Multiple Instance Hybrid Estimator for Hyperspectral Target Characterization and Sub-pixel Target Detection

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

***Abstract*—The Multiple Instance Hybrid Estimator for discriminative target characterization from imprecisely labeled hyperspectral data is presented. In many hyperspectral target detection problems, acquiring accurately labeled training data is difficult. Furthermore, each pixel containing target is likely to be a mixture of both target and non-target signatures (*i.e.*, sub-pixel targets), making extracting a pure prototype signature for the target class from the data extremely difficult. The proposed approach addresses these problems by introducing a data mixing model and optimizing the response of the hybrid sub-pixel detector within a multiple instance learning framework. The proposed approach iterates between estimating a set of discriminative target and non-target signatures and solving a sparse unmixing problem. After learning target signatures, a signature based detector can then be applied on test data. Both simulated and real hyperspectral target detection experiments show the proposed algorithm is effective at learning discriminative target signatures and achieves superior performance over state-of-the-art comparison algorithms.**

***Index Terms*—target detection, hyperspectral, endmember extraction, multiple instance learning, hybrid detector, target characterization**

I. Introduction

Hyperspectral imaging spectrometers collect electromagnetic energy scattered in the scene across hundreds or thousands of spectral bands, capturing both the spatial and spectral information [1]. The spectral information is a combination of the reflection and/or emission of sunlight across wavelength by objects on the ground, and contains the unique spectral characteristics of different materials [2, 3]. The wealth of spectral information in hyperspectral imagery enables the possibility to conduct sub-pixel analysis in application areas including target detection [4, 5], precision agriculture [6, 7], biomedical applications [8, 9] and others [10, 11].

Hyperspectral target detection generally refers to the task of locating all instances of a target given a known spectral signature within a hyperspectral scene. The reasons many classification methods are not applicable to hyperspectral target detection tasks are threefold:

Changzhe Jiao is with the Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, 65211 USA; e-mail: cjiao@mail.missouri.edu. Alina Zare is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611 USA; e-mail: azare@ufl.edu. Ronald McGarvey is with the Department of Industrial and Manufacturing Systems Engineering, University of Missouri, Columbia, MO, 65211 USA; e-mail: mcgarveyr@missouri.edu. Manuscript received October xx, 2017. This material is based upon work supported by the National Science Foundation under Grant No. IIS-1350078 - CAREER: Supervised Learning for Incomplete and Uncertain Data.

1. Precise training labels for sub-pixel targets are often difficult or infeasible to obtain. For example, the ground truth information often is obtained using a Global Positioning System (GPS) receiver placed at the target location. However, the accuracy of the GPS coordinates could drift for several meters depending on the accuracy of the GPS system. Thus, a target pixel denoted by the GPS coordinates could be a false positive point and the positive locations are not marked. The only reliable knowledge is that within a certain region around the GPS truth, some target can be found.

2. The number of training instances from the positive (target) class is often small compared to that of the negative training data such that training an effective classifier is difficult. A hyperspectral image with hundred of thousands of pixels may only have a few pixels or sub-pixel level target points.

3. Due to the relatively low spatial resolution of hyperspectral imagery and the diversity of natural scenes, many targets are mixed points (sub-pixel targets). Most of the supervised learning algorithms assume each training data point is a pure prototype of a class denoted by the label paired with this data. However, in a hyperspectral image, a target pixel could be a mixture of several background materials and the amount of the target proportion is unknown.

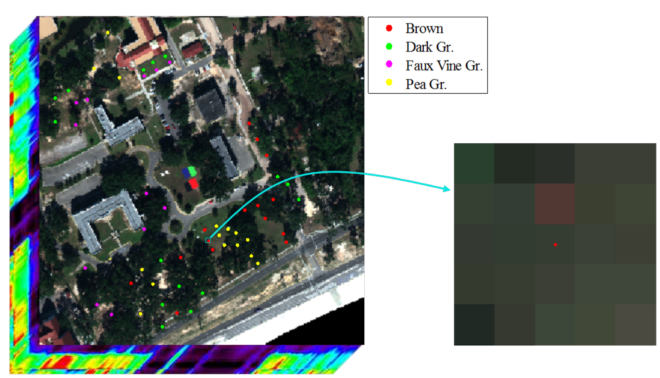


Fig. 1. Illustration of inaccurate coordinates from GPS: one target denoted as brown by GPS has one pixel drift.

For the above reasons, signature based hyperspectral target detection [12, 13] is generally applied over classification approaches. However, the performance of signature based detectors relies heavily on the quality of the target spectral signature used. Yet, obtaining an effective target signature can be a challenging problem. Common methods for obtaining target signatures such as manual selection of a target signature from the scene may result in selecting a mixed (as opposed to a pure target) spectrum. Target signatures collected from hand-held sensors on the ground over a target of interest or selected from a spectral library do not account for variation due to environmental or atmospheric conditions. Furthermore, accurately locating a target of interest is also challenging. Sub-pixel targets are difficult to see in the imagery and the co-registration of the targets to GPS coordinates could drift across several pixels.

As an example, Fig. 1 shows the scattered target locations over the MUUFL Gulfport data set collected over the University of Southern Mississippi-Gulfpark Campus [14], where there are 4 types of targets throughout the scene: Brown (15 examples), Dark Green (15 examples), Faux Vineyard Green (12 examples) and Pea Green (15 examples). The highlighted region shown in Fig. 1 is one of the brown target locations whose zoomed view is also shown, where we can clearly see that for this brown target there is one pixel drift between the real target location and ground truth location given by GPS. This is one of the rare examples we can visually see the brown target. Most of the targets are difficult to distinguish visibly.

In this paper, we model the hyperspectral target estimation task as a multiple instance concept learning problem [15, 16] and present the multiple instance hybrid estimator (MI-HE) for characterization of a target signature from imprecisely labeled hyperspectral imagery. Here, *concepts* refer to the generalized class prototypes in the feature space. In the case of hyperspectral image unmixing, a concept is the spectral signature of one assumed pure material in the scene, also known as an endmember. MI-HE explicitly addresses the aforementioned problems from the following aspects:

1. **Uncertain labels**. MI-HE adopts the idea of "bags" from multiple instance learning (MIL), introduces the multiple instance data mixing model and proposes a multiple instance concept learning framework to address the uncertain labels in hyperspectral target detection, *e.g.*, target locations coming from GPS or region of interest manually denoted by analysts.

2. **Unbalanced number of target and non-target pixels in training data**. MI-HE addresses this problem by applying a signature-based detector only to the pixels from possible target regions denoted by a GPS in the training step. The non-target regions are only needed to refine the background concepts.

3. **Mixed training data.** MI-HE models each pixel as a linear combination of target and/or non-target concepts, so the estimated target and background concepts applied to testing data can be used to perform sub-pixel detection.

Several previous methods for target characterization have been developed in the literature. The FUnctions of Multiple Instance (FUMI) algorithms [17, 18] learn representative concept from reconstruction error of the uncertainly labeled data. Compared with FUMI, MI-HE learns a discriminative target concept that maximizes the detection response of possible target regions so the estimated signatures are more discriminative for target detection. Furthermore, the FUMI algorithms do not exploit the complete label information from the training data, *i.e.*, the FUMI algorithms combine all possible target regions together into one target bag and thus discards the information that each target region contains at least one target instance. Thus is not applicable to some target detection problems, *e.g.*, negative bags cannot provide the entire background information. On the contrary, MI-HE adopts a generalized mean model to differentiate each individual target region. Compared to the discriminative target characterization algorithms, MI-ACE and MI-SMF [19] (which maximize the matched filter response in an MIL framework and estimates only one target signature), MI-HE has the potential to learn multiple signatures to address signature variability.

Expanding upon our work in [20], the MI-HE algorithm and experiments presented in this paper maintain the following improvements and advantages in comparison to our prior work: (1) introducing a discriminative concept learning term; (2) improved gradient descent optimization using Armijo's rule; (3) comprehensive experiments on multiple concepts learning (both simulated and realistic) and more types of target for Gulfport data; (4) comprehensive comparison with state-of-the-art MIL algorithms; (5) analysis of MI-HE's robustness to parameter setting.

II. Related Methods

Since the proposed MI-HE optimizes the detection statistics of some signature based detector under a MIL framework, this section provides a literature review on both the signature based detectors and MIL algorithms.

A. Signature Based Detectors

In statistical target detection 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 ($\boldsymbol{H}_0$) or target present ($\boldsymbol{H}_1$) and a detector can be designed using the generalized likelihood ratio test (GLRT) approach [21]. Following the Neyman-Pearson criterion that maximizes the probability of detection (PD) given any desired probability of false alarm (PFA), the GLRT is shown in Eq. (1),

$$\Lambda(\mathbf{x}) = \frac{f(\mathbf{x}|\text{Target present})}{f(\mathbf{x}|\text{Target absent})} \triangleq \frac{f(\mathbf{x}|H_1)}{f(\mathbf{x}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (1)$$

where $f(\mathbf{x}|H_i)$ is the likelihood function value for each hypothesis.

A large number of hyperspectral target detection methods have been developed in the literature [12, 13, 22, 23]. The adaptive coherence/cosine estimator (ACE) [24–26] and the hybrid detector [27, 28] are two such effective and widely used sub-pixel detection algorithms. The hypotheses used for the ACE are:

$$\begin{aligned} \mathbf{H}_0 &: \mathbf{x} \sim \mathcal{N}\left(0, \sigma_0^2 \Sigma_b\right) \\ \mathbf{H}_1 &: \mathbf{x} \sim \mathcal{N}\left(a\mathbf{s}, \sigma_1^2 \Sigma_b\right) \end{aligned} \quad (2)$$

where $\boldsymbol{\Sigma}_b$ is the background covariance and $\boldsymbol{s}$ is the known target signature which is scaled by a target abundance, a. The square-root of the GLRT for (2) results in the following as the ACE detector:

$$\Lambda_{ACE}(\mathbf{x},\mathbf{s}) = \frac{(\mathbf{s}-\boldsymbol{\mu}_b)^T\boldsymbol{\Sigma}_b^{-1}(\mathbf{x}-\boldsymbol{\mu}_b)}{\sqrt{(\mathbf{s}-\boldsymbol{\mu}_b)^T\boldsymbol{\Sigma}_b^{-1}(\mathbf{s}-\boldsymbol{\mu}_b)}\sqrt{(\mathbf{x}-\boldsymbol{\mu}_b)^T\boldsymbol{\Sigma}_b^{-1}(\mathbf{x}-\boldsymbol{\mu}_b)}}, \quad (3)$$

where $\boldsymbol{\mu}_b$ is the background mean subtracted from the data to ensure a zero-mean background as defined in $\boldsymbol{H}_0$.

