MULTIPLE INSTANCE HYBRID ESTIMATOR FOR LEARNING TARGET SIGNATURES

Changzhe Jiao and Alina Zare

Electrical and Computer Engineering, University of Missouri
Electrical and Computer Engineering, University of Florida

ABSTRACT

Signature-based detectors for hyperspectral target detection rely on knowing the specific target signature in advance. However, target signature are often difficult or impossible to obtain. Furthermore, common methods for obtaining target signatures, such as from laboratory measurements or manual selection from an image scene, usually do not capture the discriminative features of target class. In this paper, an approach for estimating a discriminative target signature from imprecise labels is presented. The proposed approach maximizes the response of the hybrid sub-pixel detector within a multiple instance learning framework and estimates a set of discriminative target signatures. After learning target signatures, any signature based detector can then be applied on test data. Both simulated and real hyperspectral target detection experiments are shown to illustrate the effectiveness of the method.

Index Terms— target detection, concept learning, hyperspectral, multiple instance, target characterization, subpixel.

1. INTRODUCTION

The wealth of spectral information in hyperspectral imagery provides for the ability to perform sub-pixel target detection. Signature-based hyperspectral target detectors need to know the target signature in advance. However, in a number of scenarios, obtaining an effective target signature is often a challenging problem. For instance, in some situations analysts may have approximate locations of sub-pixel targets and want to estimate their target signature and search for the target material elsewhere. In these situations, manual selection of a target signature from an image cube would not only be difficult (as only approximate locations are known) but may result in selecting a mixed (as opposed to pure target) pixels.

In this paper, we address this problem by proposing a method to characterize a target signature from imprecisely labeled hyperspectral imagery. Specifically, we model the hyperspectral target estimation task as a multiple instance concept learning problem. In multiple instance learning (MIL)

[1], training data is partitioned into sets of labeled bags (instead of being individually labeled). A positive bag must contain at least one true positive (target) data point and negative bags are composed entirely of negative data. Multiple instance concept learning is a branch of MIL that aims to learn one or a set of concepts to describe the target class. For example, Diversity Density (DD) [2] identifies a target concept that is close to the intersection of all positive bags and far from negative instances. The expectation maximization version of DD [3] improves convergence. Previous methods for MIL target characterization were proposed in [4, 5]. [4] combines all positive bags into one large positive bag and, thus, discards bag-level label information. [5] does not assume a mixing model and, thus, may fail to take advantage of the mixing model if it is known. In contrast, the proposed method maximizes the response of the hybrid sub-pixel detector, preserves bag-level labels, and assumes the linear mixing model [6].

2. 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 is grouped into K *bags*, $\mathbf{B} = \{\mathbf{B}_1, \ldots, \mathbf{B}_K\}$, with associated binary bag-level labels, $L = \{L_1, \ldots, L_K\}$ where $L_i \in \{0, 1\}$; n_i is the number of instances in bag $\mathbf{B}_i$ and $\mathbf{x}_{ij} \in \mathbf{B}_i$ denotes the jth instance in bag $\mathbf{B}_i$ with instance-level label $l_{ij} \in \{0, 1\}$. When identifying the label on certain bag or instance is important, positive bags will be indicated as $\mathbf{B}_i^+$ with associated bag level label L_i^+ containing instances $\mathbf{x}_{ij}^+$ with instance-level labels $l_{ij}^+, j = 1, \cdots, n_i^+$, where n_i^+ is the number of instances in positive bag $\mathbf{B}_i^+$. Similarly, $\mathbf{B}_i^-$, L_i^-, $\mathbf{x}_{ij}^-$, l_{ij}^- and n_i^- represent a negative bag, label for this negative bag, the jth instance in this bag, label for the jth instance and total number of instances in this bag. The number of positive and negative bags are denoted as K^+ and K^-, N^+ and N^- represent the total number of positive and negative instances, respectively. Thus $N = N^+ + N^- = \sum_{i=1}^{K^+} n_i^+ + \sum_{i=1}^{K^-} n_i^-$ and $\mathbf{B}^+$, $\mathbf{B}^-$ represent the union of all positive instances and negative instances, respectively.

