

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 applied to hyperspectral target detection task rely on knowing the specific target signature in advance. However, in many cases, obtaining such target signature is difficult or impossible. Further more, target signature obtained from laboratory measurement or manual selection from the imagery does not capture the discriminative features of target class. This paper proposed an approach for learning target signature from imprecise labeling information of the targets. The proposed approach maximizes the response of the hybrid detector under a multiple instance framework and estimates a set of discriminative target concepts. After learning target signatures, any signature based detector could be applied for target detection on test data. Both simulated and real hyperspectral target detection experiments were conducted.

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

1. INTRODUCTION

The wealth of spectral information in hyperspectral imagery provides for the ability to perform sub-pixel target detection. In statistical target detection methods, the target and background signals are modeled as random variables distributed according to some respective underlying distribution. The detection problem can then be posed as a binary hypothesis test with two competing hypotheses: target absent (H_0) or target present (H_1) and a detector can be designed using the generalized likelihood ratio test (GLRT) approach.

For detector based hyperspectral target detection, the problem is searching for a known signature within a hyperspectral scene. These target detection algorithms rely on an accurate spectral signature for the target material to be known in advance. Ideally, the target spectral signature need to be fully characterized prior to application of a target detection algorithm. However, in a number of scenarios, obtaining an effective target signature is often a challenging problem. For instance, an analyst or agent on the ground may only know

approximate locations of some possibly sub-pixel targets identified by GPS coordinates (as ground truth information). Manually picking target signature from the imagery would be difficult and mixed with background material. Furthermore, spectral signatures obtained from a spectral library or collected using laboratory instruments do not account for variation due to environmental or atmospheric conditions.

In this paper, we continue to investigate learning target signature from imprecisely labeled hyperspectral imagery, *i.e.*, model the hyperspectral target endmember extraction task as a multiple instance concept learning problem. In multiple instance learning (MIL) [1] problems, training data is grouped into bags instead of individually labeled. A positive bag must contain at least one true positive (target) data and negative bags are composed entirely negative data. Multiple instance concept learning is a branch of MIL that aims to learn one or a set of concept to describe the target category. The Diversity Density (DD) [2] tries to learn a concept for the positive class that is close to the intersection of positive bags and far always from every negative instance, *i.e.*, an area preserves both high density of positive points and low density of negative points, called diversity density. The expectation maximization version of DD (EM-DD) [3] adds a instance selection step in order to improve the convergence of DD. EM-DD selects only one instance per bag as the representative of the labeled bags then performs a quasi-newton optimization on the single-instance DD problem.

Some previous methods for hyperspectral target spectrum estimation given a MIL framework was proposed in [4, 5]. [4] combines all positive bags into one bigger positive bag and learns target signature from the reconstruction error, thus loses some bag-level labeling information. [5] maximizes the response of some matched filters and dose not assume a mixing model, thus may fail on some highly mixed detection problems. The proposed method overcomes the aforementioned drawbacks by maximizing the response of the hybrid detector (HD) [6] under a generalized mean framework.

2. MULTIPLE INSTANCE HYBRID ESTIMATOR

Without loss of generality, let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$ be training data where d is the dimensionality of an instance, \mathbf{x}_j , and N is the total number of training instances. The

This material is based upon work supported by the National Science Foundation under Grant IIS-1350078-CAREER: Supervised Learning for Incomplete and Uncertain Data.

data is grouped into K bags, $\mathbf{B} = \{\mathbf{B}_1, \dots, \mathbf{B}_K\}$, with associated binary bag-level labels, $L = \{L_1, \dots, L_K\}$ 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 j^{th} instance in bag \mathbf{B}_i with instance-level label $l_{ij} \in \{0, 1\}$. When the label on certain bag or instance matters, it will be referred as positive bag \mathbf{B}_i^+ with bag level label L_i^+ containing instances \mathbf{x}_{ij}^+ with instance-level labels $l_{ij}^+, j = 1, \dots, n_i^+$, 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 j^{th} instance in this bag, label for the j^{th} 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.