In comparison, the hypotheses used for the background structured hybrid detector (HSD) are:

$$\begin{aligned} \mathbf{H}_0 &: \mathbf{x} \sim \mathcal{N}\left(\mathbf{D}^-\mathbf{p}, \sigma_0^2\Sigma_b\right) \\ \mathbf{H}_1 &: \mathbf{x} \sim \mathcal{N}\left(\mathbf{D}\mathbf{a}, \sigma_1^2\Sigma_b\right) \end{aligned} \quad (4)$$

where $\mathbf{D} = [\mathbf{D}^+ \;\; \mathbf{D}^-]$ is the combination of endmember sets $\mathbf{D}^+$ and $\mathbf{D}^-$ that represent the target and background endmembers set, respectively. $\mathbf{a}$ and $\mathbf{p}$ are the abundance values computed by Fully Constrained Least Squares (FCLS) [29] corresponding to $\mathbf{D}$ and $\mathbf{D}^-$. The GLRT for (4) results in the HSD detector:

$$\Lambda_{HSD}(\mathbf{x},\mathbf{D}) = \frac{(\mathbf{x}-\mathbf{D}^-\mathbf{p})^T\boldsymbol{\Sigma}_b^{-1}(\mathbf{x}-\mathbf{D}^-\mathbf{p})}{(\mathbf{x}-\mathbf{D}\mathbf{a})^T\boldsymbol{\Sigma}_b^{-1}(\mathbf{x}-\mathbf{D}\mathbf{a})}, \quad (5)$$

The hybrid detector models the reconstruction error of each point as a zero mean Gaussian distribution using the entire endmember set and non-target endmember set, respectively. The ratio between the reconstruction error using the entire endmember set and only the non-target endmember set escalates the difference in the two reconstruction errors. The hybrid detector explicitly models the mixture in hyperspectral data and provides a sub-pixel detection alternative.

B. Multiple Instance Learning

Multiple instance learning (MIL) was first investigated by Dietterich *et al.* [15] in the 1990s for the prediction of drug activity (musk activity). In MIL, training data is partitioned into sets of labeled “bags” (instead of being individually labeled) in which a bag is defined to be a multi-set of data points. A positive bag must contain at least one true positive (target) data point and negative bags are composed entirely of negative data. Thus, data point-specific training labels are unavailable. Given training data in this form, the majority of MIL methods either: (1) learn target concepts for describing the target class; or (2) train a classifier that can distinguish between individual target and non-target data points and/or bags.

Since the introduction of the MIL framework [15], many methods have been proposed and developed in the literature. The majority of MIL approaches focus on learning a classification decision boundary to distinguish between positive and negative instances/bags from the ambiguously labeled data. The mi-SVM [30] models the MIL problem as a generalized mixed integer formulation of support vector machine and was solved iteratively between training a regular SVM and heuristic reassignment of the training labels. The Multiple-Instance Learning via Embedded Instance Selection (MILES) [31] relaxes the constraint in MIL that negative bags are composed of all negative instances and allows target concept to be related to negative bags for a more general application in computer vision. MILES proposed to first embed each bag to a target concept based feature space, where the set of candidate target concept comes from the union of all bags. Then a 1-norm SVM [32] was trained on the feature vectors extracted from each bag; finally, an instance selection was performed based on the SVM decision value to realize instance-level classification. The Max-Margin Multiple-Instance Dictionary Learning (MMDL) [33] adopts the idea of bag of words (BoW) model [34] and treats a set of linear SVMs as a codebook. Then each image was represented as a distribution over the codebook using spatial pyramid matching [35]. Finally another linear SVM was trained for image-level classification. The novel assumption of MMDL is that the positive instances could belong to many different clusters.

Although the above mentioned approaches are effective at training classifiers given imprecise labels, they generally do not provide an intuitive description or *representative concept* that characterizes the salient and discriminative features of the target class. The few existing approaches that estimate a target concept include Diverse Density (DD) [16] that estimates a concept by minimizing its distance to at least one instance from each positive bag and maximizing its distance from all instances in negative bags. The Expectation-Maximization (EM) version of diverse density (EMDD) [36] iteratively estimates which instances in each positive bag belong to the target class and then only uses those points from the positive bags to estimates a target concept that maximizes the diverse density. eFUMI [17, 37] treats each instance as a convex combination of positive and/or negative concepts and estimates the target and non-target concepts using an EM approach. The MI-SMF and MI-ACE [19] maximizes response of SMF (spectral matched filter [22, 38, 39]) and ACE respectively under a multiple instance learning framework and efficiently learns a discriminative target signature. However, most of these prototype-based methods only find a single target concept and are, thus, unable to account for large variation in the target class.

III. Multiple Instance Hybrid Estimator

Let $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \mathbb{R}^{d\times N}$ be training data where d is the dimensionality of an instance and N is the total number of training instances. The data are grouped and arranged into K *bags*, $\mathbf{B} = \{\mathbf{B}_1, \ldots, \mathbf{B}_K\}$, such that the first K^+ bags are positively labeled with associated binary bag-level label $L_i = 1, i = 1, \cdots, K^+$, and the rest K^- bags are negatively labeled with associated binary bag-level label $L_i = 0, i = K^+ + 1, \cdots, K$. N^+ and N^- are the total number of positive instances and negative instances, as indicated below:

$$N = N^+ + N^- = \sum_{i=1}^{K^+} N_i + \sum_{i=K^++1}^{K} N_i, \quad (6)$$

where N_i is the number of instances in bag $\mathbf{B}_i$. $\mathbf{x}_{ij} \in \mathbf{B}_i$ denotes the j^{th} instance in bag $\mathbf{B}_i$ with instance-level label $l_{ij} \in \{0, 1\}$.

Since the MIL problem states that there must be at least one positive instance in each positive bag and each negative bag must consist of only negative instances, we can approximate the probability of an individual bag to the instances in each bag, as shown in Eq. (7). Specifically, the probability for a positive bag to be positive is substituted by the instance in this bag with highest "positiveness" and the probability for a negative bag to be negative is represented by the joint probability of all instances in this bag to be negative.

$$J_1 = \prod_{i=1}^{K^+} \max_{\mathbf{x}_{ij} \in \mathbf{B}_i} \Pr(l_{ij} = +|\mathbf{B}_i) \prod_{i=K^++1}^{K} \prod_{j=1}^{N_i} \Pr(l_{ij} = -|\mathbf{B}_i). \tag{7}$$

Eq. (7) contains a $\max$ operation that is difficult to optimize numerically. Some algorithms in the literature [36, 40] adopt a noisy-OR model instead of using $\max$. However, experimental results show that the noisy-OR model is highly non-smooth and needs to be repeated with many different initializations (typically using every positive training instance) to avoid local optima. In the proposed approach, we adopt the generalized mean as an alternative of $\max$ operation, as shown in (8).

$$J_2 = \prod_{i=1}^{K^+} \left(\frac{1}{N_i} \sum_{j=1}^{N_i} \Pr(l_{ij} = +|\mathbf{B}_i)^b \right)^{\frac{1}{b}} \prod_{i=K^++1}^{K} \prod_{j=1}^{N_i} \Pr(l_{ij} = -|\mathbf{B}_i), \tag{8}$$

where $b \in [-\infty, +\infty]$ is a real number controlling the function to approximately vary from $\min$ to $\max$.

Then taking the negative logarithm and scaling the second term of Eq. (8) result in:

$$\begin{aligned} -\ln J_2 &= -\sum_{i=1}^{K^+} \frac{1}{b} \ln \left(\frac{1}{N_i} \sum_{j=1}^{N_i} \Pr(l_{ij} = +|\mathbf{B}_i)^b \right) \\ &- \rho \sum_{i=K^++1}^{K} \sum_{j=1}^{N_i} \ln \Pr(l_{ij} = -|\mathbf{B}_i), \end{aligned} \tag{9}$$

where the scaling factor ρ is usually set to be smaller than one to control the influence of negative bags.

Here each instance is modeled as a sparse linear combination of target and/or background concepts $\mathbf{D}$, $\mathbf{x}_i \approx \mathbf{D}\mathbf{a}_i$, where $\mathbf{D} = \begin{bmatrix} \mathbf{D}^+ & \mathbf{D}^- \end{bmatrix} \in \mathbb{R}^{d \times (T+M)}$, $\mathbf{D}^+ = [\mathbf{d}_1, \cdots, \mathbf{d}_T]$ is the set of T target concepts and $\mathbf{D}^- = [\mathbf{d}_{T+1}, \cdots, \mathbf{d}_{T+M}]$ is the set of M background concepts, $\mathbf{a}_i$ is the sparse vector of weights for instance $\mathbf{x}_i$. Given MIL definition, each positive bag must contain at least one instance composed of some target:

$$\text{if } L_i = 1, \exists \mathbf{x}_{ij} \in \mathbf{B}_i, i \in [1, \cdots, K^+], \text{ s.t.}$$
$$\mathbf{x}_{ij} = \sum_{t=1}^{T} a_{it}\mathbf{d}_t + \sum_{k=T+1}^{T+M} a_{ik}\mathbf{d}_k + \boldsymbol{\varepsilon}_{ij}, a_{it} \neq 0, \tag{10}$$

where $\boldsymbol{\varepsilon}_i$ is a noise term.

