Symmetric Information–Theoretic Metric Learning for Target Detection in Hyperspectral Imagery

Yanni Dong, *Member, IEEE*, Yuxiang Zhang, *Member, IEEE*, and Bo Du, *Senior Member, IEEE*

***Abstract*—Metric learning-based methods, which yield great performance and show considerable potential to improve the performance of hyperspectral image processing, aim to calculate the Mahalanobis distance metric matrix. In this article, we proposed a symmetric information-theoretic metric learning (SITML) method for hyperspectral target detection. The SITML algorithm is designed based on the classical information-theoretic metric learning (ITML) and, minimizes the differential Kullback–Leibler (KL) divergence. To enhance both of the detection performance and the generalization ability, we build metric spaces from the neighborhood of training samples to preserve the local discriminative information. Then, we conduct local pairwise constraints to maximize the Jeffery divergence (also named the symmetric KL divergence) of two multivariate Gaussian distributions to solve the problem of an asymmetric KL divergence. Finally, we use a closed-form solution to solve the optimization problem. Intensive experiments on three hyperspectral datasets indicate that SITML outperforms the classical ITML algorithm and other state-of-the-art target detection methods.**

***Index Terms*—Hyperspectral target detection, metric learning, symmetric Kullback–Leibler (KL) divergence.**

I. Introduction

HYPERSPECTRAL remote sensing covers the reflectance of a material's surface over hundreds of contiguous spectral wavelength bands. Each material has a specific wave reflectance, which can be used for distinguishing different materials [1]–[7]. Based on this spectral characteristic, hyperspectral target detection, the goal of which is to distinguish pixels as target pixels or background ones with prior information, has attracted much interest in the remote sensing processing field and has many applications, such as military reconnaissance and striking, crop yield estimation, and mineralogy resource investigation [8]–[10].

Manuscript received August 16, 2021; revised November 3, 2021 and December 7, 2021; accepted January 2, 2022. Date of publication January 11, 2022; date of current version February 7, 2022. This work was jointly supported by the National Natural Science Foundation of China under Grants 62171417, 61801444, 62071438 and 41871243, in part by the Open Research Project of The Hubei Key Laboratory of Intelligent Geo-Information Processing under Grant KLIGIP-2018A01, and in part by the Science and Technology Major Project of Hubei Province (Next-Generation AI Technologies) under Grant 2019AEA170. *(Corresponding author: Yuxiang Zhang.)*

Yanni Dong and Yuxiang Zhang are with the Hubei Subsurface Multi-Scale Imaging Key Laboratory, Institute of Geophysics and Geomatics, China University of Geosciences, China University of Geosciences, Wuhan 430074, China (e-mail: dongyanni@cug.edu.cn; zhangyx@cug.edu.cn).

Bo Du is with the National Engineering Research Center for Multimedia Software, Institute of Artificial Intelligence, School of Computer Science and Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University, Wuhan 430079, China (e-mail: dubo@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3140756

Many target detection methods have been developed. Spectral angle mapper is a very straightforward detection algorithm, measuring the "distance" between the test and target spectra [11]. Constrained energy minimization is a linear filter, in which the amount of spectral output energy is minimized under the same constraints [12], [13]. Adaptive coherence/cosine estimator (ACE), which is considered as one of the best hyperspectral target detection algorithms, is derived from generalized likelihood ratio test [14]–[16]. Orthogonal subspace projection (OSP) projects the test spectra onto the orthogonal subspace of undesired target signatures (background spectra), to maximize the signal-to-noise ratio of the residual target signature [17], [18]. Moreover, some extended versions seeking to obtain better detection performance have been presented. For example, kernel OSP maps the original space into the kernel space for handling with the problem of linearly inseparability [19], [20]. Target-constrained interference-minimized filter (TCIMF) can eliminate of the undesired targets [21].

Traditional target detection algorithms usually require rigorous assumptions on the spectral distribution. For example, the ACE algorithm assumes that the background has additional noise, leading to poor performance because it is not fully realistic for real hyperspectral images (HSIs). Additionally, the spectrum of a pixel may be a mixture of different ground objects due to the low spatial resolution of HSIs with the result that it is difficult to detect the targets of interest. For the sake of overcoming the current boundedness of traditional target detection methods, this article carries out research based on metric learning theory.

Metric learning springs from measurement theory to solve the classification problem [22]. In general, supervised metric learning, which can minimize the distance between similar samples as much as possible while effectively separating dissimilar samples, uses the prior information to learn an appropriate Mahalanobis metric matrix. Motivated by the need to properly define similarity measures, many metric learning-based methods have been proposed. The goal of the neighborhood component analysis (NCA) algorithm is to maximize the random variance, which is sensitive to the initial value and selected parameters [23]. The large margin nearest neighbor (LMNN) algorithm, extends NCA algorithm using the maximum margin setting [24]. Information-theoretic metric learning (ITML) method introduces the Kullback–Leibler (KL) divergence to measure the distance between the corresponding multivariate Gaussians of the Mahalanobis matrix and a given Mahalanobis distance function [25]. The ITML algorithm can obtain fast and scalable performance with the capability to handle many

different of constraints. Cross-view quadratic discriminant analysis (XQDA) algorithm can be specifically formulated as a generalized eigenvalue decomposition by learning the distance metric [26].

Target detection has the same purpose, which aims to label pixels as target or background [27], [28]. Thus, these aspects motivate us to extend distance metric learning to hyperspectral target detection. Based on this, we proposed some methods to obtain better performance of hyperspectral target detection. Random forest metric learning algorithm combines semimultiple metrics with random forests to better separate the desired targets and background [29]. Adaptive information-theoretic metric learning with local constraints, which was previously presented in [30], relaxes the fixed threshold of the ITML method with local decision constraints to retain the locality information.

Generally, metric learning-based methods are able to offer a natural solution for measuring the similarity of samples. Generally, the advantages of metric learning-based methods are as follows.

1) They obtain the metric matrix, reflecting the notion of similarity for the target detection task.
2) They need fewer hypotheses, leading to more powerful generalization ability.
3) Most algorithms, such as NCA algorithm [23], have the ability to handle with the high dimensional data with few samples.
4) They are robust to noisy data and can be trained with weak constraints.

However, for target detection tasks, most existing metric learning methods have two major challenges. Existing metric learning-based methods either ignore local discriminative information or use fixed pairwise constraints to learn the distance metric, which makes it less effective at handling data with complex distributions. In addition, it is difficult for most metric learning methods to determine the global minimum value of the objective formula. Thus, considering the problems of classical target detection methods and traditional metric learning methods, we present a novel symmetric ITML (SITML) method for hyperspectral target detection. This article has three main contributions, which can be summarized as follows.

1) To preserve the local discriminative information, while considering the locality of the data distribution and improving the detection performance, we consider each training sample as a center and build a within-class difference space and between-class difference space from its neighborhood.
2) To solve the problem of asymmetric KL divergence, which works in a weakly supervised manner and leads to an inaccurate description of distance, we apply local pairwise constraints to maximize the Jeffery divergence as the objective function by using the abovementioned neighborhood of each training sample.
3) To implement and handle large scale problems, we obtain the solution via a closed-form solution instead of using Bregman's method to solve the optimization problem.

The rest of this article is organized as follows. A brief overview of metric learning theory and the ITML algorithm is given in Section II. In Section III, SITML is introduced in detail. In Section IV, extensive experimental results are conducted on one synthetic and two real HSIs. Finally, Section V concludes this article.

II. Related Work

In this section, the basic concepts of metric learning theory and the ITML algorithm are briefly reviewed.

A. Metric Learning Theory

Suppose that each sample is expressed by an *L*th dimensional training set $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{n} \in \mathrm{R}^L$, where *i* corresponds to the sample index, *L* is the number of feature dimensions and *n* is the number of training samples. Given two samples $\mathbf{x}_i$ and $\mathbf{x}_j$, the Euclidean distance is defined as $\mathbf{d_E}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)}$, which means the square root of the inner product of the difference of two samples. The Euclidean distance cannot handle with the scaling and dimensionality of the features. Therefore, most distance metric learning-based methods calculate a Mahalanobis distance metric matrix $\mathbf{M} \in R^{L\times L}$ to ensure that $\mathbf{d_M}$ is a meaningful distance as follows:

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)}. \tag{1}$$

Because the metric matrix $\mathbf{M}$ is a symmetric and positive semidefinite matrix, (1) can be decomposed as $\mathbf{M} = \mathbf{W}\mathbf{W}^{\mathrm{T}}$, where $\mathbf{W}$ is a nonsquare matrix and can map data from high-dimensional space into a low-dimensional space for dimensionality reduction, and $\mathbf{W} \in \mathbf{R}^{L\times D}(D \ll L)$ [31].

