{1\times(M+1)}\left\{ P(z_i=0)[\mathbf{0}_{d\times1} \ \mathbf{E}^-]^T[\mathbf{0}_{d\times1} \ \mathbf{E}^-] + P(z_i=1)\mathbf{E}^T\mathbf{E} \right\}^{-1} \cdot \left\{ \left[P(z_i=0)[\mathbf{0}_{d\times1} \ \mathbf{E}^-]^T + P(z_i=1)\mathbf{E}^T \right] x_i + \frac{1}{a}\begin{bmatrix} 0 \\ \mathbf{V} \end{bmatrix} \right\}}{\mathbf{1}_{1\times(M+1)}\left\{ P(z_i=0)[\mathbf{0}_{d\times1} \ \mathbf{E}^-]^T[\mathbf{0}_{d\times1} \ \mathbf{E}^-] + P(z_i=1)\mathbf{E}^T\mathbf{E} \right\}^{-1}\mathbf{1}_{(M+1)\times1}} \right\} \tag{A12}$$

$$\mathbf{E} = \left\{ (1-u)w_i \sum_{i=1}^{N^+}\left[P(z_i=0)\mathbf{x}_i\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}^T + P(z_i=1)\mathbf{x}_i\mathbf{P}_i^T \right] + (1-u)\sum_{i=1}^{N^-}\left[\mathbf{x}_i\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}^T \right] + u \cdot \mathbf{R}\,(\boldsymbol{\mu}_0)_{1\times(M+1)} \right\}$$

$$\cdot \left\{ (1-u)w_i \sum_{i=1}^{N^+}\left[P(z_i=0)\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}^T + P(z_i=1)\mathbf{P}_i\mathbf{P}_i^T \right] + (1-u)\sum_{i=1}^{N^-}\left[\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}^T \right] + u \right\}^{-1} \tag{A13}$$

Then, take the partial derivative of (A7) with respect to $\mathbf{e}_t$ and $\mathbf{e}_k$, respectively

$$\frac{\partial F}{\partial \mathbf{e}_t} = \sum_{i=1}^{N^+}\left[(-1)(1-u)w_i p_{it}\left(\mathbf{x}_i - p_{it}\mathbf{e}_t - \sum_{k=1}^{M}p_{ik}\mathbf{e}_k \right) \right] + u(\mathbf{e}_t - \boldsymbol{\mu}_0) \tag{A8}$$

$$\frac{\partial F}{\partial \mathbf{e}_k} = \sum_{i=1}^{N^+}\left[(-1)(1-u)w_i p_{ik}\left(\mathbf{x}_i - p_{it}\mathbf{e}_t - \sum_{k=1}^{M}p_{ik}\mathbf{e}_k \right) \right] + (1-u)\sum_{i=1}^{N^-}\left[-p_{ik}\left(\mathbf{x}_i - \sum_{k=1}^{M}p_{ik}\mathbf{e}_k \right) \right] + u(\mathbf{e}_k - \boldsymbol{\mu}_0). \tag{A9}$$

These can then be combined into matrix form and the expression set to 0

$$\frac{\partial F}{\partial \mathbf{E}} = \sum_{i=1}^{N^+}\left[(-1)(1-u)w_i(\mathbf{x}_i - \mathbf{E}\mathbf{P}_i)\mathbf{P}_i^T \right]$$

$$+ (1-u)\sum_{i=1}^{N^-}\left[-\left(\mathbf{x}_i - \mathbf{E}\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix} \right)\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}^T \right]$$

$$+ u\left(\mathbf{E} - \mathbf{R}(\boldsymbol{\mu}_0)_{1\times(M+1)} \right) = 0. \tag{A10}$$

Finally, the update equation for $\mathbf{E}$ is shown in

$$\mathbf{E} = \left\{ (1-u)w_i \sum_{i=1}^{N^+}\left[\mathbf{x}_i\mathbf{P}_i^T \right] \right.$$

$$+ (1-u)\sum_{i=1}^{N^-}\left[\mathbf{x}_i\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}^T \right] + u \cdot \mathbf{R}(\boldsymbol{\mu}_0)_{1\times(M+1)} \right\}$$

$$\cdot \left\{ (1-u)w_i\sum_{i=1}^{N^+}[\mathbf{P}_i\mathbf{P}_i^T] + (1-u)\sum_{i=1}^{N^-}\left[\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}\begin{bmatrix} 0 \\ \mathbf{P}_i^- \end{bmatrix}^T \right] + u \right\}^{-1}. \tag{A11}$$

APPENDIX B

Following an approach similar to what was used in cFUMI, update equations for eFUMI can be obtained by optimizing the objective function shown in (9). The resulting update equations for the proportion values are in (A12), shown at the top of the page. Here, it is difficult to write $\mathbf{P}_i$ in matrix form because, in $\left\{ P(z_i=0)[\mathbf{0}_{d\times1} \ \mathbf{E}^-]^T[\mathbf{0}_{d\times1} \ \mathbf{E}^-] + P(z_i=1)\mathbf{E}^T\mathbf{E} \right\}^{-1}$, $\mathbf{P}_i$ is related to $P(z_i)$ in an inverse matrix. Therefore, the proportion of points from positive bags must be updated point by point.