Similarly, each negatively labeled bag should not contain any target:

$$\text{if } L_i = 0, \forall \mathbf{x}_{ij} \in \mathbf{B}_i, i \in [K^+ + 1, \cdots, K], \text{ s.t.} \tag{11}$$
$$\mathbf{x}_{ij} = \sum_{k=T+1}^{T+M} a_{ik}\mathbf{d}_k + \boldsymbol{\varepsilon}_{ij}. \tag{12}$$

Given the above data model, we introduce the hybrid detector to estimate if instances from positive bags are the positive target points. Specifically, define the following term,

$$\Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i, L_i = 1) = \exp\left(-\beta \frac{\|\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij} - \mathbf{D}^-\mathbf{p}_{ij}\|^2} \right), \tag{13}$$

where β is a scaling parameter, $\mathbf{a}_{ij}$ and $\mathbf{p}_{ij}$ are the sparse representation of $\mathbf{x}_{ij}$ given entire concept set $\mathbf{D}$ and background concept set $\mathbf{D}^-$. The residual vectors,

$$\begin{cases} \mathbf{r}_{ij} = (\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}) \\ \mathbf{q}_{ij} = (\mathbf{x}_{ij} - \mathbf{D}^-\mathbf{p}_{ij}) \end{cases}, \tag{14}$$

are the reconstruction residual vectors corresponding to $\mathbf{D}$ and $\mathbf{D}^-$, respectively. Further, $\mathbf{a}_{ij} = [\mathbf{a}_{ij}^+;\ \mathbf{a}_{ij}^-]$, where $\mathbf{a}_{ij}^+$ and $\mathbf{a}_{ij}^-$ are subsets of $\mathbf{a}_{ij}$ corresponding to $\mathbf{D}^+$ and $\mathbf{D}^-$, respectively. Since $\mathbf{D}$ is a super set of $\mathbf{D}^-$, theoretically the reconstruction error of $\mathbf{x}_{ij}$ using $\mathbf{D}$ (the numerator) should be always smaller than that using $\mathbf{D}^-$ (the denominator). Specifically, solving the sparse representation $\mathbf{a}$ given a dictionary set $\mathbf{D}$ is modeled as a Lasso problem [41, 42] shown in Eq. (15):

$$\mathbf{a}^* = \arg\min \frac{1}{2}\|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_1, \tag{15}$$

where λ is a scaling vector to control the sparsity of $\mathbf{a}$. The solving of l_1 regularized least squares have been investigated extensively in the literature [43–45]. Here we adopt the iterative shrinkage-thresholding algorithm (ISTA) [46, 47] for solving the sparse codes $\mathbf{a}$.

The definition of $\Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i, L_i = 1)$ in (13) indicates that if a point $\mathbf{x}_{ij} \in \mathbf{B}_i, L_i = 1$, is a true positive point, it should not be well represented by only the non-target concepts, so the residual error approximated by the entire concepts, $\|\mathbf{r}_{ij}\|^2$, will be much smaller than that by the background concepts $\|\mathbf{q}_{ij}\|^2$, thus $\Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i, L_i = 1) = \exp\left(-\beta\frac{\|\mathbf{r}_{ij}\|^2}{\|\mathbf{q}_{ij}\|^2}\right) \to 1$. Otherwise, if $\mathbf{x}_{ij} \in \mathbf{B}_i, L_i = 1$ is a false positive point, $\|\mathbf{r}_{ij}\|^2 \approx \|\mathbf{q}_{ij}\|^2$, thus $\Lambda(\mathbf{x}_{ij}, \mathbf{D}|\mathbf{B}_i, L_i = 1) = \exp\left(-\beta\frac{\|\mathbf{r}_{ij}\|^2}{\|\mathbf{q}_{ij}\|^2}\right) \to 0$.

For points from negative bags, following Eq. (12), we model the reconstruction error of points $\mathbf{x}_{ij} \in \mathbf{B}_i, L_i = 0$ as a zero mean Gaussian distribution, shown as Eq. (16),

$$\Pr(l_{ij} = -|\mathbf{B}_i, L_i = 0) = \exp\left(-\|\mathbf{x}_{ij} - \mathbf{D}^-\mathbf{p}_{ij}\|^2\right), \tag{16}$$

where $\mathbf{p}_{ij}$ is the sparse representation of $\mathbf{x}_{ij}$ given $\mathbf{D}^-$ and is solved by Eq. (15). Here instead of applying the hybrid detector, we use a least squares to represent the residual error of $\mathbf{x}_{ij}$. This indicates that the negative points should be fully represented by just the non-target concepts, $\mathbf{D}^-$. The intuitive understanding of this assumption is that minimizing the least squares of all of the negative points provides a good description of the background. Moreover, because there are typically many more negative points in negatively labeled bags than true positive points in positive bags, the target concept estimation may be biased if the hybrid detector was also applied to negative instances.

Thus far, we have constructed an objective function that learns a set of concepts that maximize the hybrid sub-pixel detector statistics of the positive bags and characterize the negative bags. However, given the objective function so far, there is no guarantee that the estimated target concept captures only the discriminative features of the positive class and is discriminative from the negative class. Inspired by the discriminative terms proposed by the Dictionary Learning with Structured Incoherence [48] and the Fisher Discrimination Dictionary Learning (FDDL) algorithm [49, 50], we propose a cross incoherence term $Q(\mathcal{X}, \mathbf{D}^{+}, \mathcal{A})$ shown in Eq. 17 to complete the objective, where $\mathcal{X}$ is the concatenation of all instances from negatively labeled bags, $\mathbf{D}^{+}$ is the target concept set which is the subset of $\mathbf{D} = [\mathbf{D}^{+}\ \mathbf{D}^{-}]$, $\mathcal{A} = [\mathcal{A}^{+}\ \mathcal{A}^{-}]$ is the sparse codes matrix of $\mathcal{X}$ with respect to the entire concepts $\mathbf{D}$.

$$\begin{aligned} Q(\mathcal{X}, \mathbf{D}^{+}, \mathcal{A}) &= \frac{\alpha}{2} \| Diag\left((\mathbf{D}^{+}\mathcal{A}^{+})^{T} \mathcal{X} \right) \|_2^2 \\ &= \frac{\alpha}{2} \sum_{i=K^{+}+1}^{K} \sum_{j=1}^{N_i} \left((\mathbf{D}^{+}\mathbf{a}_{ij}^{+})^{T} \mathbf{x}_{ij} \right)^2 \end{aligned} \quad (17)$$

The understanding of the proposed cross incoherence term is presented by examining the reconstruction of the negative data set $\mathcal{X}$. First of all, $\mathcal{X}$ should be well represented by the non-target concept set $\mathbf{D}^{-}$, *i.e.*, $\mathcal{X} \approx \mathbf{D}^{-}\mathbf{P}$. This is fulfilled by inclusion of the term in Eq. (16). Second, since $\mathbf{D} = [\mathbf{D}^{+}\ \mathbf{D}^{-}]$ is a superset of $\mathbf{D}^{-}$, the reconstruction error of $\mathcal{X}$ by the entire concept set $\mathbf{D}$ is also small, *i.e.*, $\mathcal{X} \approx \mathbf{D}^{+}\mathcal{A}^{+} + \mathbf{D}^{-}\mathcal{A}^{-} = \mathbf{R}^{+} + \mathbf{R}^{-}$. In order to have a target concept $\mathbf{D}^{+}$ that is distinct from the negative data, it is expected that the reconstruction of $\mathcal{X}$ with respect to the target concept, $\mathbf{R}^{+}$, should either maintain small energy or else have a bad representation of $\mathcal{X}$, and thus Eq. 17 is optimized.

The final objective function is shown in Eq. 18, which contains three terms: generalized mean (GM) term (first), background data fidelity term (second) and the cross incoherence (discriminative) term (third):

$$\begin{aligned} J_3 = & -\sum_{i=1}^{K^{+}} \frac{1}{b} \ln \left(\frac{1}{N_i} \sum_{j=1}^{N_i} \exp\left(-\beta \frac{\|\mathbf{x}_{ij} - \mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij} - \mathbf{D}^{-}\mathbf{p}_{ij}\|^2} \right)^{b} \right) \\ & + \rho \sum_{i=K^{+}+1}^{K} \sum_{j=1}^{N_i} \|\mathbf{x}_{ij} - \mathbf{D}^{-}\mathbf{p}_{ij}\|^2 \\ & + \frac{\alpha}{2} \sum_{i=K^{+}+1}^{K} \sum_{j=1}^{N_i} \left((\mathbf{D}^{+}\mathbf{a}_{ij}^{+})^{T} \mathbf{x}_{ij} \right)^2, \end{aligned} \quad (18)$$

IV. Optimization

The optimization of Eq. (18) can be decomposed into two sub-problems, updating the concepts $\mathbf{D}$ and the sparse representation $\mathbf{a}$ alternatively. The algorithm stops when the number of preset iterations is reached or the change between two iterations is smaller than a preset threshold. The method is summarized in Alg. 1. For readability, the derivation of update equations are described in Sec. IX.

Algorithm 1 MI-HE algorithm

Initialize $\mathbf{D}^{0}$, $iter = 0$
repeat
  for $t = 1, \cdots, T$ **do**
    Solve $\mathbf{a}_{ij}$, $\mathbf{p}_{ij}$ according (34), $\forall i \in \{1, \cdots, K\}, j \in \{1, \cdots, N_i\}$
    Update $\mathbf{d}_t$ using gradient descent according to (27)
    $\mathbf{d}_t \leftarrow \frac{1}{\|\mathbf{d}_t\|_2}\mathbf{d}_t$
  end for
  for $k = T+1, \cdots, T+M$ **do**
    Solve $\mathbf{a}_{ij}$, $\mathbf{p}_{ij}$ according (34), $\forall i \in \{1, \cdots, K\}, j \in \{1, \cdots, N_i\}$
    Update $\mathbf{d}_k$ using gradient descent according to (28)
    $\mathbf{d}_k \leftarrow \frac{1}{\|\mathbf{d}_k\|_2}\mathbf{d}_k$
  end for
  $iter \leftarrow iter + 1$
until Stopping criterion reached
return $\mathbf{D}$

V. Experimental Results

In the following, MI-HE is evaluated and compared to several MIL concept learning methods on simulated data and to a real hyperspectral target detection data set. The simulated data experiments are included to illustrate the properties of MI-HE and provide insight into how and when the methods are effective.