Equation (1) can be rewritten as

$$\begin{aligned} d_{\mathbf{W}}(\mathbf{x}_i, \mathbf{x}_j) &= \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{W}\mathbf{W}^T(\mathbf{x}_i - \mathbf{x}_j)} \\ &= \sqrt{(\mathbf{W^T}\mathbf{x}_i - \mathbf{W^T}\mathbf{x}_j)^T(\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j)}. \end{aligned} \tag{2}$$

Once the nonsquare matric $\mathbf{W}$ is learned, for any test pixel spectral vector $\mathbf{x}_i \in R^L$, the metric feature representation can be calculated in the metric feature space by $\mathbf{x'}_i = \mathbf{W}^T\mathbf{x}_i$. By sorting the nearest neighbors of the target pixel or applying a detector in such a metric feature space, the target detection results can be obtained.

B. Overview of the ITML Algorithm

The ITML method aims to measure the "closeness" between the Mahalanobis matrix $\mathbf{M}$ and a given Mahalanobis matrix $\mathbf{M}_0$ via a natural information-theoretic technique. That is, the ITML method can minimize the differential relative entropy between two multivariate Gaussians via a natural information-theoretic approach. The ITML algorithm has two main advantages: it can handle many constraints while optionally incorporating the prior information on the objective function; and it can obtain fast and scalable results. Motivated by this article, some extended methods, such as the LogDet divergence-based metric learning method, have been proposed [32].

Given a Mahalanobis distance parameterized by $\mathbf{M}$, its corresponding multivariate Gaussian can be obtained, which can be parameterized by mean vector $\mu$and covariance matrix $\mathbf{\Sigma}$ as:

$$p(\mathbf{x};\mathbf{M}) = \frac{1}{(2\pi)^{\frac{l}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}}\exp(-\frac{1}{2}d_{\mathbf{M}}(\mathbf{x},\mu)) = \frac{1}{(2\pi)^{\frac{l}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}}\exp(-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)) \quad (3)$$

where $|\mathbf{\Sigma}|$ is the determinant of $\mathbf{\Sigma}$. The distance between two Mahalanobis distance functions parameterized by $\mathbf{M}_0$ and $\mathbf{M}$ by the differential relative entropy of the corresponding multivariate Gaussians (also known as the KL divergence) can be measured as

$$KL(p(\mathbf{x};\mathbf{M_0})||p(\mathbf{x};\mathbf{M})) = \int p(\mathbf{x};\mathbf{M}_0)\log\frac{p(\mathbf{x};\mathbf{M}_0)}{p(\mathbf{x};\mathbf{M})}d\mathbf{x}. \quad (4)$$

According to [25] and [33], ITML assumes that the means of the Gaussians are the same, and the divergence is formulated as

$$KL(p(\mathbf{x};\mathbf{M_0})||p(\mathbf{x};\mathbf{M})) = \frac{1}{2}\left[tr(\mathbf{\Sigma}_{\mathbf{M}}^{-1}\mathbf{\Sigma}_{\mathbf{M_0}}) + \log\left|\mathbf{\Sigma}_{\mathbf{M}}\mathbf{\Sigma}_{\mathbf{M_0}}^{-1}\right| - D\right] \quad (5)$$

where D is constant that represents the dimension. $tr(\mathbf{\Sigma}_{\mathbf{M}}^{-1}\mathbf{\Sigma}_{\mathbf{M_0}})$ denotes the trace of $\mathbf{\Sigma}_{\mathbf{M}}^{-1}\mathbf{\Sigma}_{\mathbf{M_0}}$. Then, the LogDet optimization problem minimizing the differential relative entropy of multivariate Gaussians, can be solved.

Considering the feasible solution, ITML algorithm incorporates the slack variables ξ_0 and ξ (where ξ is the vector of slack variables, and ξ_0 is the initialization vector) into the formula (5), which can ensure a viable existence of matrix $\mathbf{M}$

$$\begin{aligned}&\min_{M\geq 0,\xi}\left[tr(\mathbf{\Sigma}_{\mathrm{M}}^{-1}\mathbf{\Sigma}_{\mathbf{M}_0}) + \log\left|\mathbf{\Sigma}_{\mathbf{M}}\mathbf{\Sigma}_{\mathbf{M}_0}^{-1}\right| - D\right]\\ &+\gamma\cdot\left[tr(\mathbf{\Sigma}_{diag(\xi)}^{-1}\mathbf{\Sigma}_{diag(\xi_0)}) + \log\left|\mathbf{\Sigma}_{diag(\xi)}\mathbf{\Sigma}_{diag(\xi_0)}^{-1}\right| - D\right]\\ &s.t. d_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) \leq \xi_c(i,j)(\mathbf{x}_i,\mathbf{x}_j) \in similarsamples\\ &d_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) \geq \xi_c(i,j)(\mathbf{x}_i,\mathbf{x}_j) \in dissimilarsamples\end{aligned} \quad (6)$$

where $c(i,j)$ denotes the (i,j)th constraint. γ denotes the tradeoff parameter.

To solve the optimization problem of formula (6), the methods can be extended from [34], which forms the basis for the algorithm by repeatedly computing Bregman projections. This projection can be performed via the update $\mathbf{M}_{t+1} = \mathbf{M}_t + \beta\mathbf{M}_t(\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^{\mathrm{T}}\mathbf{M}_t$, where β is the projection parameter and t is the number of iterations.

III. Proposed Method

Motivated by the ITML algorithm, we propose the SITML method. The proposed SITML algorithm uses the local metric spaces to preserve the local discriminative information. Then, local pairwise constraints are constructed to maximize the Jeffery divergence to both improve the detection performance and handle the problem of asymmetric KL divergence. Finally, a closed-form solution is calculated to obtain the solution to the optimization problem and, acquire the metric matrix. By using the metric matrix, we can transform data into the Mahalanobis metric feature space to achieve better detection performance.

A. Local Metric Space Formulation

Let $y_{ij} \in (+1,-1)$ denotes the relations of the training samples. If $\mathbf{x}_i$ and $\mathbf{x}_j$ are the same class, then $y_{ij}=1$, and if $\mathbf{x}_i$ and $\mathbf{x}_j$ are the different classes, then $y_{ij}=-1$. The insight that for each center sample $\mathbf{x}_i$, there are k_1 nearest neighbors in the same class (within-classes) and k_2 nearest neighbors in different classes (between-classes) can be obtained. The sets of possible differences for sample $\mathbf{x}_i$ and its k_1 within-class difference space, are denoted as $N_{k_1}^{+}(\mathbf{x}_i)$. Similarly, the sets of possible differences for sample $\mathbf{x}_i$ and its k_2 between-class difference space, are denoted as $N_{k_2}^{-}(\mathbf{x}_i)$. We can obtain the following matrices:

$$\begin{aligned}\mathbf{S} &= \left\{(\mathbf{x}_i-\mathbf{x}_j)|\mathbf{x}_i\in\mathbf{X},\mathbf{x}_j\in N_{k_1}^{+}(\mathbf{x}_i)\right\}\\ \mathbf{D} &= \left\{(\mathbf{x}_i-\mathbf{x}_j)|\mathbf{x}_i\in\mathbf{X},\mathbf{x}_j\in N_{k_2}^{-}(\mathbf{x}_i)\right\}.\end{aligned} \quad (7)$$

$\mathbf{S}$ represents the within-class difference space, and $\mathbf{D}$ represents the between-class difference space. The dimension of matrices S and D is $L\times L$. By constructing a neighborhood of training samples, the local distinguishable information can be retained, which can be used to consider the locality of the data distribution and improve the detection performance.

B. Maximization of the Jeffery Divergence

Although ITML is an effective algorithm for different applications, the KL divergence of ITML, which works in a weakly supervised manner may lead to inaccurate description of distance, is asymmetric ($KL(P_1||P_2) \neq KL(P_2||P_1)$). In addition, by constantly changing the parameters of the predicted distribution, different values of the KL divergence can be obtained. Thus, we apply the symmetric KL divergence, which is also known as Jeffrey divergence, to build a framework from the neighborhood of each training sample. Compared to the KL divergence, the Jeffery divergence can better judge the similarity of samples.