Given this notation, the proposed multiple instance hybrid estimator (MI-HE) aims to maximize the probability of

the labels of the bags as in (1). Since, in MIL, each positive bag must contain at least one positive instance, we can then substitute the probability for a bag to be positive with the probability of the most likely point in the bag to be a target, *i.e.*, $\max_{j \in n_i^+} \Pr(\mathbf{x}_{ij}^+ = +|\mathbf{B}_i^+)$. Since each negative bag is assumed to only contain non-target instances, then the probability for a negative bag to be negative can be represented by the joint probability of all instances in this bag to be negative,

$$J_1 = \prod_{i=1}^{K^+} \Pr(L_i^+ = +|\mathbf{B}_i^+) \prod_{i=1}^{K^-} \Pr(L_i^- = -|\mathbf{B}_i^-). \qquad (1)$$

$$= \prod_{i=1}^{K^+} \max_{j \in n_i^+} \Pr(l_{ij}^+ = +|\mathbf{B}_i^+) \prod_{i=1}^{K^-} \prod_{j=1}^{n_i^-} \Pr(l_{ij}^- = -|\mathbf{x}_{ij}^-). (2)$$

Eq. (2) contains a $\max$ operation that is difficult to optimize numerically. Some algorithms in the literature adopt a noisy-OR model instead of a $\max$ [2, 3]. However, experimental results show that the noisy-OR is non-smooth and generally needs to be optimized repeatedly with many different initializations to avoid local optima. In the proposed approach, we adopt the generalized mean as an alternative,

$$J_2 = \prod_{i=1}^{K^+} \left(\frac{1}{n_i^+} \sum_{j=1}^{n_i^+} \Pr(l_{ij}^+ = +|\mathbf{B}_i^+)^p \right)^{\frac{1}{p}} \prod_{i=1}^{K^-} \prod_{j=1}^{n_i^-} \Pr(l_{ij}^- = -|\mathbf{x}_{ij}^-),$$
$$(3)$$

where $p \in [-\infty, +\infty]$ is a parameter that varies the operation from a $\min$ to a $\max$, respectively. We then optimize the negative logarithm of J_2 and add a scaling factor, $\rho < 1$, to the second term to control the influence of the negative bags (when $N^- \neq N^+$) as shown in (4).

We now must define $\Pr(l_{ij}^+ = +|\mathbf{B}_i^+)$ and $\Pr(l_{ij}^- = -|\mathbf{x}_{ij}^-)$. As in [7], each instance is modeled as a sparse linear combination of target and/or background signatures $\mathbf{D}$, $\mathbf{x}_j \approx \mathbf{D}\boldsymbol{\alpha}_j$, where $\boldsymbol{\alpha}_j$ is the sparse vector of weights. Each positive bag contains at least one instance with target:

$$\text{if } L_i = 1, \exists \mathbf{x}_j \in \mathbf{B}_i^+ \text{ s.t.}$$
$$\mathbf{x}_j = \sum_{t=1}^{T} \alpha_{jt} \mathbf{d}_t^+ + \sum_{k=1}^{M} \alpha_{jk} \mathbf{d}_k^- + \varepsilon_j, \alpha_{jt} \neq 0, \quad (5)$$

where ε_j is a noise term. A negatively labeled bag $\mathbf{B}_i^-$ should not contain any target:

$$\text{if } L_i = 0, \forall \mathbf{x}_j \in \mathbf{B}_i^-, \mathbf{x}_j = \sum_{k=1}^{M} \alpha_{jk} \mathbf{d}_k^- + \varepsilon_j. \qquad (6)$$