The proposed multiple instance hybrid estimator (MI-HE) starts from the bag-level likelihood measurement and wants to maximize the probability of given bags,

$$J = \prod_{i=1}^{K^+} \Pr(L_i^+ = 1 | \mathbf{B}_i^+) \cdot \prod_{i=1}^{K^-} \Pr(L_i^- = 0 | \mathbf{B}_i^-). \quad (1)$$

Since the MIL problem defines there must be at least one (could be more) positive instance in each positive bag and the negative bag must consist of all negative instances, we can transfer the probability of individual bag to the instances in each bag, shown as Eq. (2). 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 join probability of all instances in this bag to be negative.

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

Eq. (2) contains a max operation that is difficult to optimize numerically. Some algorithms in the literature [2, 3] 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 repeat 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,

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

where $p \in [-\infty, +\infty]$ is a real number controls the function approximately to be from min to max.

Then taking negative logarithm and scaling the negative part of the likelihood function result in Eq. (4): where the scaling factor ρ is usually set to be smaller than one to control the influence of negative bags.

Similar to [7], here each instance is modeled as a sparse linear combination of target and/or background concepts \mathbf{D} , $\mathbf{x}_j \approx \mathbf{D}\boldsymbol{\alpha}_j$, where $\boldsymbol{\alpha}_j$ is the sparse vector of weights for instance \mathbf{x}_j . Each positive bag contains at least one instance composed of some target:

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

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

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

Given the above data model, we introduce the hybrid detector to discriminate the probability for instances from positive bags to be positive. Specifically, the probability for \mathbf{x}_{ij}^+ in \mathbf{B}_i^+ is defined as:

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

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

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

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 $\boldsymbol{\alpha}$.

For points from negative bags, following Eq. (6), we model the reconstruction error of points $\mathbf{x}_{ij}^- \in \mathbf{B}_i^-$ as a zero mean Gaussian distribution with unknown variance, shown as Eq. (9),

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

where $\boldsymbol{\alpha}_{ij}^-$ is the sparse representation of \mathbf{x}_{ij}^- given \mathbf{D}^- and is solved by solving Eq. (8).

The objective function (4) is optimized by gradient descent with sparse coding shown in Alg. 1

3. EXPERIMENTS

MI-HE was run on simulated data generated from four spectra selected from the ASTER spectral library [11] shown in

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

Algorithm 1 MI-HE algorithm

```

1: Initialize  $\mathbf{D}^0$ ,  $iter = 0$ 
2: repeat
3:   for  $t = 1, \dots, T$  do
4:     Solving  $\{\alpha_i^+\}_{i=1}^N, \{\alpha_i^-\}_{i=1}^N$  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  $k = 1, \dots, M$  do
9:     Solving  $\{\alpha_i^+\}_{i=1}^N, \{\alpha_i^-\}_{i=1}^N$  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  $\mathbf{D}$ 

```

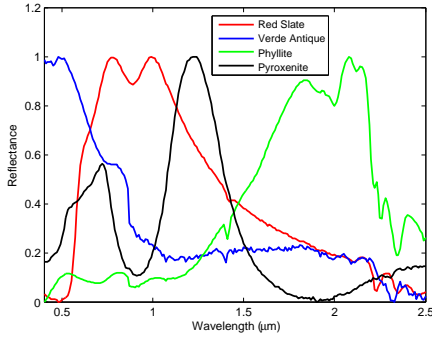


Fig. 1. Signatures from ASTER library used to generate simulated data

Fig. 1. Specifically, the Red Slate, Verde Antique, Phyllite and Pyroxenite spectra from the rock class with 211 bands and wavelengths ranging from $0.4\mu\text{m}$ to $2.5\mu\text{m}$ 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.

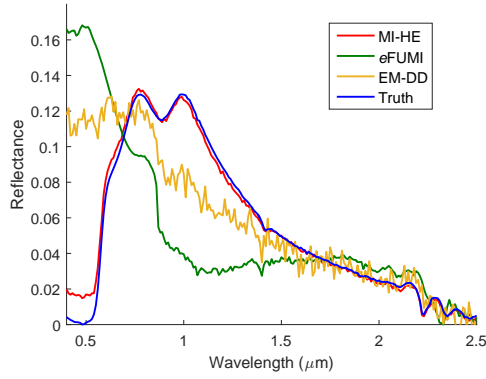
Three sets of highly-mixed noisy data with varied mean target proportion value ($p_{t,mean}$) were generated, respectively. Specifically, this synthetic data has 15 positive and 5 negative bags with each bag has 500 points. If it is positively labeled, there are 100 target points with $p_{t,mean}$ target proportion. The parameter $p_{t,mean}$ that controls the mean target proportion value was set to 0.3, 0.5 and 0.7 respectively to get different 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. What is different from the synthetic data in [4] is there are only two background endmembers, Phyllite and Pyroxenite, involved in the negative bags; in the first 5 positive bags, target endmember Red Slate and all three background endmembers are presented; in the second 5 positive bags target endmember Red Slate and two background endmembers, Phyllite and Pyroxenite, are present; in the last 5 positive bags, only target endmember and one background endmember Pyroxenite are present.