Suppose $P_{\mathbf{S}}$ is the distribution of the within-class difference space and $P_{\mathbf{D}}$ is the distribution of the between-class difference space. Given a Mahalanobis distance, we achieve the corresponding multivariate Gaussian distributions of $P_{\mathbf{S}}$ and $P_{\mathbf{D}}$, where $\mathbf{\Sigma}_{\mathbf{S}}$ and $\mathbf{\Sigma}_{\mathbf{D}}$ are the covariance matrices, respectively. To compute the within-class and between-class covariance matrices $\mathbf{\Sigma}_{\mathbf{S}}$ and $\mathbf{\Sigma}_{\mathbf{D}}$, respectively, we apply the maximum likelihood estimation as

$$\begin{aligned}\mathbf{\Sigma}_{\mathbf{S}} &= \frac{1}{|\mathbf{S}|}\sum_{i=1}^{n}\left[\sum_{\mathbf{x}_j\in N_{k_1}^{+}(\mathbf{x}_i)}(\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^T\right]\\ \mathbf{\Sigma}_{\mathbf{D}} &= \frac{1}{|\mathbf{D}|}\sum_{i=1}^{n}\left[\sum_{\mathbf{x}_j\in N_{k_2}^{-}(\mathbf{x}_i)}(\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^T\right]\end{aligned} \quad (8)$$

where n is the number of training samples.

Considering $\mathbf{M} = \mathbf{W}\mathbf{W}^T$, the Jeffery divergence between the transformed distributions $P_{\mathbf{S_W}}$ and $P_{\mathbf{D_W}}$ which is with zero mean and covariance matrices $\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W}$ and $\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W}$, can be formulated as

$$\begin{aligned} J(\mathbf{W}) &= \mathrm{KL}(P_{\mathbf{S_W}}, P_{\mathbf{D_W}}) + \mathrm{KL}(P_{\mathbf{D_W}}, P_{\mathbf{S_W}}) \\ &= \frac{1}{2}\Big[tr((\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W}) \\ &\quad + \log\Big|(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1}\Big|) - D\Big] \\ &\quad + \frac{1}{2}\Big[tr((\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W}) \\ &\quad + \log\Big|(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}\Big|) - D\Big] \\ &= \frac{1}{2}\Big[tr((\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W}) \\ &\quad + (\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W}))\Big] - D \end{aligned} \tag{9}$$

where *D* is constant. Thus, we can calculate (9) by removing *D* and obtaining the transformed metric matrix as

$$\begin{aligned} &\underset{\mathbf{W}\in\mathbf{R}^{L\times D}}{agrmax}\, J(\mathbf{W}) \\ &\quad = tr((\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W}) \\ &\quad + (\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})). \end{aligned} \tag{10}$$

According to our previous work [30], ITML is subject to a fixed threshold, which cannot be used to handle data subject to complex distributions. Thus, we use a local decision function $f(d_{ij})$, instead of a fixed threshold. Combined with (6) and (10), we can obtain the final objective function as

$$\begin{aligned} &\underset{\mathbf{W}\in\mathbf{R}^{L\times D}}{agrmax}\, J(\mathbf{W}) = tr((\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W}) \\ &\qquad + (\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})). \\ &\qquad + \gamma\cdot\Big[tr(\mathbf{\Sigma}^{-1}_{diag(\xi)}\mathbf{\Sigma}_{diag(\xi_0)}) \\ &\qquad + \log\Big|\mathbf{\Sigma}_{diag(\xi)}\mathbf{\Sigma}^{-1}_{diag(\xi_0)}\Big| - D\Big] \\ &s.t.\ d_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) \le f_{same}(\mathbf{d}_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)) \\ &\qquad + \xi_c(i,j)(\mathbf{x}_i,\mathbf{x}_j) \in similarsamples \\ &\quad d_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j) \ge f_{diff}(\mathbf{d}_{\mathbf{M}}(\mathbf{x}_i,\mathbf{x}_j)) \\ &\qquad + \xi_c(i,j)(\mathbf{x}_i,\mathbf{x}_j) \in dissimilarsamples \\ &\quad \xi_c(i,j) \ge 0, \forall(i,j), d \ge 0 \end{aligned} \tag{11}$$

where

$$\begin{aligned} &f_{same}(d_{ij}) \\ &\quad = d_{ij} - (d_{ij}^{(1/N_{same})}/d_{max})(\mathbf{x}_i,\mathbf{x}_j) \in similarsamples \\ &f_{diff}(d_{ij}) \\ &\quad = d_{ij} + (d_{max}/d_{ij}^{(1/N_{diff})})(\mathbf{x}_i,\mathbf{x}_j) \in dissimilarsamples \end{aligned} \tag{12}$$

where d_{ij} is the distance of $(\mathbf{x}_i,\mathbf{x}_j)$. d_{max} is the maximal distance among all the pairs $(\mathbf{x}_i,\mathbf{x}_j)$. N_{same} is the scale factor that controls the shrinkage of similar samples. And N_{diff} is the scale factor that controls the expansion of dissimilar samples. To guarantee that the local decision function $f_{same}(d_{ij})$ can shrink faster, we set $N_{same} = 1$ in this article. Likewise, to guarantee the local decision function $f_{diff}(d_{ij})$ can expand faster, we set $N_{diff} = 1/\log_2(d_{max}/(d_{max}-2))$.

C. Optimization Problem and Its Solution

Our goal is to find the linear transformation $\mathbf{W}$ that maximizes the Jeffrey divergence $J(\mathbf{W})$. To achieve this, according to [35], we take the derivative of the objective function $J(\mathbf{W})$ with respect to $\mathbf{W}$ as

$$\begin{aligned} &\frac{\partial}{\partial\mathbf{W}}J(\mathbf{W}) \\ &= -2\mathbf{\Sigma_S}\mathbf{W}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1}\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1} \\ &\quad + 2\mathbf{\Sigma_D}\mathbf{W}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1} \\ &\quad - 2\mathbf{\Sigma_D}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1} \\ &\quad + 2\mathbf{\Sigma_S}\mathbf{W}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}. \end{aligned} \tag{13}$$

The optimal matrix $\mathbf{W}$ should satisfy $\frac{\partial}{\partial\mathbf{W}}J(\mathbf{W}) = 0$, thus we can obtain the following formulas:

$$\begin{aligned} \mathbf{\Sigma_S}^{-1}\mathbf{\Sigma_D}\mathbf{W} &= \mathbf{W}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W}) \\ \mathbf{\Sigma_D}^{-1}\mathbf{\Sigma_S}\mathbf{W} &= \mathbf{W}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W}). \end{aligned} \tag{14}$$

In addition, we assume that $\mathbf{\Lambda} \in \mathbf{R}^{D\times D}$ is the diagonal matrix, where $\lambda_1, \lambda_2, \cdots \lambda_D$ are the eigenvalues of $\mathbf{\Sigma_S}^{-1}\mathbf{\Sigma_D}$. Thus, we have

$$\mathbf{\Sigma_S}^{-1}\mathbf{\Sigma_D}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}. \tag{15}$$

Because the inverse of a nonsingular matrix shares identical eigenvectors and reciprocal eigenvalues, once we have the eigenvectors and eigenvalues of the matrix, we can obtain analogous information on the inverse of the matrix. $\mathbf{\Sigma_D}^{-1}\mathbf{\Sigma_S}$ is the inverse matrix of $\mathbf{\Sigma_S}^{-1}\mathbf{\Sigma_D}$, and hence

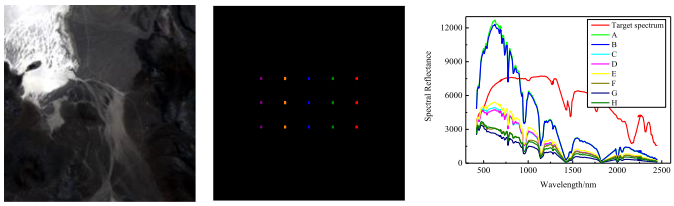
$$\mathbf{\Sigma_D}^{-1}\mathbf{\Sigma_S}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}^{-1}. \tag{16}$$

Combined with (14)–, we can obtain the following formula:

$$\begin{aligned} (\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W}) &= \mathbf{\Lambda} \\ (\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W}) &= \mathbf{\Lambda}^{-1}. \end{aligned} \tag{17}$$

Finally, the objective function can be simplified as

$$\begin{aligned} &\underset{\mathbf{W}\in\mathbf{R}^{L\times D}}{agrmax}\, J(\mathbf{W}) \\ &\quad = tr((\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})) \\ &\qquad + tr((\mathbf{W}^T\mathbf{\Sigma_D}\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{\Sigma_S}\mathbf{W})) \\ &\quad = tr(\mathbf{\Lambda}) + tr(\mathbf{\Lambda}^{-1}) \\ &\quad = \sum_{i=1}^{D}(\lambda_i + \lambda_i^{-1}) \end{aligned} \tag{18}$$



(a) (b) (c)

Fig. 1. (a) AVIRIS LCVF dataset. (b) Implanted target locations. (c) Target spectrum and some typical spectra of background samples.