A. Simulated Data

As we discussed before, eFUMI combines all positive bags as one big positive bag and all negative bags as one big negative bag and learns target concept from the big positive bag that is different from the negative bag. So if the negative bags maintain incomplete knowledge of the background, *e.g.*, some non-target concept appears only in the subset of positive bags, eFUMI will perform poorly. However, MI-HE which maintains bag structure will be able to distinguish the target.

Given this hypothesis, simulated data was generated from four spectra selected from the ASTER spectral library [51]. Specifically, the Red Slate, Verde Antique, Phyllite and Pyroxenite spectra from the rock class with 211 bands and wavelengths ranging from 0.4μm to 2.5μm (as shown in Fig. 2 in solid lines) were used as endmembers to generate hyperspectral data. Red Slate was labeled as the target endmember.

Four sets of highly-mixed noisy data with varied mean target proportion value ($\boldsymbol{\alpha}_{t\_mean}$) were generated, a detailed generation process can be found in [17]. Specifically, this synthetic data has 15 positive and 5 negative bags with each bag has 500 points. If it is a positively labeled bag, there are 200 highly-mixed target points containing mean target (Red Slate) proportion from to 0.1 to 0.7 respectively to vary the level of target presence from weak to high. Gaussian white noise was added so that signal-to-noise ratio of the data was set to $20dB$. To highlight the ability of MI-HE to leverage individual bag-level labels, we use different subsets

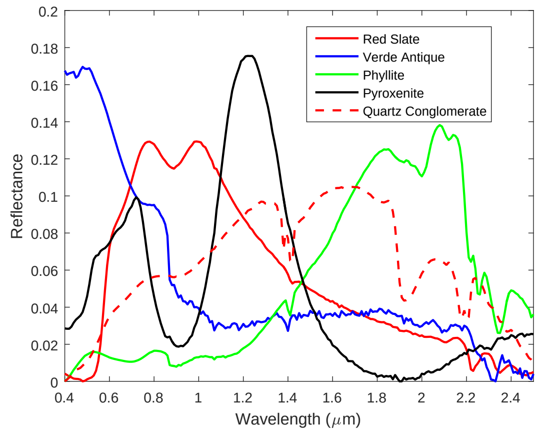
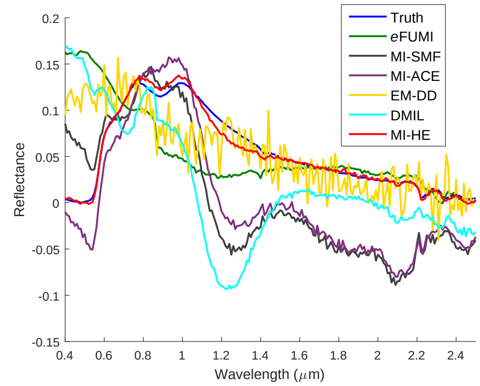


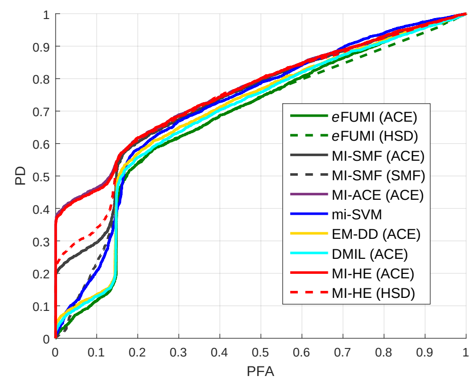
Fig. 2. Signatures from ASTER library used to generate simulated data

TABLE I
LIST OF CONSTITUENT ENDMEMBERS FOR SYNTHETIC DATA WITH INCOMPLETE BACKGROUND KNOWLEDGE

Bag No.	Bag Label	Target Endmember	Background Endmember
1-5	$+$	Red Slate	Verde Antique, Phyllite, Pyroxenite
6-10	$+$	Red Slate	Phyllite, Pyroxenite
11-15	$+$	Red Slate	Pyroxenite
16-20	$-$	N/A	Phyllite, Pyroxenite



(a) Estimated target signatures



(b) ROC curves

Fig. 3. MI-HE and comparisons on synthetic data with incomplete background knowledge, $\boldsymbol{\alpha}_{t\_mean} = 0.1$. MI-SMF and MI-ACE are not expected to recover the true signature.

of background endmembers to build synthetic data as shown in Tab. I. Tab. I shows that the negatively labeled bags only contain 2 negative endmembers and there exists one confusing background endmember in the first 5 positive bags which is Verde Antique. It is expected that the proposed MI-HE will be able to learn the target signature correctly and eFUMI will confuse both Red Slate and Verde Antique as target signatures.

The parameter settings of MI-HE for this experiment are $T = 1, M = 9, \rho = 0.8, b = 5, \beta = 5$ and $\lambda = 1 \times 10^{-3}$. MI-HE was compared to state-of-the-art MIL algorithms eFUMI [17, 37], MI-SMF and MI-ACE [19], DMIL [52, 53], EM-DD [36] and mi-SVM [30]. The mi-SVM algorithm was added to these experiments to include a comparison MIL approach that does not rely on estimating a target signature.

Fig. 3(a) shows the estimated target signature from data with 0.1 mean target proportion value. It clearly shows that the proposed MI-HE is able to correctly distinguish Red Slate as target concept from the incomplete background knowledge. Also, the other comparison algorithms can also estimate a target concept close to the groundtruth Red Slate spectrum. However, eFUMI is always confused with the other non-target endmember, Verde Antique, that exists in some positive bags but is excluded from the background bags.

For simulated detection analysis, estimated target concepts from the training data were then applied to the test data generated separately following the same generating procedure. The detection was performed using the HSD or ACE detection statistic. For MI-HE and eFUMI, both methods were applied since those two algorithms can come out a set of background concept from training simultaneously; for MI-SMF, both SMF and ACE were applied since MI-SMF's objective is maximizing the multiple instance spectral matched filter; for the rest multiple instance target concept learning algorithms, MI-ACE, EM-EE, DMIL, only ACE was applied. For the testing procedure of mi-SVM, a regular SVM testing process was performed using LIBSVM [54], and the decision values (signed distances to hyperplane) of test data determined from trained SVM model were taken as the confidence values. For the signature based detectors, the background data mean and covariance were estimated from the negative instances of the training data.

For quantitative evaluation, Fig. 3(b) show the receiver operating characteristic (ROC) curves using estimated target signature, where it can be seen that the eFUMI is confused with the testing Verde Antique data at very low PFA rate. Tab. II shows the area under the curve (AUC) of proposed MI-HE and comparison algorithms. The results reported are the median results over five runs of the algorithm on the same data. From Tab. II, it can be seen that for MI-HE and MI-ACE, the best performance on detection was achieved using ACE detector. The reason that MI-HE's detection using HSD detector is a little worse is that HSD relies on knowing the complete background concept to properly represent each non-target testing data, the missing non-target concept (Verde Antique) makes the non-target testing data containing Verde

TABLE II
DETECTION STATISTICS (AUCS) FOR SIMULATED HYPERSPECTRAL DATA WITH INCOMPLETE BACKGROUND KNOWLEDGE, BOLD FOR THE BEST, UNDERLINE FOR THE SECOND BEST

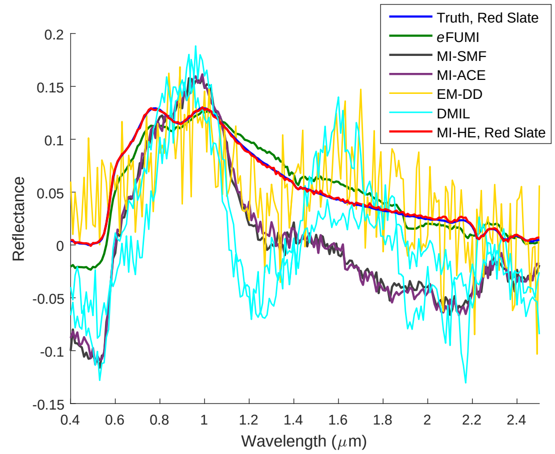
Algorithm	α_{t_mean}			
	0.1	0.3	0.5	0.7
MI-HE (HSD)	0.743	<u>0.931</u>	0.975	0.995
MI-HE (ACE)	<u>0.763</u>	**0.952**	**0.992**	**0.999**
*e*FUMI (ACE)	0.675	0.845	0.978	<u>0.998</u>
*e*FUMI (HSD)	0.671	0.564	0.978	<u>0.998</u>
MI-SMF (SMF)	0.719	0.923	0.972	0.993
MI-SMF (ACE)	0.735	**0.952**	**0.992**	**0.999**
MI-ACE (ACE)	**0.764**	**0.952**	**0.992**	**0.999**
mi-SVM	0.715	0.815	0.866	0.900
EM-DD (ACE)	0.695	0.918	<u>0.983</u>	<u>0.998</u>
DMIL (ACE)	0.687	0.865	0.971	0.996

Antique maintain a large reconstruction error, and thus large detection statistics.