Given this model, we introduce the hybrid subpixel detector to estimate the probability instances from positive bags are positive. Specifically, the probability for $\mathbf{x}_{ij}^+$ in $\mathbf{B}_i^+$ is a target point is defined as:

$$\Pr(l_{ij}^+ = +|\mathbf{B}_i^+) = \exp\left(-\beta \frac{\|\mathbf{x}_{ij}^+ - \mathbf{D}\boldsymbol{\alpha}_{ij}^+\|^2}{\|\mathbf{x}_{ij}^+ - \mathbf{D}^-\boldsymbol{\alpha}_{ij}^{+b}\|^2} \right), \quad (7)$$

where $\mathbf{D} = \begin{bmatrix} \mathbf{D}^+ & \mathbf{D}^- \end{bmatrix} \in \mathbb{R}^{d \times (T+M)}, \mathbf{D}^+ = \begin{bmatrix} \mathbf{d}_1^+, \cdots, \mathbf{d}_T^+ \end{bmatrix}$ is the set of T target signatures and $\mathbf{D}^- = \begin{bmatrix} \mathbf{d}_1^-, \cdots, \mathbf{d}_M^- \end{bmatrix}$ is the set of M background signatures, β is a scaling parameter; $\boldsymbol{\alpha}_{ij}^+$ and $\boldsymbol{\alpha}_{ij}^{+b}$ are the sparse representation of $\mathbf{x}_{ij}^+$ given entire signatures set $\mathbf{D}$ and background signatures set $\mathbf{D}^-$, respectively. Specifically, solving for a sparse $\boldsymbol{\alpha}$ given a dictionary set $\mathbf{D}$ is modeled as the Lasso problem [8, 9] shown in (8):

$$\hat{\boldsymbol{\alpha}} = \arg\min \frac{1}{2}\|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1, \qquad (8)$$

where λ is a scaling vector to control the sparsity of $\boldsymbol{\alpha}$. Here we adopt the iterative shrinkage-thresholding algorithm (ISTA) [10] for solving the sparse codes $\boldsymbol{\alpha}$.

Algorithm 1 MI-HE algorithm
1: Initialize $\mathbf{D}^0$, $iter = 0$
2: **repeat**
3:   **for** $t = 1, \cdots, T$ **do**
4:     Solving $\{\boldsymbol{\alpha}_i^+\}_{i=1}^N$, $\{\boldsymbol{\alpha}_i^-\}_{i=1}^N$ by ISTA
5:     Update $\mathbf{d}_t$ by optimizing (4) using gradient descent
6:     $\mathbf{d}_t \leftarrow \frac{1}{\|\mathbf{d}_t\|_2}\mathbf{d}_t$
7:   **end for**
8:   **for** $k = 1, \cdots, M$ **do**
9:     Solving $\{\boldsymbol{\alpha}_i^+\}_{i=1}^N$, $\{\boldsymbol{\alpha}_i^-\}_{i=1}^N$ by ISTA
10:    Update $\mathbf{d}_k$ by optimizing (4) using gradient descent
11:    $\mathbf{d}_k \leftarrow \frac{1}{\|\mathbf{d}_k\|_2}\mathbf{d}_k$
12: **end for**
13: $iter \leftarrow iter + 1$
14: **until** Stopping criterion meets
15: **return** $\mathbf{D}$

For points from negative bags, following (6), we model the reconstruction error of points $\mathbf{x}_{ij}^- \in \mathbf{B}_i^+$ as a zero mean Gaussian distribution with unknown variance,

$$\Pr(l_{ij}^- = -|\mathbf{x}_{ij}^-) = \exp\left(\|\mathbf{x}_{ij}^- - \mathbf{D}^-\boldsymbol{\alpha}_{ij}^-\|^2 \right), \qquad (9)$$

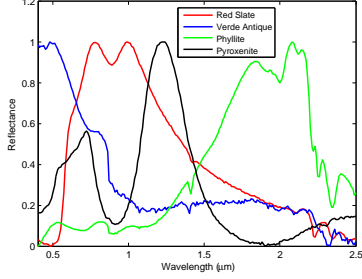where $\boldsymbol{\alpha}_{ij}^-$ is the sparse representation of $\mathbf{x}_{ij}^-$ given $\mathbf{D}^-$. The objective function (4) is then optimized by gradient descent with sparse coding as outlined in Alg. 1

3. EXPERIMENTS

MI-HE was applied to simulated data generated from four spectra selected from the ASTER spectral library [11] shown in Fig. 1. Specifically, the Red Slate, Verde Antique, Phyllite and Pyroxenite spectra from the rock class with 211 bands and wavelengths ranging from $0.4\mu m$ to $2.5\mu m$ were used as endmembers to generate hyperspectral data. Red Slate was labeled as the target endmember. [4] provides a precise description of how the simulated data was generated.