TABLE I
DETAILS OF THE IMPLANTED TARGET PANELS FOR THE AVIRIS LCVF DATASET

Color of Target Panel	Sample Index	Line Index	Fraction
	50	(75, 76) (100, 101) (125, 126)	10%
	75	(75, 76) (100, 101) (125, 126)	20%
	100	(75, 76) (100, 101) (125, 126)	30%
	125	(75, 76) (100, 101) (125, 126)	40%
	150	(75, 76) (100, 101) (125, 126)	50%

To solve (11), we set the linearly independent $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_i$ as eigenvectors of $\boldsymbol{\Sigma}_{\mathbf{S}}^{-1}\boldsymbol{\Sigma}_{\mathbf{D}}$. That is, we can solve (11) via a closed-form solution [36], [37]. The eigenvalues λ_i of $\boldsymbol{\Sigma}_{\mathbf{S}}^{-1}\boldsymbol{\Sigma}_{\mathbf{D}}$ can be obtained with generalized eigenvalue decomposition by solving the following formula:

$$\boldsymbol{\Sigma}_{\mathbf{D}}\boldsymbol{\omega}_i = \lambda_i \boldsymbol{\Sigma}_{\mathbf{S}}\boldsymbol{\omega}_i. \tag{19}$$

Then, we can obtain matrix $\mathbf{W}$, the columns of which are the vectors $\boldsymbol{\omega}_i$, where $i = 1, 2, \ldots D$. That is, we can acquire linear projection matrix $\mathbf{W}$ (or $\mathbf{M}$). For any arbitrary $\mathbf{x}_i \in R^{L\times n}$, we can obtain the transformed data via

$$\mathbf{x'}_i = \mathbf{W}^T\mathbf{x}_i. \tag{20}$$

Finally, we can detect targets by using any detector.

IV. EXPERIMENTAL RESULTS

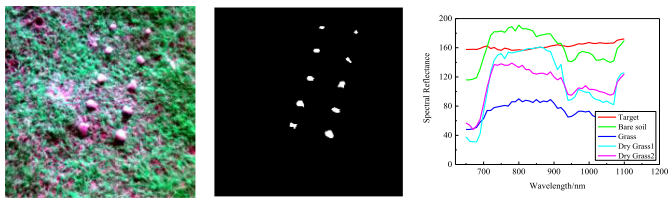
In this section, we present the three popular HSI datasets employed, describe the experimental settings of the parameters, and analyze the effectiveness of SITML.

A. Hyperspectral Datasets Description

The first synthetic image was acquired by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor, covering the lunar crater volcanic field (LCVF) in USA, with 200×200 pixels and 188 bands, which was created in [38] and [39]. In addition, the alunite spectrum was selected from the U.S. Geological Survey digital spectral library and used as the target spectrum. Fig. 1(a) and (b) shows the image scene and the locations of target panels, which are implanted. There are two mixed pixels, which are mixed with prior target spectrum $\mathbf{t}$ and background spectra $\mathbf{b}$ of the original position as: $\mathbf{x} = p\mathbf{t} + (1-p)\mathbf{b}$, in each target panel. p represents the fraction, which varies from 10% to 50% from left to right. Table I gives the details. Fig. 1(c) shows the target spectrum and typical spectra of the background samples. Eight background pixels are randomly chosen as the background samples.

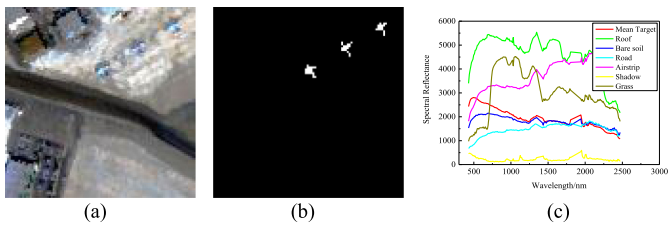
TABLE II
AVERAGE COMPUTATIONAL TIME (IN SECONDS) ANALYSIS OF DIFFERENT ALGORITHMS FOR THE THREE DATASETS

Datasets / Methods	AVIRIS LCVF	Nuance CRI stone	AVIRIS San Diego airport
ACE	0.8929	0.6696	0.3177
OSP	0.1031	0.0924	0.0410
TCIMF	1.0161	2.6874	0.6877
LMNN	2.1320	0.9172	4.2955
XQDA	0.0167	0.0094	0.0212
ITML	0.0299	0.0216	0.0578
SITML	0.0315	0.0202	0.0552



(a) (b) (c)

Fig. 2. (a) Nuance CRI stone dataset. (b) Target locations. (c) Target spectrum and some typical spectra of background samples.



(a) (b) (c)

Fig. 3. (a) AVIRIS San Diego airport dataset. (b) Target locations. (c) Target spectrum and some typical spectra of background samples.

The second dataset was acquired by the Nuance CRI hyperspectral sensor, including 400×400 pixels and 1186 target pixels of stone, as shown in Fig. 2(a). Fig. 2(b) and (c) shows the locations of the stones and the spectral signatures, respectively, [40], [41]. The number of spectral bands is 46. Ideally, spectra should be from the spectral library for target detection. In this experiment, the red line of Fig. 3(c) represents the priori target spectra, which are chosen from the center of a stone. Twenty background spectra are randomly chosen as the background samples.

The third dataset was collected by AVIRIS, located in San Diego, USA. Fig. 3(a) shows the image scene, including 189 spectral bands. The image size is 100×100 pixels [42]–[45]. The 58 target pixels of the three aircrafts are the targets of interest, which are shown in Fig. 3(b). Three pixels labeled at the center of each aircraft are selected as target samples, while ten background pixels are randomly chosen as background samples. Fig. 3(c) displays the target spectrum and some typical spectra of background samples.

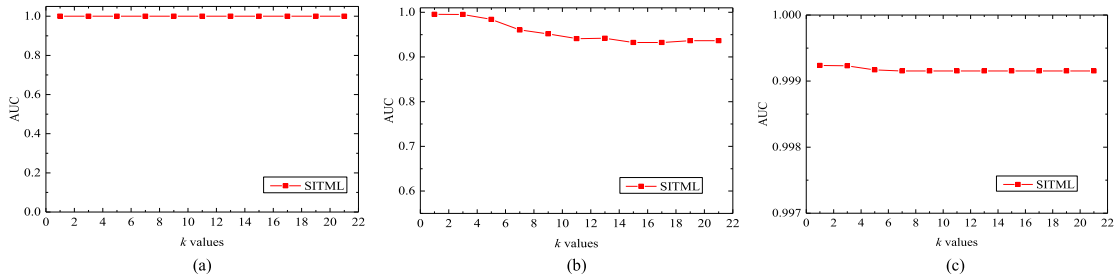


Fig. 4. AUC values of SITML with respect to parameter k. (a) AVIRIS LCVF dataset. (b) Nuance CRI stone dataset. (c) AVIRIS San Diego airport dataset.