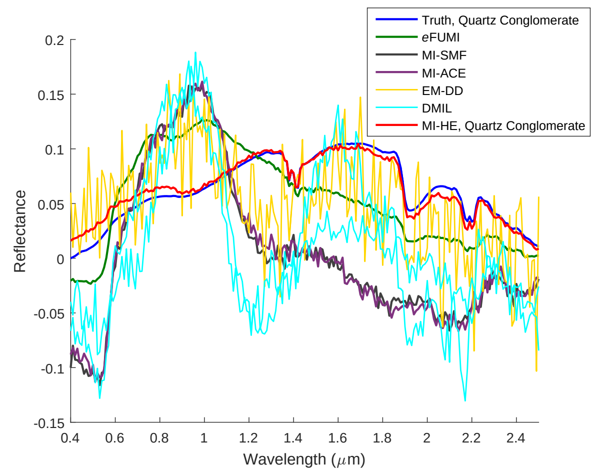
To formulate a multiple concept leaning problem, another rock endmember, Quartz Conglomerate, from ASTER spectral library was used as the second target concept (as shown in Fig. 2 in dashed line). Three sets of highly-mixed noisy data with varied mean target proportion value from [0.1, 0.1] to [0.3, 0.3] were generated. The synthetic data has 5 positive bags containing both target concept (Red Slate and Quartz Conglomerate) and non-target concept (Verde Antique, Phyllite, Pyroxenite); 5 negative bags containing only the background concept (Verde Antique, Phyllite, Pyroxenite). The mean target proportion value was set to [0.1, 0.1], [0.2, 0.2] and [0.3, 0.3], respectively to vary the level of target presence from weak to high. The other settings for simulated data are the same as above mentioned. It is expected that the proposed MI-HE is able to learn multiple target concept at one time.

The parameter settings of MI-HE for this experiment are $T = 2, M = 9, \rho = 0.8, b = 5, \beta = 5$ and $\lambda = 1 \times 10^{-3}$. Fig. 4(a) and 4(b) show the estimated target concepts by proposed MI-HE and comparisons, where we can see that the proposed MI-HE is able to accurately estimate multiple target concepts simultaneously. Compared with MI-HE, although DMIL is also a multiple concept learning algorithm, target concept as estimated by DMIL is noisy, and not a representative prototype of the target class. The remaining comparison algorithms are single target concept learning which are always confused by the multiple target concept problem.

Fig. 5 shows the ROCs using estimated target signature, and Tab. III shows the AUCs of proposed MI-HE and comparison algorithms. The results reported are the median results over five runs of the algorithm on the same data. Since ACE is a single signature based detector, given multiple target concepts, the maximum detection statistics across all estimated target concept was picked for each testing data. From Tab. III, it can be seen that the proposed MI-HE outperforms all the comparison single concept learning algorithms as well as the multiple MI concept learning algorithm DMIL.



(a) Estimated target signatures, Red Slate



(b) Estimated target signatures, QuartzConglomerate

Fig. 4. Estimated target signatures by MI-HE and comparisons on synthetic data with multiple target concepts, $\alpha_{t\_mean} = [0.1, 0.1]$. Not all comparisons algorithms are expected to recover true target signatures.

TABLE III
DETECTION STATISTICS (AUCS) FOR SIMULATED HYPERSPECTRAL DATA WITH MULTIPLE TARGET CONCEPTS, BOLD FOR THE BEST, UNDERLINE FOR THE SECOND BEST

Algorithm	α_{t_mean}		
	[0.1, 01]	[0.2, 0.2]	[0.3, 0.3]
MI-HE (HSD)	**0.875**	**0.982**	**0.998**
MI-HE (ACE)	**0.875**	**0.982**	**0.998**
*e*FUMI (ACE)	<u>0.872</u>	<u>0.980</u>	**0.998**
*e*FUMI (HSD)	0.865	0.976	<u>0.997</u>
MI-SMF (SMF)	0.866	0.977	<u>0.997</u>
MI-SMF (ACE)	0.865	0.976	<u>0.997</u>
MI-ACE (ACE)	0.866	0.976	<u>0.997</u>
mi-SVM	0.711	0.890	0.970
EM-DD (ACE)	0.858	0.979	**0.998**
DMIL (ACE)	0.850	0.971	0.994

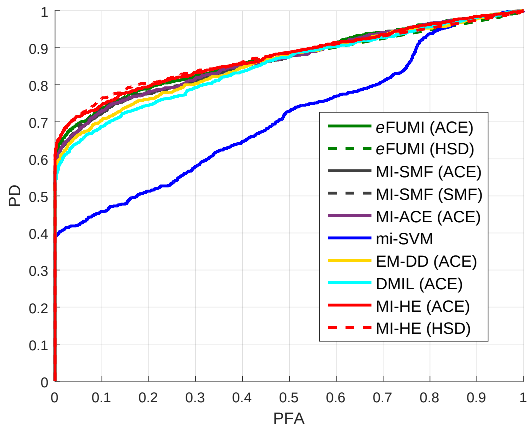


Fig. 5. ROCs of MI-HE and comparisons on synthetic data with multiple target concepts, $\boldsymbol{\alpha}_{t\_mean} = [0.1, 0.1]$.

VI. MUUFL Gulfport Hyperspectral Data

For experiments on real hyperspectral target detection data, the MUUFL Gulfport hyperspectral data set collected over the University of Southern Mississippi-Gulfpark Campus was used. This data set contains 325×337 pixels with 72 spectral bands corresponding to wavelengths from $367.7nm$ to $1043.4nm$ at a $9.5-9.6nm$ spectral sampling interval with spatial resolution 1 pixel/m^2 [14]. The first four and last four bands were removed due to sensor noise. Two sets of this data (Gulfport Campus Flight 1 and Gulfport Campus Flight 3) were selected as cross-validated training and testing data for these two data sets have the same altitude and spatial resolution. Throughout the scene, there are 64 man-made targets in which 57 were considered in this experiment which are cloth panels of four different colors: Brown (15 examples), Dark Green (15 examples), Faux Vineyard Green (FVGr) (12 examples) and Pea Green (15 examples). The spatial location of the targets are shown as scattered points over an RGB image of the scene in Fig. 6. Some of the targets are in the open ground and some are occluded by the live oak trees. Moreover, the targets also vary in size, for each target type, there are targets that are $0.25m^2$, $1m^2$ and $9m^2$ in area, respectively, resulting a very challenging, highly mixed sub-pixel target detection problem.

A. MUUFL Gulfport Hyperspectral Data, Individual Target Type Detection

For this part of the experiments, each individual target type was treated as a target class, respectively. For example, when "Brown" is selected as target class, a 5×5 rectangular region corresponding to each of the 15 ground truth locations denoted by GPS was grouped into a positive bag to account for the drift coming from GPS. This size was chosen based on the accuracy of the GPS device used to record the ground truth locations. The remaining area that does not contain a brown target was grouped into a big negative bag. This constructs the detection problem for "Brown" target. Similarly, there are 15, 12, 15 positive labeled bags for Dark Green, Faux Vineyard Green and Pea Green, respectively. The parameter settings of MI-HE for this experiment are $T=1, M=9, \rho=0.3, b=5, \beta=1$ and $\lambda = 5 \times 10^{-3}$.

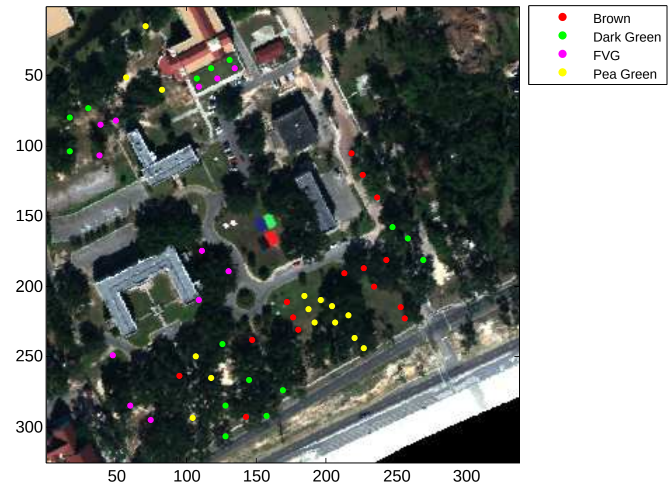


Fig. 6. MUUFL Gulfport data set RGB image and the 579 target locations

MI-HE and comparison algorithms were evaluated on this data using the Normalized Area Under the receiver operating characteristic Curve (NAUC) in which the area was normalized out to a false alarm rate (FAR) of 1×10^{-3} false alarms/m^2 [55]. During detection on the test data, the background mean and covariance were estimated from the negative instances of the training data. The results reported are the median results over five runs of the algorithm on the same data.

Fig. 7(a) shows the estimated target concept by proposed MI-HE and comparisons for Brown target type training on flight 3. We can see that the proposed MI-HE is able to recover the target concept quite close to groundtruth spectra manually selected from the scene. Fig. 7(b) shows the detection ROCs given target spectra estimated on flight 3 and cross validated on flight 1. Tab. IV shows the NAUCs for MI-HE and comparison algorithms cross validated on all four types of target, where it can be seen the proposed MI-HE generally outperforms the comparisons for most of the target types.

B. MUUFL Gulfport Hyperspectral Data, All Four Target Types Detection

For training and detection for the four target types together, the positive bags were generated by grouping each of the 5×5 regions denoted by the groundtruth that it contains any of the four types of target. Thus, for each flight there are 57 target points and 57 positive bags were generated. The remaining area that does not contain any target was grouped into a big negative bag. The parameter settings of MI-HE for this experiment are $T=9, M=11, \rho=0.3, b=5, \beta=1$ and $\lambda = 5 \times 10^{-3}$.

Fig. 8(a) and 8(b) shows the detection ROCs given target spectra estimated one flight and cross validated on another flight, which show that the detection statistics by proposed MI-HE using HSD are significantly better than the comparison algorithms. Tab. V summarizes the NAUCs as a quantitative comparison.