Three sets of highly-mixed noisy data with varied mean target proportion value (p_{t_mean}) were generated. Specifically, this synthetic data has 15 positive and 5 negative bags

$$- \ln J = - \sum_{i=1}^{K^+} \frac{1}{p} \ln \left(\frac{1}{n_i^+} \sum_{j=1}^{n_i^+} \Pr(l_{ij}^+ = +|\mathbf{B}_i^+)^p \right) - \rho \sum_{i=1}^{K^-} \sum_{j=1}^{n_i^-} \ln \Pr(l_{ij}^- = -|\mathbf{x}_{ij}^-), \qquad (4)$$

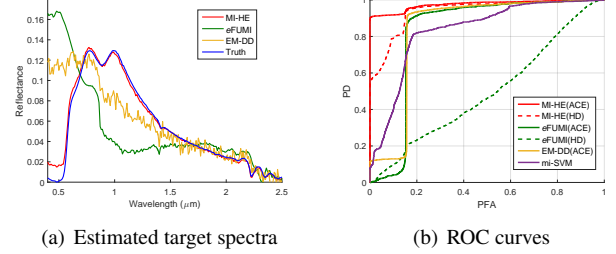
Fig. 1. Signatures used to generate simulated data

(a) Estimated target spectra (b) ROC curves

Fig. 2. Comparison of MI-HE, eFUMI, EM-DD and mi-SVM on synthetic data with mean target proportion 0.3

Table 1. List of Background Endmembers for Synthetic Data

Bag Number	Bag Label	Background Endmember
1-5	+	Verde Antique, Phyllite, Pyroxenite
6-10	+	Phyllite, Pyroxenite
11-15	+	Pyroxenite
16-20	−	Phyllite, Pyroxenite

with each bag has 500 points. If it is positively labeled, there are 100 target points containing mean target (Red Slate) proportion p_{t_mean}. The parameter p_{t_mean} that controls the mean target proportion value was set to 0.3, 0.5 and 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 $30dB$. To highlight the ability of MI-HE to leverage individual bag-level labels, we use different subsets of background endmembers to build synthetic data as shown in Tab. 1.

MI-HE was compared to eFUMI and EM-DD. As we mentioned before, eFUMI combines all positive bags into one big positive bag. Given the aforementioned synthetic data, eFUMI confuses both Red Slate and Verde Antique as target signatures since Verde Antique is missing in the training negative bags (and eFUMI does not preserve bag-level labels). However, the proposed MI-HE is able to learn the target signature correctly. Fig. 2(a) shows the target signature estimated by MI-HE and comparison algorithms eFUMI [4] and EM-DD [3]. We can see that MI-HE is able to learn a target signature that matches the groundtruth, but eFUMI mistakes Verde Antique as target signature and EM-DD learns one noisy, non-pure target signature.

For quantitative evaluation, receiver operating curves (ROC) from detection on testing data generated using the same method are shown in Fig. 2(b). Tab. 2 shows the area of probability of detection (PD) under the curve (AUC), where we can see proposed MI-HE outperforms the comparison

Table 2. Simulated Hyperspectral Data Experiments. Results listed in Median AUC over 5 Runs

Algorithm	P_{t_mean}		
	0.3	0.5	0.7
MI-HE (HD)	0.944	0.981	0.996
MI-HE (ACE)	**0.974**	**0.997**	**0.999**
eFUMI (ACE)	0.810	0.843	0.843
eFUMI (HD)	0.471	0.467	0.504
EM-DD (ACE)	0.841	0.876	0.998
mi-SVM	0.834	0.884	0.891

algorithms eFUMI, EM-DD and mi-SVM [12]. For MI-HE and eFUMI, both the structured-background adaptive coherence/cosine estimator (ACE) [13, 14] and Hybrid Sub-pixel Detector (HD) [6] were applied; for EM-DD, only ACE was applied since EM-DD does not learn a set of background signatures simultaneously. The results reported are the median results over five runs of the algorithm on the same data.