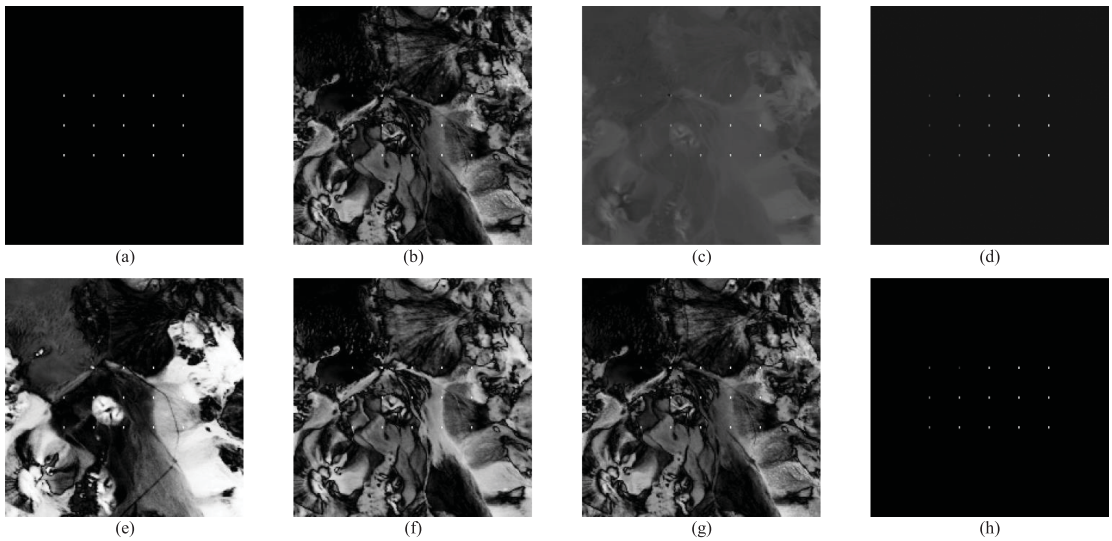


Fig. 5. 2-D plots on the AVIRIS LCVF dataset. (a) Ground truth. (b) Adaptive coherence/cosine estimator. (c) Orthogonal subspace projection. (d) Target-constrained interference-minimized filter. (e) Large margin nearest neighbor. (f) Cross-view quadratic discriminant analysis. (g) Information-theoretic metric learning. (h) Symmetric information-theoretic metric learning.

B. Experimental Setting

To prove the significance of the proposed SITML algorithm, the two-dimensional (2-D) detection maps, obtained by the detection test statistic results, are used to qualitatively analyze the detection performance. Higher brightness implies that target locations are more obvious and the ability to detect targets of interest is higher. Moreover, the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) value are regarded as the classic comparison measurements [46], [47]. The ROC curves, which illustrate the relations between the false positive rates and the probability of detection at a set of different thresholds, are widely used in the target detection applications as a standard performance evaluation tool. Generally, at the same level of false positive rates, an algorithm that has a higher probability of detection performs better. That is, the ROC curve of a good detection method should be closer to the upper left of the coordinate axis. However, the ROC curves of different algorithms may alternate with each other, making it difficult to judge which method performs better. The AUC value represents the average behavior used to assess the accuracy. The method with the larger AUC value is considered to be the better method. To further prove the ability of the SITML algorithm, the situations of separation between the target and background were evaluated. For each detector, we normalize the target and background columns from 0 to 1. The positions of the columns represent the separability of the target and background.

To evaluate the effectiveness of SITML, we compared it with three classical target detectors, i.e., ACE, OSP, and TCIMF; and three effective metric learning-based methods, i.e., LMNN, XQDA, and ITML algorithms; on three HSIs.

To conduct impartial comparison, for each HSI, SITML, and all comparison algorithms adopt the same target spectra and background ones as the priori samples for each algorithm except the ACE method, which only needs the same target samples. For the metric learning-based methods, i.e., LMNN, XQDA, ITML, and SITML methods, we use the ACE as the target detector. Thus, we also use the ACE as the comparison algorithm to prove the ability of SITML.

The parameters of the LMNN, XQDA, and ITML algorithms are adjusted according to the relevant [48]–[50]. As mentioned in Section III, for the proposed SITML algorithm, k_1 and k_2 are the within-class and between-class nearest neighbors, respectively. To simplify the analyses, we set $k = k_1 = k_2$. To investigate the impact of the choice of the nearest neighbors on the performance of the proposed SITML algorithm, Figs. 4(a)–(c) illustrate the

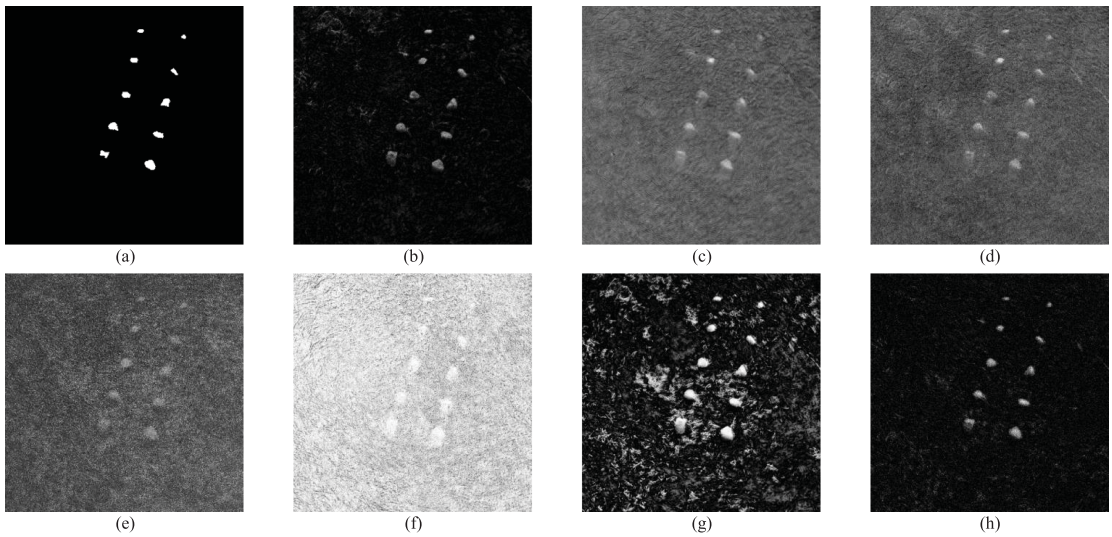


Fig. 6. 2-D plots on the Nuance CRI stone dataset. (a) Ground truth. b) Adaptive coherence/cosine estimator. (c) Orthogonal subspace projection. (d) Target-constrained interference-minimized filter. (e) Large margin nearest neighbor. (f) Cross-view quadratic discriminant analysis. (g) Information-theoretic metric learning. (h) Symmetric information-theoretic metric learning.

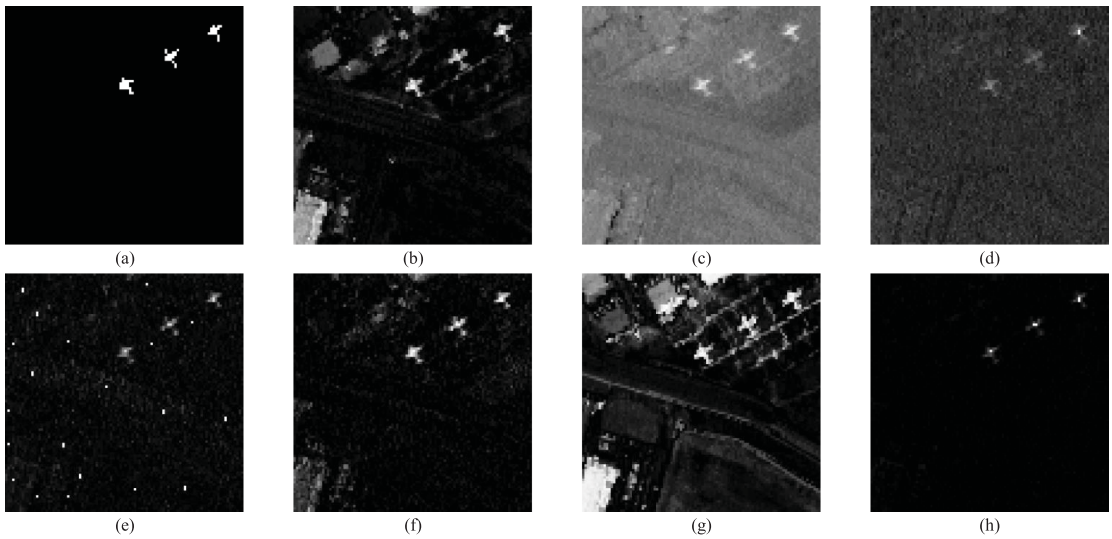


Fig. 7. 2-D plots on the AVIRIS San Diego airport dataset. (a) Ground truth. b) Adaptive coherence/cosine estimator. (c) Orthogonal subspace projection. (d) Target-constrained interference-minimized filter. (e) Large margin nearest neighbor. (f) Cross-view quadratic discriminant analysis. (g) Information-theoretic metric learning. (h) Symmetric information-theoretic metric learning.