The proportion update equation for points from negative bags is the same as cFUMI shown in (A6).

The update equation for the endmember matrix is then found in (A13), shown at the top of the page.

APPENDIX C

Algorithm 3 describes how each simulated data point is generated according to its bag-level label and instance-level label following the model in (10) and (11). Algorithm 4 describes one method to generate simulated data given the target signature ($\mathbf{e}_T \in \mathbb{R}^d$), background signatures ($\mathbf{E}_{minus} \in \mathbb{R}^{d\times M}$), number of positive bags (num_{nbags}), number of negative bags (num_{nbags}), number of points in each bag (num_{points}), the number of positive points in each positive bag (n_{tar}), the minimum number of background endmembers per data point (N_b), the mean target proportion value (p_{t_mean}), and a parameter to control proportion variance (σ). The code produces the following outputs: $\mathbf{x}$: a synthetic data point, $\mathbf{X}$: full synthetic data matrix, $\mathbf{labels}_{bag}$: binary bag-level label vector, and $\mathbf{labels}_{point}$: binary instance-level label vector.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

16   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

Algorithm 3 Pseudo Code for Generating Linearly Mixed Data Point Given Bag-Level and Point-Level Label

---

Input: $\mathbf{e}_t, \mathbf{E}_{minus}, label_{bag}, label_{point}, N_b, p_{t\_mean}, \sigma$
Output: x
1: **if** $label_{bag}\&label_{point}$ **then**
2:     Uniformly draw integer m between $[N_b, M]$
3:     **if** $m == 0$ **then**
4:         $\boldsymbol{\alpha} = 1$
5:     **else**
6:         $\boldsymbol{\alpha} = \sigma \cdot [p_{t\_mean}, (1 - p_{t\_mean}/m) \times \mathbf{1}_{1 \times m}]$
7:     **end if**
8:     Randomly select m endmembers (denoted as $\mathbf{E}_{subminus}$) from $\mathbf{E}_{minus}$
9:     $\mathbf{p} \leftarrow$ sample $m + 1$ random values from Dirichlet Distribution given parameter $\boldsymbol{\alpha}$
10:     Generate point $\mathbf{x}$ following the Linear Mixing Model (10) using $\mathbf{p}$ and $[\mathbf{e}_t, \mathbf{E}_{subminus}]$
11: **else**
12:     Uniformly draw integer m between $[\max(1, N_b), M]$
13:     $\boldsymbol{\alpha} = \sigma \cdot \mathbf{1}_{1 \times m}$
14:     Randomly select m endmembers (denoted as $\mathbf{E}_{subminus}$) from $\mathbf{E}_{minus}$
15:     $\mathbf{p} \leftarrow$ sample m random values from Dirichlet Distribution given parameter $\boldsymbol{\alpha}$
16:     Generate point $\mathbf{x}$ following the Linear Mixing Model (11) using $\mathbf{p}$ and $\mathbf{E}_{subminus}$
17: **end if**

Algorithm 4 Pseudo Code for Generating Synthetic Data as Bags

---

Input: $\mathbf{e}_t, \mathbf{E}_{minus}, num_{pbags}, num_{nbags}, num_{points}, n_{tar}, N_b, p_{t\_mean}, \sigma$
Output: $\mathbf{X}, \mathbf{labels}_{bag}, \mathbf{labels}_{point}$
1: **for** $i \leftarrow 1$ to num_{pbags} **do**
2:   **for** $j \leftarrow 1$ to n_{tar} **do**
3:     $\mathbf{labels}_{bag}(i,j) = 1$
4:     $\mathbf{labels}_{point}(i,j) = 1$
5:     $x_{i,j}^+ \leftarrow$ Algorithm 3 given parameters set $\{\mathbf{e}_t, \mathbf{E}_{minus}, \mathbf{labels}_{bag}(i,j), \mathbf{labels}_{point}(i,j), N_b, p_{t\_mean}$ and $\sigma\}$
6:   **end for**
7:   **for** $j \leftarrow n_{tar} + 1$ to num_{points} **do**
8:     $\mathbf{labels}_{bag}(i,j) = 1$
9:     $\mathbf{labels}_{point}(i,j) = 0$
10:     $x_{i,j}^- \leftarrow$ Algorithm 3 given parameters set $\{\mathbf{e}_t, \mathbf{E}_{minus}, \mathbf{labels}_{bag}(i,j), \mathbf{labels}_{point}(i,j), N_b, p_{t\_mean}$ and $\sigma\}$
11:   **end for**
12: **end for**
13: **for** $i \leftarrow num_{pbags} + 1$ to $num_{pbags} + num_{nbags}$ **do**
14:   **for** $j \leftarrow 1$ to num_{points} **do**
15:     $\mathbf{labels}_{bag}(i,j) = 0$
16:     $\mathbf{labels}_{point}(i,j) = 0$
17:     $x_{i,j}^- \leftarrow$ Algorithm 3 given parameters set $\{\mathbf{e}_t, \mathbf{E}_{minus}, \mathbf{labels}_{bag}(i,j), \mathbf{labels}_{point}(i,j), N_b, p_{t\_mean}$ and $\sigma\}$
18:   **end for**
19: **end for**