TABLE IV

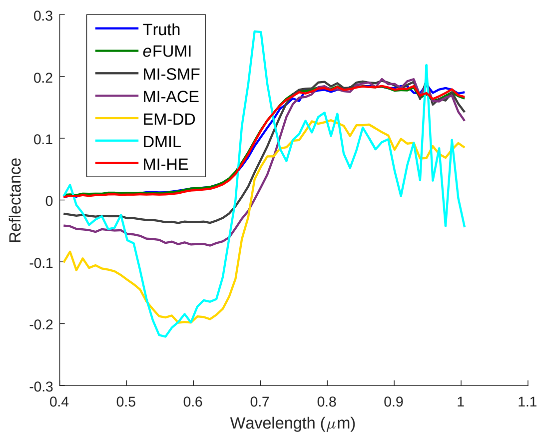
DETECTION STATISTICS (NAUCS) FOR GULFPORT DATA WITH INDIVIDUAL TARGET TYPE, BOLD FOR THE BEST, UNDERLINE FOR THE SECOND BEST

Alg.	Train on Flight 1; Test on Flight 3				Train on Flight 3; Test on Flight 1			
	Brown	Dark Gr.	Faux Vine Gr.	Pea Gr.	Brown	Dark Gr.	Faux Vine Gr.	Pea Gr.
MI-HE (HSD)	**0.499**	**0.453**	**0.655**	0.267	**0.781**	**0.532**	**0.655**	0.350
MI-HE (ACE)	0.433	0.379	0.104	0.267	0.710	0.360	0.111	0.266
***e*FUMI (ACE)**	0.423	0.377	0.654	0.267	0.754	0.491	0.605	0.393
***e*FUMI (HSD)**	0.444	0.436	0.653	0.267	0.727	0.509	0.500	0.333
MI-SMF (SMF)	0.419	0.354	0.533	0.266	0.657	0.405	0.650	0.384
MI-SMF (ACE)	0.448	0.382	0.579	0.316	0.760	0.501	0.613	0.388
MI-ACE(ACE)	0.474	0.390	0.485	**0.333**	0.760	0.483	0.593	0.380
mi-svm	0.206	0.195	0.412	0.265	0.333	0.319	0.245	0.274
EM-DD(ACE)	0.411	0.381	0.486	0.279	0.760	0.503	0.541	**0.416**
DMIL(ACE)	0.419	0.383	0.191	0.009	0.743	0.310	0.081	0.083

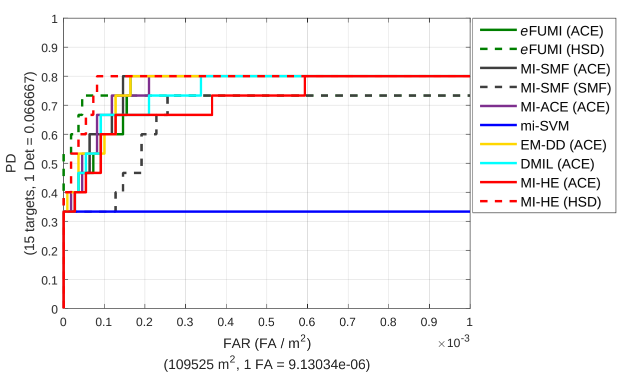
TABLE V

DETECTION STATISTICS (NAUCS) FOR GULFPORT DATA WITH ALL FOUR TARGET TYPES, BOLD FOR THE BEST, UNDERLINE FOR THE SECOND BEST

Alg.	Tr. Fl. 1; Te. Fl. 3	Tr. Fl. 3; Te. Fl. 1	Alg.	Tr. Fl. 1; Te. Fl. 3	Tr. Fl. 3; Te. Fl. 1
MI-HE (HSD)	**0.304**	**0.449**	**MI-SMF(ACE)**	0.219	0.327
M-IHE (ACE)	0.257	0.254	**MI-SMF(SMF)**	0.198	0.277
***e*FUMI (ACE)**	0.214	0.325	**mi-SVM**	0.235	0.269
***e*FUMI (HSD)**	0.256	0.331	**EM-DD(ACE)**	0.211	0.310
MI-ACE (ACE)	0.226	0.340	**DMIL(ACE)**	0.198	0.225

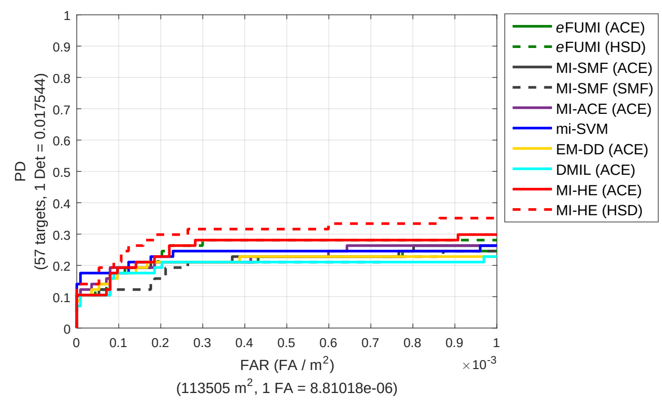


(a) Estimated target signatures

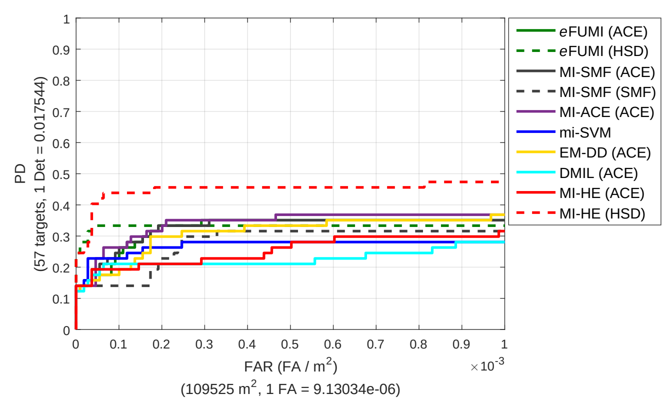


(b) ROC curves

Fig. 7. MI-HE and comparisons on Gulfport Data Brown, training flight 3 testing flight 1



(a) Train on flight 1, detect on flight 3



(b) Train on flight 3, detect on flight 1

Fig. 8. ROCs of MI-HE and comparisons on Gulfport Data all types detection

VII. Analysis of MI-HE Parameter Settings on Simulated Data

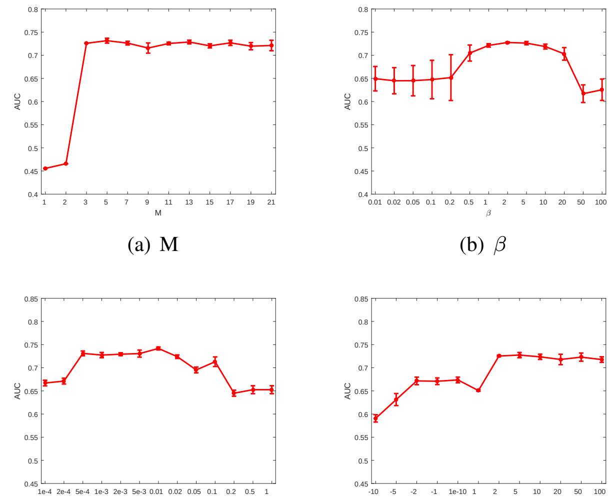
In order to provide deeper insights into the sensitivity of MI-HE performance relative to variations in input parameters, we tested MI-HE on simulated hyperspectral data across a range of parameter values. Specifically, the varying parameters and ranges examined are shown in Tab. VI. Red Slate was again selected as target endmember and the other three, Verde Antique, Phyllite and Pyroxenite were used as non-target endmembers as shown in Fig. 2. Tab. VII shows the bags labeling and constituent endmembers. The synthetic data has $K^+ = 5$ positive and $K^- = 5$ negative bags with each bag containing 100 points. If it is a positively labeled bag, there are 50 highly-mixed target points containing mean target (Red Slate) proportion $\boldsymbol{\alpha}_{t\_mean} = 0.1$ and are mixed with at least one randomly picked background endmember. Gaussian white noise was added so that the signal-to-noise ratio of the data was set to $20dB$. The reference parameters for MI-HE were set to $T = 1, M = 7, \rho = 0.8, b = 5, \beta = 5$ and $\lambda = 1\times 10^{-3}$.

TABLE VI
Testing Ranges for Analysis of MI-HE Parameter Settings

Type	Range
M	$1, 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21$
β	$0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100$
λ	$1\times10^{-4}, 2\times10^{-4}, 5\times10^{-4}, 1\times10^{-3}, 2\times10^{-3}$, $5\times10^{-3}, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1$
b	$-10, -5, -2, -1, 1\times10^{-10}, 1, 2, 5, 10, 20, 50, 100$

TABLE VII
List of Constituent Endmembers for Synthetic Data

Bag No.	Bag Label	Constituent Endmembers
1-5	$+$	Red Slate, Verde Antique, Phyllite, Pyroxenite
6-10	$-$	Verde Antique Phyllite, Pyroxenite



(a) M (b) β (c) λ (d) b

Fig. 9. Detection statistics (AUCs) of MI-HE plots with different parameter settings

Fig. 9 shows the detection performance (mean AUCs and variance over five runs) of MI-HE for the sensitivity analysis on the simulated data described above. Several interesting inferences can be drawn from Fig. 9 regarding how MI-HE responds to its parameter settings. For the setting of M, it can be seen from Fig. 9(a) that MI-HE performs consistently when the number of background concepts is set to greater than or equal to 3, which is the true number of background concepts of the simulated data. Since MI-HE adopts a linear mixing model instead of convex mixing for the data mixture, the conclusion is MI-HE is not sensitive to the setting of M given M is not set smaller than the necessary constituent background concepts of input data.

For the setting of β, Fig. 9(b) shows MI-HE performs well with β in the range [1, 10] on this data. Since β is a scaling factor for the hybrid detector, $\Pr(\mathbf{x}_{ij}|\mathbf{D}, \mathbf{B}_i^+) = \exp\left(-\beta\frac{\|\mathbf{x}_{ij}-\mathbf{D}\boldsymbol{a}_{ij}^+\|^2}{\|\mathbf{x}_{ij}-\mathbf{D}^-\boldsymbol{p}_{ij}\|^2}\right)$, the scaling effect controls the gradient of the objective function (18). Since the range $\beta \in$ [1, 10] provides a moderate gradient for the exponential function, so it can be concluded that MI-HE requires β to be set to a suitable range, *e.g.*, [1, 10].