AUC values of the SITML method for different parameter k. Obviously, the detection performance behaves very stably and effectively when k is sufficiently small. Consequently, we set $k = k_1 = k_2 = 5$ for all the experiments.

C. Experimental Results and Analysis

1) Detection Performance: To analyze the results qualitatively, the 2-D results on the AVIRIS LCVF dataset, which are normalized from 0 to 1, are shown in Fig. 5(a)–(h). For the AVIRIS LCVF dataset, except for the TCIMF and SITML algorithm, none of the algorithms show good distinguishable detection maps. The proposed the SITML algorithm obtains the highest detection values for the target pixels when they suppress the background effectively. This indicates that SITML algorithm is more robust than the comparison algorithms. Fig. 6(a)–(h) show the 2-D plots of the detection results on the Nuance CRI stone dataset. We can again observe that the target locations of the ACE and SITML algorithms are more obvious while suppressing the background than those of the other comparison algorithms. That is, only the ACE and SITML algorithms obtain satisfactory detection results, while the OSP, TCIMF, LMNN, XQDA, and ITML algorithms hardly work on this dataset because of the high false positive rates. Similar to the above two datasets, we plot the 2-D plots of the results on the AVIRIS San Diego airport dataset, as shown in Fig. 7. The SITML

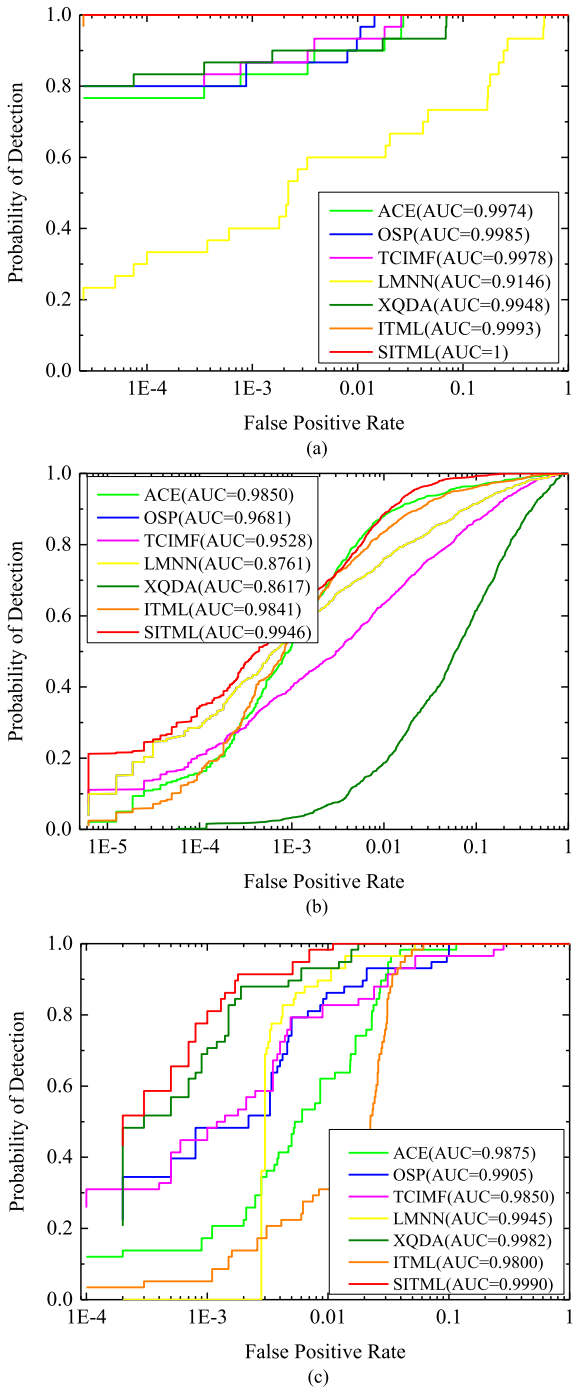


Fig. 8. ROC curves and AUC values for the three datasets. (a) AVIRIS LCVF. (b) Nuance CRI stone. (c) AVIRIS San Diego airport.

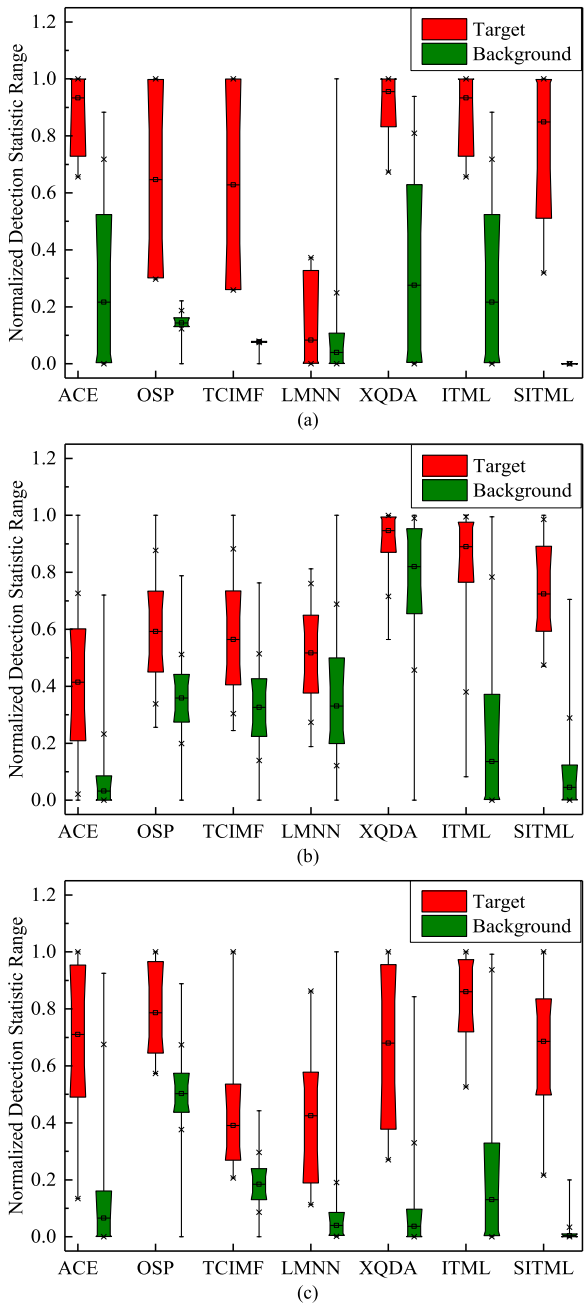


Fig. 9. Separability maps for the three datasets. (a) AVIRIS LCVF. (b) Nuance CRI stone. (c) AVIRIS San Diego airport.

algorithm outputs high statistical values of target pixels while suppressing the background effectively. As for the other comparison algorithms, i.e., ACE, OSP, TCIMF, LMNN, XQDA, and ITML, are insufficient at suppressing the background, although they can recognize all the targets, which may cause high false positive rates. This is probably because SITML considers the local pairwise constraints derived from the neighborhood of each training sample, which can maximize the Jeffery divergence and significantly improve the performance of the detector.

In order to assess the detection performance quantitatively, we plot the ROC curves in log-scale and obtain the AUC values by using the detection test statistic results of different algorithms, as shown in Fig. 8. Furthermore, we use the AUC values to measure the general behavior to further evaluate the performance.