For experiments on real hyperspectral target detection data, the MUUFL Gulfport Hyperspectral data set was used. This data set was collected over the University of Southern Mississippi-Gulfpark Campus and contains 325×337 pixels

Table 3. Gulfport Brown Detection. Results listed in Median NAUC over 5 Runs at FAR $= 1 \times 10^{-3} m^2$

Algorithm	Train: Flight 1, Test: Flight 3	Train: Flight 3, Test: Flight 1
MI-HE (HD)	**0.552**	**0.792**
MI-HE (ACE)	0.442	0.699
eFUMI (ACE)	0.437	0.746
eFUMI (HD)	0.417	0.765
EM-DD (ACE)	0.420	0.749
mi-SVM	0.353	0.333
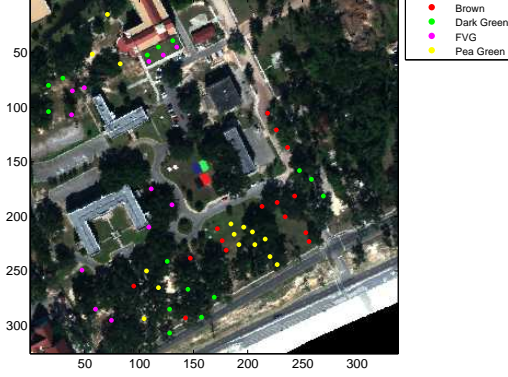
Fig. 3. MUUFL Gulfport data set RGB image and the 64 target locations

with 72 bands corresponding to wavelengths from $367.7nm$ to $1043.4nm$ at a $9.5 - 9.6nm$ spectral sampling interval. The spatial resolution is 1 m. [15] provides more detailed information about the data. Two flights over the area from this data (Gulfport Campus Flight 1 and Gulfport Campus Flight 3) were selected as cross-validated training and testing data. Throughout the scene, there are 64 emplaced man-made targets shown in Fig. 3. The targets are cloth panels of four different colors: Brown (15 examples), Dark Green (15 examples), Faux Vineyard Green (FVG) (12 examples) and Pea Green (15 examples). We take Brown as our target type and for each target in the training flight, a 5×5 rectangular region around each ground truth point for each target were labeled as positive bags to account the drift coming from inaccurate GPS groundtruthing.

For the quantitative analysis of the estimated brown signatures, the normalized area of probability of detection (PD) under the curve (NAUC) at false alarm rate (FAR) $1 \times 10^{-3} m^2$ were computed and were shown in Tab. 3, where the proposed MI-HE combined with HD preserves the best detection performance. Results reported are the median results over five runs.

4. CONCLUSION AND FUTURE WORK

The proposed MI-HE is able to learn target signatures with better quality and achieve competitive and state-of-the-art hyperspectral target detection results when compared to existing multiple instance concept learning methods. The future work includes optimizing the objective function by quasi-newton method to improve its convergence and adding a discriminative term to promote the discriminativeness of the estimated target signature.

5. REFERENCES

[1] T. G. Dieterich *et al.*, "Solving the multiple-instance problem with axis-parallel rectangles," *Artificial Intell.*, vol. 89, no. 1-2, pp. 31–17, 1997.

[2] O. Maron and T. Lozano-Perez, "A framework for multiple-instance learning.," *NIPS*, vol. 10, 1998.

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

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

[5] A. Zare *et al.*, "Multiple instance hyperspectral target characterization," Under Review.

[6] 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.

[7] C. Jiao and A. Zare, "Multiple instance dictionary learning using functions of multiple instances," in *Int. Conf. Pattern Recognition (ICPR)*.

[8] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[9] S. Chen *et al.*, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[10] I. Daubechies *et al.*, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. on Pure and Applied Math.*, vol. 57, pp. 1413–1457, 2004.

[11] AM Baldridge *et al.*, "The ASTER spectral library version 2.0," *Remote Sens. of Environment*, vol. 113, no. 4, pp. 711–715, 2009.

[12] S. Andrews *et al.*, "Support vector machines for multiple-instance learning," in *NIPS*, 2002, pp. 561–568.

[13] S. Kraut and L.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.

[14] W. Basener, "Clutter and anomaly removal for enhanced target detection," in *Proc. SPIE*, 2010, vol. 7695, p. 769525.

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