The setting of sparsity level, λ, results in the step length of soft shrinkage. This value should be related with the magnitude of the input data. For the simulated data tested here, the proportion values were generated from the Dirichlet distribution within range [0, 1]. As shown by Fig. 9(c), the appropriate range of λ for this simulated data is $[5\times10^{-4}, 0.02]$ which is reasonable. However, in general, prior knowledge is needed for setting specific λ for a real dataset. Currently we set λ to be approximately the $1/1000$ of the l_2-norm mean of the training data.

The generalized mean model is controlled by the parameter b to realize the minimum operation ($b \to -\infty$), the average operation ($b \in (0, 1]$) or the maximum operation ($b \to +\infty$). Since the proposed model aims to predict the true positive instance from positive bags and assumes the "soft maximum" operation for this generalized mean model, it is expected that the model will work well with b greater than 1. Fig. 9(d) verifies this hypothesis showing that the algorithm works well for b great than 1. For all experiments shown in this paper, the parameter b was set to 5 and observed to work well.

Although there are several parameters for MI-HE, these parameters come from the models assumed to underlie MI-HE. From this sensitivity analysis of MI-HE with different parameter settings, it can be concluded that there is a general range of model stability for each parameter. Moreover, the above analysis provides a intuitive understanding and heuristic approach for setting the parameters of MI-HE to stable ranges.

VIII. Conclusion

In this work the MI-HE target concept learning framework for MIL problems is proposed and investigated. MI-HE is able to learn multiple discriminative target concepts from ambiguously labeled data. After learning target concepts, target detection can be conducted by applying the estimated target concepts to any signature based detector. Comprehensive experiments show that the proposed MI-HE is effective in learning discriminative target concept and achieves superior performance over comparison algorithms in several scenarios.

IX. Derivation of MI-HE update equations

Similar to the Dictionary Learning using Singular Value Decomposition (K-SVD) approach [56], the optimization of target and background concepts is performed by taking gradient descent with respect to one atom at a time and holding the rest fixed. Denote f_{GM} as the generalized mean part of Eq. (18):

$$\begin{aligned} f_{GM} &= -\sum_{i=1}^{K^+}\frac{1}{b}\ln\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^b\right) \\ &= -\sum_{i=1}^{K^+}\frac{1}{b}\ln\frac{1}{N_i}-\sum_{i=1}^{K^+}\frac{1}{b}\ln\left(\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^b\right) \end{aligned} \quad (19)$$

Removing the constant part in f_{GM} and taking the partial derivative with respect to $\mathbf{d}$ results in (20):

$$\begin{aligned} \frac{\partial f_{GM}}{\partial \mathbf{d}} &= -\sum_{i=1}^{K^+}\frac{1}{\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^b} \\ &\cdot\left(\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^{b-1}\cdot\frac{\partial\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)}{\partial\mathbf{d}}\right), \end{aligned} \quad (20)$$

where $\mathbf{d}$ is a symbolic notation for any atom in $\mathbf{D}$.

Then take the partial derivative on the fidelity (second) term of the objective function Eq. (18) with respect to the background concept, $\mathbf{d}_k$:

$$\begin{aligned} &\frac{\partial -\rho\sum_{i=K^++1}^{K}\sum_{j=1}^{N_i}\ln\Pr(l_{ij}=-|\mathbf{B}_i)}{\partial\mathbf{d}_k} \\ &= \frac{\partial\rho\sum_{i=K^++1}^{K}\sum_{j=1}^{N_i}\|\mathbf{x}_{ij}-\mathbf{D}^-\mathbf{p}_{ij}\|^2}{\partial\mathbf{d}_k} \\ &= \rho\sum_{i=K^++1}^{K}\sum_{j=1}^{N_i}-2p_{ijk}(\mathbf{x}_{ij}-\mathbf{D}^-\mathbf{p}_{ij}) \\ &= \rho\sum_{i=K^++1}^{K}\sum_{j=1}^{N_i}-2p_{ijk}\mathbf{q}_{ij}, \end{aligned} \quad (21)$$

where p_{ijk} is the kth element in $\mathbf{p}_{ij}$ corresponding to $\mathbf{d}_k$.

The partial derivative of the cross incoherence (third) term corresponding to $\mathbf{d}_t$ is:

$$\frac{\partial Q(\mathcal{X},\mathbf{D}^+,\mathcal{A})}{\partial\mathbf{d}_t} = \alpha\sum_{i=K^++1}^{K}\sum_{j=1}^{N_i}(\mathbf{D}^+\mathbf{a}_{ij}^+)^T\mathbf{x}_{ij}\cdot a_{ijt}^+\mathbf{x}_{ij} \quad (22)$$

The partial derivatives of the negative objective function Eq. (18) with respect to $\mathbf{d}_t$ and $\mathbf{d}_k$ are shown in (23) and (24).

$$\begin{aligned} \frac{\partial J_3}{\partial\mathbf{d}_t} &= -\sum_{i=1}^{K^+}\frac{1}{\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^b} \\ &\cdot\left(\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^{b-1}\cdot\frac{\partial\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)}{\partial\mathbf{d}_t}\right) \\ &+\alpha\sum_{i=K^++1}^{K}\sum_{j=1}^{N_i}(\mathbf{D}^+\mathbf{a}_{ij}^+)^T\mathbf{x}_{ij}\cdot a_{ijt}^+\mathbf{x}_{ij} \end{aligned} \quad (23)$$

$$\begin{aligned} \frac{\partial J_3}{\partial\mathbf{d}_k} &= -\sum_{i=1}^{K^+}\frac{1}{\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^b} \\ &\cdot\left(\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^{b-1}\cdot\frac{\partial\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)}{\partial\mathbf{d}_k}\right) \\ &+\rho\sum_{i=K^++1}^{K}\sum_{j=1}^{N_i}-2p_{ijk}\mathbf{q}_{ij} \end{aligned} \quad (24)$$

The next step is taking the partial derivative of the hybrid detector in (13) with respect to $\mathbf{d}_t$ and $\mathbf{d}_k$ shown as Eq. (25) and (26) respectively:

$$\begin{aligned} &\frac{\partial\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)}{\partial\mathbf{d}_t} \\ &= \exp\left(-\beta\frac{\|\mathbf{x}_{ij}-\mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij}-\mathbf{D}^-\mathbf{p}_{ij}\|^2}\right)\frac{\partial\left(-\beta\frac{\|\mathbf{x}_{ij}-\mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij}-\mathbf{D}^-\mathbf{p}_{ij}\|^2}\right)}{\partial\mathbf{d}_t} \\ &= \Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)\frac{2\beta a_{ijt}^+(\mathbf{x}_{ij}-\mathbf{D}\mathbf{a}_{ij})}{\|\mathbf{x}_{ij}-\mathbf{D}^-\mathbf{p}_{ij}\|^2} \\ &= \Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)\frac{2\beta a_{ijt}^+\mathbf{r}_{ij}}{\|\mathbf{q}_{ij}\|^2} \end{aligned} \quad (25)$$

$$\begin{aligned} &\frac{\partial\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)}{\partial\mathbf{d}_k} \\ &= \exp\left(-\beta\frac{\|\mathbf{x}_{ij}-\mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij}-\mathbf{D}^-\mathbf{p}_{ij}\|^2}\right)\frac{\partial\left(-\beta\frac{\|\mathbf{x}_{ij}-\mathbf{D}\mathbf{a}_{ij}\|^2}{\|\mathbf{x}_{ij}-\mathbf{D}^-\mathbf{p}_{ij}\|^2}\right)}{\mathbf{d}_k} \\ &= \Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)\frac{2\beta a_{ijk}^-\mathbf{r}_{ij}\|\mathbf{q}_{ij}\|^2-2\beta p_{ijk}\|\mathbf{r}_{ij}\|^2\mathbf{q}_{ij}}{\|\mathbf{q}_{ij}\|^4} \end{aligned} \quad (26)$$

Substituting the gradient of hybrid detector with respect to $\mathbf{d}_t$ and $\mathbf{d}_k$ in Eq. (25) and (26) to Eq. (23) and (24), respectively, we can get the resultant gradient of the objective function (18) over $\mathbf{d}_t$ and $\mathbf{d}_k$:

$$\begin{aligned} \triangle\mathbf{d}_t &= -\sum_{i=1}^{K^+}\frac{1}{\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^b} \\ &\left(\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^b\cdot\frac{2\beta a_{ijt}^+\mathbf{r}_{ij}}{\|\mathbf{q}_{ij}\|^2}\right) \\ &+\alpha\sum_{i=K^++1}^{K}\sum_{j=1}^{N_i}(\mathbf{D}^+\mathbf{a}_{ij}^+)^T\mathbf{x}_{ij}\cdot a_{ijt}^+\mathbf{x}_{ij} \end{aligned} \quad (27)$$

$$\begin{aligned} \triangle\mathbf{d}_k &= -\sum_{i=1}^{K^+}\frac{1}{\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^b} \\ &\left(\sum_{j=1}^{N_i}\Lambda(\mathbf{x}_{ij},\mathbf{D}|\mathbf{B}_i)^b\cdot 2\beta\frac{a_{ijk}^-\mathbf{r}_{ij}\|\mathbf{q}_{ij}\|^2-p_{ijk}\|\mathbf{r}_{ij}\|^2\mathbf{q}_{ij}}{\|\mathbf{q}_{ij}\|^4}\right) \\ &-\rho\sum_{i=K^++1}^{K}\sum_{j=1}^{N_i}2p_{ijk}\mathbf{q}_{ij} \end{aligned} \quad (28)$$

The optimization of sparse representation can be viewed as a l_1 regularized least squares problem, also known as lasso problem [41, 42], denoted as $\mathcal{L}$. The lasso problem is shown in Eq. (15), where given concept (or dictionary) set $\mathbf{D}$ and preset sparsity level λ, $\mathbf{a}^*$ is the optimal sparse representation of the input data $\mathbf{x}$. Here we adopt the iterative shrinkage-thresholding algorithm (ISTA) [46, 47] for solving for the sparse codes $\mathbf{a}$.