As shown in Fig. 8(a), on the AVIRIS LCVF dataset, it can be observed that the ITML and SITML algorithms achieve better detection performances since their ROC curves are always above those of the other comparison algorithms. However, the ROC curves of the ITML and SITML algorithms alternate with each other, making it difficult to judge which method performs best. However, the AUC value, achieved by the SITML algorithm

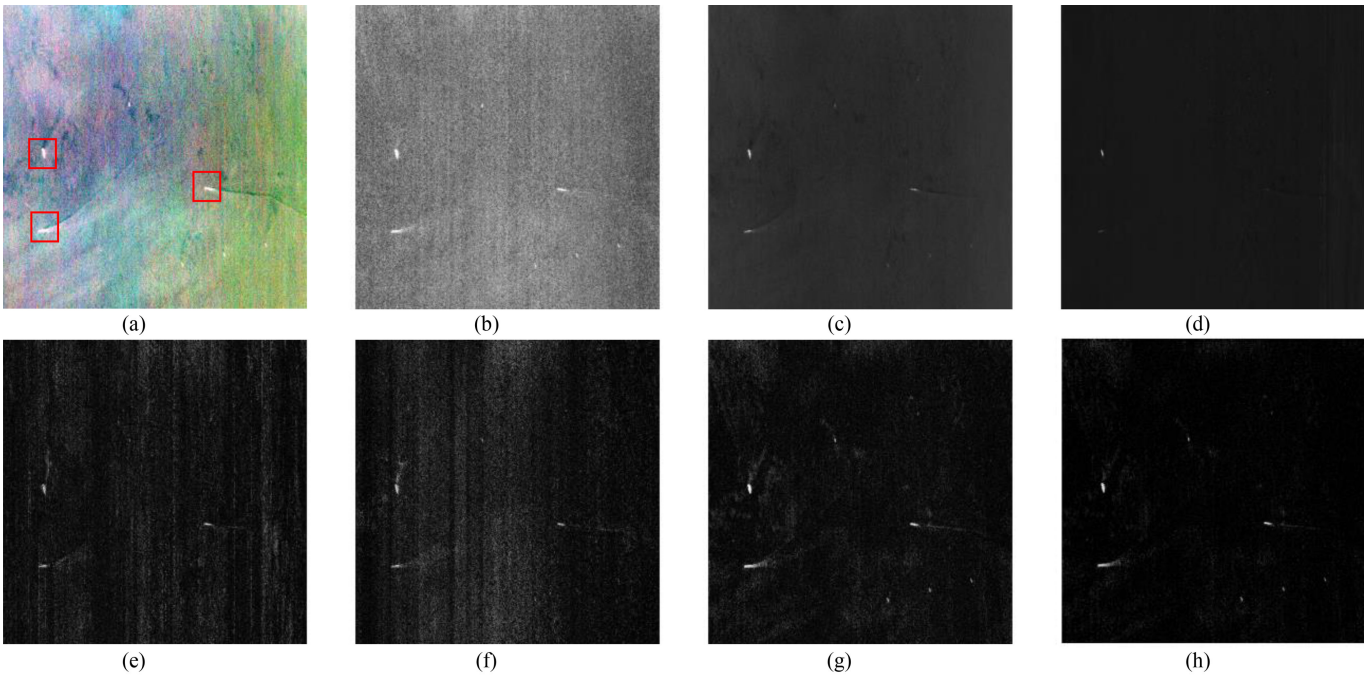


Fig. 10. 2-D plots on OHS Qingdao vessel dataset. (a) False-color composite map. (b) Adaptive coherence/cosine estimator. (c) Orthogonal subspace projection. (d) Target-constrained interference-minimized filter. (e) Large margin nearest neighbor. (f) Cross-view quadratic discriminant analysis. (g) Information-theoretic metric learning. (h) Symmetric information-theoretic metric learning.

improves from 0.9993 to 1 compared with the ITML algorithm. This indicates that SITML can obtain better performance than the ITML algorithm.

As shown in Fig. 8(b), for the Nuance CRI stone dataset, the false positive rate of the SITML algorithm decreases to the 10^{-1} level when the probability of detection is 100%, while the false positive rates of the comparison algorithms are approximately 1 when the probability of detection is 100%. Compared with the other algorithms, the ROC curve of SITML illustrates the better performance under the FAR 10^{-3} level, which is the primary range of interest. The AUC value of the proposed SITML algorithm is the largest value compared with the other algorithms, showing that SITML has the best performance compared with the other methods.

As shown in Fig. 8(c), the ROC of SITML is closest to the upper left and always above those of the comparison algorithms on the AVIRIS San Diego airport dataset. The SITML algorithm has the best probability of detection under the same false positive rate, which indicate that SITML has the highest detection performance. As expected from the ROC curve analysis, the AUC values of the comparison algorithms are less than that of SITML.

2) Separability Analysis: The separability between the target and background is shown in Fig. 9. For all three datasets, the proposed SITML algorithm can always restrain the background information and has the best separability between the target and background. For the AVIRIS LCVF dataset in Fig. 9(a), we can observe that there is a small gap between the target and background for the ACE, OSP, TCIMF, XQDA, and ITML, but SITML illustrates the best separability. For the Nuance CRI stone dataset in Fig. 9(b), the gap between the two columns of each comparison algorithm is not very obvious except for ITML. Although the gap between the target and background for ITML is obvious, it cannot restrain background information very well. For the AVIRIS San Diego airport dataset, ITML and SITML show the best separability, while ITML cannot suppress the background very well.

3) Computational Time: To compare the computational time of different algorithms, all the experiments are implemented on the same computer. Table II gives the average running time of each method for ten times on the AVIRIS LCVF dataset, Nuance CRI stone dataset and AVIRIS San Diego airport dataset. The computational time is related to the pixel size, the number of bands and the number of prior samples. In general, the running time of the proposed SITML algorithm has comparable speed performance compared with the TCIMF and LMNN, while it is a little slower than the running time of the other comparison algorithms. That is, although SITML needs to construct a within-class difference space and between-class difference space from the neighborhood of each sample, SITML is still competitive with the comparison algorithms. That occurs because SITML solve the optimization problem via a closed-form solution.

Although the efficiency of our method is not dominant, it outperforms the comparison algorithms on all the datasets. We believe that the computational time of the proposed SITML algorithm can be accelerated through GPU acceleration.

V. Practical Application and Analysis

In this section, we further verify the practicability of the proposed SITML method, using the newest data collected by the orbita hyperspectral (OHS) satellite. OHS satellite is the first hyperspectral satellite with surface coating technology for sensors in the world. Compared with other hyperspectral satellites,

OHS breaks through the bottleneck of hyperspectral satellites and opens a new era of remote sensing image processing [51], [52].

This article uses the HSI of the Qingdao coast acquired on April 17, 2019, and applies one of the scenes as an example for experiment. The raw data has been preprocessed, including radiation correction and geometric correction. The spectral coverage of this scene is 466–940 nm, including 32 bands from visible to near-infrared wavelengths. And the spatial resolution is 10 m. The image window size is 1000×1000 pixels. Three vessels are the main targets of interest, which are marked in red box in Fig. 10(a). In this experiment, six pixels labeled at the center of each vessel are selected as target samples, eight background spectra are randomly chosen as the background samples.

We qualitatively analyze the performance of the proposed SITML method compared with the comparison methods by drawing the 2-D plots according to the detection test statistic results, which are shown in Fig. 10(b)–(h). On the OHS Qingdao vessel dataset, the ACE and OSP can identify the desired targets, while failing to restrain the background pixels. In contrast, the TCIMF can effectively suppress the background pixels, while failing to highlight the desired targets. Compared with the LMNN, XQDA and ITML methods, the proposed SITML algorithm can further suppress the background while the target positions of SITML are more prominent. In this experiment, our proposed SITML algorithm is superior to the comparison methods in qualitative evaluation.

VI. Conclusion

Target detection is very important for HSI processing, but its performance is far from satisfying. Metric learning-based methods are straightforward and effective at offering a natural solution to measure the similarity of samples. In this article, we present the SITML method to detect targets of interest. To preserve the discriminative information, our method constructs a within-class difference space and between-class difference space from the neighborhood of each sample. To solve the problem of asymmetric KL divergence, we use the Jeffery divergence to build the metric learning framework, which can be optimized by a closed-form solution. To demonstrate the effectiveness of the SITML algorithm, extensive experiments are implemented on three HSIs. Six benchmark algorithms are compared with the proposed algorithm via 2-D plots, ROC curve analysis, AUC values and separability maps. The experimental results indicate that the SITML algorithm is superior to the comparison algorithms. In future work, we will focus on code optimization to accelerate the calculation time to achieve further superiority in real applications. Additionally, we will compare the 3D ROC curves [53] of different algorithms to further evaluate the effectiveness of SITML.

Acknowledgment

The authors would like to thank Dr. Shanwei Liu and Dr. Mingming Xu for generously providing the OHS Qingdao vessel dataset for our experiment. The authors would also like to thank Prof. Jun Li and all anonymous reviewers for providing insightful and helpful comments.

References

[1] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho, through the use of a constrained energy minimization technique," *Remote Sens. Environ.*, vol. 59, no. 1, pp. 64–76, 1997.