REFERENCES

[1] D. Gillis *et al.*, "A generalized linear mixing model for hyperspectral imagery," in *Proc. SPIE*, Apr. 2008, vol. 6966, pp. 1–11.

[2] C. Cortes and V. Vapnik, "Support vector machine," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[3] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[4] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 4, pp. 580–585, Jul./Aug. 1985.

[5] J. Dula, A. Zare, D. Ho, and P. Gader, "Landmine classification using possibilistic k-nearest neighbors with wideband electromagnetic induction data," in *Proc. SPIE*, Jun. 2013, vol. 8709, Art. ID. 87091F.

[6] H. Frigui and P. Gader, "Detection and discrimination of land mines based on edge histogram descriptors and fuzzy k-nearest neighbors," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2006, pp. 1494–1499.

[7] H. Frigui and P. Gader, "Detection and discrimination of land mines in ground-penetrating radar based on edge histogram descriptors and a possibilistic k-nearest neighbor classifier," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 1, pp. 185–199, 2009.

[8] A. Ng and M. Jordan, "On discriminative versus generative classifiers: A comparison of logistic regression and naive Bayes," in *Proc. NIPS*, 2002, vol. 14, pp. 841–848.

[9] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 1. New York, NY, USA: Springer-Verlag, 2006.

[10] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[11] N. Keshava, J. Kerekes, D. Manolakis, and G. Shaw, "An algorithm taxonomy for hyperspectral unmixing," in *Proc. SPIE*, 2000, vol. 4049, pp. 42–63.

[12] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[13] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 58–78, 2009.

[14] M. Parente and A. Plaza, "Survey of geometric and statistical unmixing algorithms for hyperspectral images," in *Proc. 2nd Workshop WHISPERS*, 2010, pp. 1–4.

[15] J. M. Bioucas-Dias and A. Plaza, "An overview on hyperspectral unmixing: Geometrical, statistical, and sparse regression based approaches," in *Proc. IEEE IGARSS*, 2011, pp. 1135–1138.

[16] T.-H. Chan, W. K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, Nov. 2011.

[17] J. Wang and C.-I. Chang, "Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2601–2616, Sep. 2006.

[18] C.-I. Chang, C.-C. Wu, C.-S. Lo, and M.-L. Chang, "Real-time simplex growing algorithms for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1834–1850, Apr. 2010.

[19] M.-D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.

[20] A. Ifarraguerri and C.-I. Chang, "Multispectral and hyperspectral image analysis with convex cones," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 756–770, Mar. 1999.

[21] J. M. P. Nascimento and J. M. Bioucas Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 175–187, Jan. 2005.

[22] M. Berman *et al.*, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.

[23] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009.

[24] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[25] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jul. 2007.

[26] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.

[27] Y. Zhong, R. Feng, and L. Zhang, "Non-local sparse unmixing for hyperspectral remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1889–1909, Jun. 2014.

[28] X. Lu, X. Wu, Y. Yuan, P. Yan , and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, May 2013.

[29] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.

[30] F. Chen and Y. Zhang, "Sparse hyperspectral unmixing based on constrained $\ell_p - \ell_2$ optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1142–1146, Sep. 2013.

[31] Z. Shi, W. Tang, Z. Duren, and Z. Jiang, "Subspace matching pursuit for sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3256–3274, Jun. 2014.

[32] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operators," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2025–2041, Sep. 2002.

[33] D. M. Rogge *et al.*, "Integration of spatial-spectral information for the improved extraction of endmembers," *Remote Sens. Environ.*, vol. 110, no. 3, pp. 287–303, Oct. 2007.

[34] A. Zare, O. Bchir, H. Frigui, and P. Gader, "Spatially-smooth piece-wise convex endmember detection," in *Proc. WHISPERS*, 2010, pp. 1–4.

[35] A. Zare and P. Gader, "Piece-wise convex spatial-spectral unmixing of hyperspectral imagery using possibilistic and fuzzy clustering," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2011, pp. 741–746.