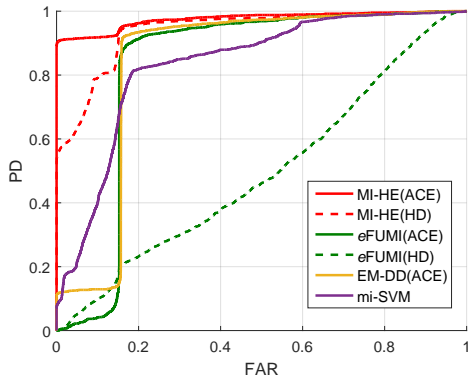
As we mentioned before, eFUMI combines all positive bags into one big positive bags. Given the aforementioned synthetic data, theoretically eFUMI will confuse both Red Slate and Verde Antique as target concept since Verde Antique is missing in the training negative bags. However, the proposed MI-HE should be able to learn the target concept correctly since it aims to maximize the detection statistics of each positive bag. Fig. 2(a) shows the estimated target signature estimated by MI-HE and comparison algorithms eFUMI [4] and EM-DD [3], where we can see that MI-HE is able to learn an exact target signature as groundtruth, but eFUMI mistakes Verde Antique as target concept and EM-DD learns one noisy and 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. 1 shows the normalized area of probability of detection (PD) under the curve (NAUC) at false alarm rate (FAR) 1×10^{-3} , where we can see proposed MI-HE outperforms the comparison algorithms eFUMI, EM-DD and mi-SVM [12]. For MI-HE and eFUMI, both detectors, structured-background adaptive coherence/cosine estimator (ACE) [13, 14] and HD [6] were applied; for EM-DD, only ACE was applied since EM-DD does not learn a set of background concept 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-Gulfport Campus and contains 325×337 pixels 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



(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. Simulated Hyperspectral Data Experiments. Results listed in Median NAUC over 5 Runs at $\text{FAR} = 1 \times 10^{-3}$

Algorithm	$P_{t,mean}$		
	0.3	0.5	0.7
MI-HE (HD)	<u>0.537</u>	<u>0.670</u>	0.952
MI-HE (ACE)	0.895	0.978	0.998
eFUMI (ACE)	0.0	0.0	0.0
eFUMI (HD)	0.0	0.0	0.0
EM-DD (ACE)	0.113	0.158	<u>0.986</u>
mi-SVM	0.075	0.0	0.0

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

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	<u>0.765</u>
EM-DD (ACE)	0.420	0.749
mi-SVM	0.353	0.333

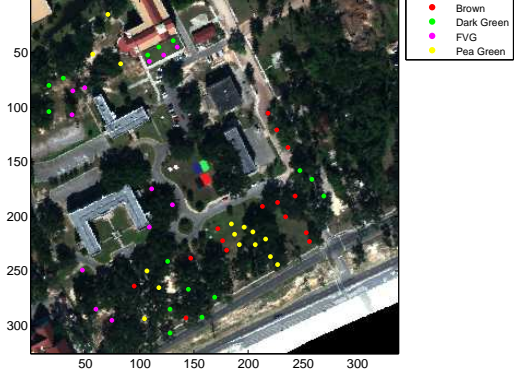


Fig. 3. MUUFL Gulfport data set RGB image and the 64 target locations

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 GPS groundtruthing.

For the quantitative analysis of the estimated brown signatures, the NAUCs at $\text{FAR} 1 \times 10^{-3}$ were computed and were shown in Tab. 2, 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 signature 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. Dietterich *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. 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 Baldrige *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.