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

[3] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "DAEN: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, Jul. 2019.

[4] D. G. Manolakis, R. B. Lockwood, and T. W. Cooley, *Hyperspectral Imaging Remote Sensing: Physics, Sensors, and Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[5] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5085–5102, Jun. 2021.

[6] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 911–923, Feb. 2018.

[7] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018.

[8] K. Tiwari, M. K. Arora, and D. Singh, "An assessment of independent component analysis for detection of military targets from hyperspectral images," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 13, no. 5, pp. 730–740, Oct. 2011.

[9] E. M. Winter *et al.*, "Mine detection experiments using hyperspectral sensors," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 5415, pp. 1035–1041, 2004.

[10] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.

[11] F. A. Kruse, "The spectral image processing system (SIPS)–interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, vol. 44, no. 2/3, pp. 145–163, May–Jun. 1993.

[12] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Ph.D. dissertation, Dept. Elect. Eng., Univ. Maryland, Baltimore County, Baltimore, MD, USA, 1993.

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

[14] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, 1992.

[15] D. Manolakis, D. Marden, and G. A. J. L. l. j. Shaw, "Hyperspectral image processing for automatic target detection applications," vol. 14, no. 1, pp. 79–116, 2003.

[16] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Is there a best hyperspectral detection algorithm?," *Int. Soc. Opt. Photon.*, vol. 7334, 2009, Art. no. 733402.

[17] S. Bernabe, S. Sanchez, A. Plaza, S. López, J. A. Benediktsson, and R. Sarmiento, "Hyperspectral unmixing on GPUs and multi–core processors: A comparison," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1386–1398, Apr. 2013.

[18] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[19] L. Capobianco, A. Garzelli, and G. Camps–Valls, "Target detection with semisupervised kernel orthogonal subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3822–3833, Nov. 2009.

[20] H. Kwon and N. M. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 178–194, Feb. 2006.

[21] H. Ren and C. I. Chang, "Target–constrained interference–minimized approach to subpixel target detection for hyperspectral images," *Opt. Eng.*, vol. 39, no. 12, pp. 3138–3145, Dec. 2000.

[22] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, Sep. 1936.

[23] J. Goldberger, G. E. Hinton, S. T. Roweis, and R. R. Salakhutdinov, "Neighbourhood components analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, pp. 513–520.

[24] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.

[25] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information–theoretic metric learning," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 209–216.
[26] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re–identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2197–2206.
[27] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Hyperspectral remote sensing image subpixel target detection based on supervised metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4955–4965, Aug. 2014.
[28] Y. Dong, W. Shi, B. Du, X. Hu, and L. Zhang, "Asymmetric weighted logistic metric learning for hyperspectral target detection," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2021.3070909.
[29] Y. Dong, B. Du, and L. Zhang, "Target detection based on random forest metric learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 4, pp. 1830–1838, Apr. 2015.
[30] Y. Dong, B. Du, L. Zhang, and X. Hu, "Hyperspectral target detection via adaptive information-theoretic metric learning with local constraints," *Remote Sens.,* vol. 10, no. 9, Sep. 2018, Art. no. 1415.
[31] P. Moutafis, M. Leng, and I. A. Kakadiaris, "An overview and empirical comparison of distance metric learning methods," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 612–625, Mar. 2017.
[32] J. Mei, M. Liu, H. R. Karimi, and H. Gao, "LogDet divergence–based metric learning with triplet constraints and its applications," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4920–4931, Nov. 2014.
[33] J. V. Davis and I. Dhillon, "Differential entropic clustering of multivariate Gaussians," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 337–344.
[34] B. Kulis, M. Sustik, and I. Dhillon, "Learning low-rank kernel matrices," in *Proc. Int. Conf. Mach. Learn.*, 2006, pp. 505–512.
[35] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Lyngby, Denmark: Technical Univ. Denmark 2012.
[36] P. Zadeh, R. Hosseini, and S. Sra, "Geometric mean metric learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2464–2471.
[37] S. Wang and R. Jin, "An information geometry approach for distance metric learning," in *Proc. Artif. Intell. Statist. Conf.*, 2009, pp. 591–598.
[38] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Hyperspectral remote sensing image subpixel target detection based on supervised metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4955–4965, Aug. 2014.
[39] H. Yu, L. Gao, W. Liao, and B. Zhang, "Group sparse representation based on nonlocal spatial and local spectral similarity for hyperspectral imagery classification," *Sensors*, vol. 18, no. 6, pp. 1695, 2018.
[40] L. Zhang, B. Du, and Y. Zhong, "Hybrid detectors based on selective endmembers," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2633–2646, Jun. 2010.
[41] Y. Dong, B. Du, L. Zhang, and L. Zhang, "Dimensionality reduction and classification of hyperspectral images using ensemble discriminative local metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2509–2524, May 2017.
[42] Y. Gu, C. Wang, S. Wang, and Y. Zhang, "Kernel-based regularized-angle spectral matching for target detection in hyperspectral imagery," *Pattern Recognit. Lett.*, vol. 32, no. 2, pp. 114–119, Jan. 2011.
[43] Y. Niu and B. Wang, "Extracting target spectrum for hyperspectral target detection: An adaptive weighted learning method using a self-completed background dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1604–1617, Mar. 2017.
[44] Y. Liu, G. Gao, and Y. Gu, "Tensor matched subspace detector for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 1967–1974, Apr. 2017.
[45] B. Tu, X. Yang, X. Ou, G. Zhang, J. Li, and P. Antonio, "Ensemble entropy metric for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2021.3116681.
[46] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, Apr. 2008.
[47] Y. Zhang, B. Du, and L. Zhang, "A sparse representation-based binary hypothesis model for target detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1346–1354, Mar. 2015.
[48] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 209–216.
[49] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.
[50] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2197–2206.
[51] Y. Jiang, J. Wang, L. Zhang, G. Zhang, X. Li, and J. Wu, "Geometric processing and accuracy verification of Zhuhai-1 hyperspectral satellites," *Remote Sens.,* vol. 11, no. 9, 2019, Art. no. 996.
[52] Q. Li, R. Zhong, and Y. Wang, "A method for the destriping of an orbita hyperspectral image with adaptive moment matching and unidirectional total variation," *Remote Sens.,* vol. 11, no. 18, 2019, Art. no. 2098.
[53] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, Jun. 2021.

Yanni Dong (Member, IEEE) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2012, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Lab of Information Engineering in Surveying, Mapping and Remote sensing, Wuhan University, in 2017.

She is currently an Associate Professor with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China. She was a Hong Kong Scholar with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, in 2019. Her current research interests include hyperspectral image processing, pattern recognition and machine learning.

Dr. Dong is a Reviewer of more than twenty international journals, including the IEEE Transactions on Cybernetics, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON MULTIMEDIA, PR, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. She regularly serves as PC Member of IJCAI and AAAI.

Yuxiang Zhang (Member, IEEE) received the B.S. degree in sciences and techniques of remote sensing and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2011 and 2016, respectively.

From 2019 to 2020, she was a Visiting Scholar with the University of Sydney, Sydney, NSW, Australia. She is currently an Associate Professor with the Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China. Her current research interests include hyperspectral image processing, pattern recognition, and machine learning.

Bo Du (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from the State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2010.

He is currently a Professor with the School of Computer Science and Institute of Artificial Intelligence, Wuhan University. He is also the Director of the National Engineering Research Center for Multimedia Software, Wuhan University. He has authored or coauthored more than 80 research papers published in the IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON CYBERNETICS (TCYB), IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS), and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL), etc. 13 of them are ESI hot papers or highly cited papers. His major research interests include pattern recognition, hyperspectral image processing, and signal processing.

Dr. Du is an Associate Editor for *Neural Networks, Pattern Recognition and Neurocomputing*. He is also a Reviewer of 20 Science Citation Index (SCI) magazines including IEEE TPAMI, TCYB, TGRS, TIP, JSTARS, and GRSL. He was the recipeint of the Highly Cited Researcher 2019 by the Web of Science Group, the IJCAI (International Joint Conferences on Artificial Intelligence) Distinguished Paper Prize, the IEEE Data Fusion Contest Champion, and the IEEE Workshop on Hyperspectral Image and Signal Processing Best Paper Award, in 2018. He is also regularly a Senior PC Member of IJCAI and AAAI. He was the Area Chair for ICPR.