[36] M. Xu, B. Du, and L. Zhang, "Spatial-spectral information based abundance-constrained endmember extraction methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1939–1404, Jun. 2014.

[37] T. C. Glenn, "Context dependent detection in hyperspectral imagery," Ph.D. dissertation, Univ. Florida, Gainesville, FL, USA, Dec. 2013.

[38] B. Du and L. Zhang, "Target detection based on a dynamic subspace," *Pattern Recognit.*, vol. 47, no. 1, pp. 344–358, Jan. 2014.

[39] B. D. R. Zhao and L. Zhang, "A robust nonlinear hyperspectral anomaly detection approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1227–1234, Apr. 2014.

[40] B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, Nov. 2014.

[41] S. M. Kay, *Fundamental of Statistical Signal Processing: Volume II—Detection Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[42] S. Kraut and L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2538–2541, Sep. 1999.

[43] S. Kraut, L. Scharf, and L. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.

[44] W. F. Basener, "Clutter and anomaly removal for enhanced target detection," in *Proc. SPIE*, Apr. 2010, vol. 7695, Art. ID. 769525.

[45] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.

[46] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Perez, "Solving the multiple-instance problem with axis-parallel rectangles," *Artif. Intell.*, vol. 89, no. 1/2, pp. 31–17, Jan. 1997.

[47] O. Maron and T. Lozano-Perez, "A framework for multiple-instance learning" in *Proc. NIPS*, 1998, vol. 10, pp. 570–576.

[48] Q. Zhang and S. Goldman, "EM-DD: An improved multiple-instance learning technique," in *Proc. NIPS*, 2002, vol. 2, pp. 1073–1080.

[49] J. Bolton and P. Gader, "Multiple instance learning for hyperspectral image analysis," in *Proc. IEEE IGARSS*, 2010, pp. 4232–4235.

[50] J. Bolton, P. Gader, H. Frigui, and P. Torrione, "Random set framework for multiple instance learning," *Inf. Sci.*, vol. 181, no. 11, pp. 2061–2070, Jan. 2011.

[51] V. Raykar, B. Krishnapuram, J. Bi, M. Dundar, and R. Rao, "Bayesian multiple instance learning: Automatic feature selection and inductive transfer," in *Proc. ICML*, 2008, pp. 808–815.

[52] A. Zare and P. Gader, "Pattern recognition using functions of multiple instances," in *Proc. 20th ICPR*, 2010, pp. 1092–1095.

[53] A. Zare *et al.*, "Sub-pixel target spectra estimation and detection using functions of multiple instances," in *Proc. 3rd WHISPERS*, 2011, pp. 1–4.

[54] A. Zare and C. Jiao, "Extended functions of multiple instances for target characterization," in *Proc. WHISPERS*, 2014, pp. 1–4.

[55] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

[56] A. Baldridge, S. Hook, C. Grove, and G. Rivera, "The ASTER spectral library version 2.0," *Remote Sens. Environ.*, vol. 113, no. 4, pp. 711–715, Apr. 2009.

[57] S. Holzwarth *et al.*, "Hysens-DAIS 7915/ROSIS imaging spectrometers at DLR," in *Proc. 3rd EARSel, Workshop Imag. Spectroscopy, Herrsching*, May 2003, pp. 3–14.

[58] A. Giannandrea *et al.*, "The SHARE 2012 data campaign," in *Proc. SPIE*, 2013, vol. 8743, Art. ID. 87430F.

[59] P. Gader, A. Zare, R. Close, J. Aitken, and G. Tuell, "MUUFL gulfport hyperspectral and lidar airborne data set," Univ. Florida, Gainesville, FL, USA, Tech. Rep. REP-2013-570, Oct. 2013.

[60] T. Glenn, A. Zare, P. Gader, and D. Dranishnikov, "Bullwinkle: Scoring code for sub-pixel targets (version 1.0) [software]," 2013. [Online]. Available: http://engineers.missouri.edu/zarea/code/

[61] I. Psorakis, T. Damoulas, and M. A. Girolami, "Multiclass relevance vector machines: Sparsity and accuracy," *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1588–1598, Oct. 2010.

Changzhe Jiao (S'13) received the B.S. and M.S. degrees in control theory from Xidian University, Xi'an, China, in 2007 and 2012, respectively.

He is currently a Graduate Research Assistant working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, USA. His research interests include multiple instance learning and hyperspectral image analysis.

Alina Zare (S'07–M'08–SM'13) received the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 2008.

She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO, USA. Her research interests include machine learning, computational intelligence, Bayesian methods, sparsity promotion, image analysis, pattern recognition, hyperspectral image analysis, and remote sensing.

Dr. Zare is a recipient of the 2014 National Science Foundation CAREER award and the 2014 National Geospatial-Intelligence Agency New Investigator Program Award.