The gradient of (15) with respect to $\mathbf{a}$ without considering the l_1 penalty term is:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{a}} = -\mathbf{D}^T \left(\mathbf{x} - \mathbf{D}\mathbf{a}\right). \tag{29}$$

Then $\mathbf{a}$ at q^{th} iteration can be updated using gradient descent shown in (30):

$$\mathbf{a}^q = \mathbf{a}^{q-1} - \delta \frac{\partial \mathcal{L}}{\partial \mathbf{a}}, \tag{30}$$

followed by a soft-thresholding step:

$$\mathbf{a}^* = S_\lambda \left(\mathbf{a}^q\right), \tag{31}$$

where S_λ: $\mathbb{R}^n \to \mathbb{R}^n$ is the *soft-thresholding* operator defined by

$$S_\lambda \left(\mathbf{a}[k]\right) = sign(\mathbf{a}[k]) \max(|\mathbf{a}[k]| - \lambda, 0), k = 1, \cdots, n \tag{32}$$

Following a similar proof to that in [57], when the step length δ satisfies (33), the update of $\mathbf{a}$ using a gradient descent method with step length η monotonically decreases the value of the objective function, where $Eig_{max}(\mathbf{D^T D})$ denotes the maximum eigenvalue of $\mathbf{D}^T\mathbf{D}$. For simplicity, δ was set to $Eig_{max}\left(\mathbf{D}^T\mathbf{D}\right)$ for all input data:

$$\delta \in \left(0, \frac{1}{Eig_{max}\left(\mathbf{D}^T\mathbf{D}\right)}\right) \tag{33}$$

Finally the resultant update equation for the sparse representation of instance $\mathbf{x}$ given concept set $\mathbf{D}$ is:

$$\mathbf{a}^* = S_\lambda \left(\mathbf{a}^q + \frac{1}{Eig_{max}\left(\mathbf{D}^{-T}\mathbf{D}\right)} \left(\mathbf{D}^T(\mathbf{x} - \mathbf{D}\mathbf{a}^q)\right)\right) \tag{34}$$

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1350078 - CAREER: Supervised Learning for Incomplete and Uncertain Data.

References

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, 2002.

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

[3] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Applied Earth Observations Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[4] S. E. Yuksel, J. Bolton, and P. Gader, "Multiple-instance hidden markov models with applications to landmine detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6766–6775, 2015.

[5] A. Zare, J. Bolton, P. Gader, and M. Schatten, "Vegetation mapping for landmine detection using long-wave hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 172–178, Jan 2008.

[6] G. Mahajan, R. Sahoo, R. Pandey, V. Gupta, and D. Kumar, "Using hyperspectral remote sensing techniques to monitor nitrogen, phosphorus, sulphur and potassium in wheat (triticum aestivum l.)," *Precision Agriculture*, vol. 15, no. 5, pp. 499–522, 2014.

[7] Z. Wang, L. Lan, and S. Vucetic, "Mixture model for multiple instance regression and applications in remote sensing," *IIEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2226–2237, 2012.

[8] R. Pike, G. Lu, D. Wang *et al.*, "A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 653–663, 2016.

[9] A. Pardo, E. Real, V. Krishnaswamy *et al.*, "Directional kernel density estimation for classification of breast tissue spectra," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 64–73, 2017.

[10] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proc. IEEE*, vol. 97, no. 6, pp. 1031–1055, 2009.

[11] M. Lara, L. Lleó, B. Diezma-Iglesias, J.-M. Roger, and M. Ruiz-Altisent, "Monitoring spinach shelf-life with hyperspectral image through packaging films," *J. Food Eng.*, vol. 119, no. 2, pp. 353–361, 2013.

[12] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, 2002.

[13] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, 2014.

[14] P. Gader, A. Zare *et al.*, "MUUFL gulfport hyperspectral and lidar airborne data set," University of Florida, Gainesville, FL, REP-2013-570, Tech. Rep., Oct. 2013.

[15] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial Intell.*, vol. 89, no. 1, pp. 31–71, 1997.

[16] O. Maron and T. Lozano-Perez, "A framework for multiple-instance learning." in *Advances Neural Inf. Process. Syst. (NIPS)*, vol. 10, 1998, pp. 570–576.

[17] C. Jiao and A. Zare, "Functions of multiple instances for learning target signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4670 – 4686, 2015.

[18] ——, "Multiple instance dictionary learning using functions of multiple instances," in *Int. Conf. Pattern Recognition (ICPR)*, 2016, pp. 2688–2693.

[19] A. Zare, C. Jiao, and T. Glenn, "Discriminative multiple instance hyperspectral target characterization," *IEEE Trans. Pattern Anal. Mach. Intell.*, In Press.

[20] C. Jiao and A. Zare, "Multiple instance hybrid estimator for learning target signatures," in *Proc. IEEE Intl. Geosci. Remote Sens. Symp. (IGARSS)*, 2017, pp. 1–4.

[21] S. M. Kay, *Fundamental of Statistical Signal Processing: Volume II – Detection Theory*. Prentice-Hall, 1993.

[22] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory J.*, vol. 14, no. 1, pp. 79–116, 2003.
[23] D. Manolakis, E. Truslow, M. Pieper *et al.*, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 24–33, 2014.
[24] 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, Sept. 1999.
[25] S. Kraut, L. Scharf, and L. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, 2001.
[26] W. F. Basener, "Clutter and anomaly removal for enhanced target detection," in *Proc. SPIE*, vol. 7695, 2010, p. 769525.
[27] J. Broadwater, R. Meth, and R. Chellappa, "A hybrid algorithm for subpixel detection in hyperspectral imagery," in *Proc. IEEE Intl. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 3, 2004, pp. 1601–1604.
[28] J. Broadwater and R. Chellappa, "Hybrid detectors for subpixel targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1891–1903, Nov. 2007.
[29] D. C. Heinz *et al.*, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, 2001.
[30] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," in *Advances Neural Inf. Process. Syst. (NIPS)*, 2002, pp. 561–568.
[31] Y. Chen, J. Bi, and J. Z. Wang, "MILES: Multiple-instance learning via embedded instance selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1931–1947, 2006.
[32] J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani, "1-norm support vector machines," in *Advances Neural Inf. Process. Syst. (NIPS)*, vol. 16, no. 1, 2004, pp. 49–56.
[33] X. Wang, B. Wang, X. Bai, W. Liu, and Z. Tu, "Max-margin multiple-instance dictionary learning," in *Int. Conf. Mach. Learning (ICML)*, 2013, pp. 846–854.
[34] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.
[35] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, vol. 2, 2006, pp. 2169–2178.
[36] Q. Zhang and S. A. Goldman, "EM-DD: An improved multiple-instance learning technique," in *Advances Neural Inf. Process. Syst. (NIPS)*, vol. 2, 2002, pp. 1073–1080.
[37] A. Zare and C. Jiao, "Extended functions of multiple instances for target characterization," in *IEEE Workshop Hyperspectral Image Signal Process.: Evolution in Remote Sens. (WHISPERS)*, 2014, pp. 1–4.
[38] J. Theiler and B. R. Foy, "Effect of signal contamination in matched-filter detection of the signal on a cluttered background," *IEEE Geosci. Remote Sens. Lett*, vol. 3, no. 1, pp. 98–102, Jan 2006.
[39] N. M. Nasrabadi, "Regularized spectral matched filter for target recognition in hyperspectral imagery," *IEEE Signal Process. Lett*, vol. 15, pp. 317–320, 2008.
[40] O. Maron and A. L. Ratan, "Multiple-instance learning for natural scene classification." in *Int. Conf. Mach. Learning (ICML)*, vol. 98, 1998, pp. 341–349.
[41] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Statistical Soc. Series B (Methodological)*, pp. 267–288, 1996.
[42] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.
[43] S. Mallat, *A wavelet tour of signal processing, Third Edition: The Sparse Way*. Academic Press, 2008.
[44] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Mach. Learning*, vol. 4, no. 1, pp. 1–106, 2012.
[45] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Found. Trends Comput. Graphics Vision*, vol. 8, no. 2-3, pp. 85–283, 2014.
[46] M. A. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, 2003.
[47] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Applied Math.*, vol. 57, pp. 1413–1457, 2004.
[48] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, 2010, pp. 3501–3508.
[49] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Int. Conf. Comput. Vision*. IEEE, 2011, pp. 543–550.
[50] ——, "Sparse representation based fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vision*, vol. 109, no. 3, pp. 209–232, 2014.
[51] A. Baldridge, S. Hook, C. Grove, and G. Rivera, "The ASTER spectral library version 2.0," *Remote Sensing of Environment*, vol. 113, no. 4, pp. 711–715, 2009.
[52] A. Shrivastava, V. M. Patel, j. K. Pillai, and R. Chellappa, "Generalized dictionaries for multiple instance learning," *Int. J. of Comput. Vision*, vol. 114, no. 2, pp. 288–305, Septmber 2015.
[53] A. Shrivastava, J. K. Pillai, V. M. Patel, and R. Chellappa, "Dictionary-based multiple instance learning," in *IEEE Int. Conf. Image Process.*, 2014, pp. 160–164.
[54] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technology*, vol. 2, no. 3, pp. 1–27, 2011.
[55] T. Glenn, A. Zare, P. Gader, and D. Dranishnikov, "Bullwinkle: Scoring code for sub-pixel targets (ver-

sion 1.0) [software]," 2013, http://engineers.missouri.edu/zarea/code/.

[56] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[57] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*. Springer Science & Business Media, 2007.