Sparse Transfer Manifold Embedding for Hyperspectral Target Detection

Lefei Zhang, *Student Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, Dacheng Tao, *Senior Member, IEEE*, and Xin Huang

***Abstract*— Target detection is one of the most important applications in hyperspectral remote sensing image analysis. However, the state-of-the-art machine-learning-based algorithms for hyperspectral target detection cannot perform well when the training samples, especially for the target samples, are limited in number. This is because the training data and test data are drawn from different distributions in practice and given a small-size training set in a high-dimensional space, traditional learning models without the sparse constraint face the over-fitting problem. Therefore, in this paper, we introduce a novel feature extraction algorithm named sparse transfer manifold embedding (STME), which can effectively and efficiently encode the discriminative information from limited training data and the sample distribution information from unlimited test data to find a low-dimensional feature embedding by a sparse transformation. Technically speaking, STME is particularly designed for hyperspectral target detection by introducing sparse and transfer constraints. As a result of this, it can avoid over-fitting when only very few training samples are provided. The proposed feature extraction algorithm was applied to extensive experiments to detect targets of interest, and STME showed the outstanding detection performance on most of the hyperspectral datasets.**

***Index Terms*— Dimension reduction (DR), elastic net, hyperspectral, target detection, transfer learning.**

I. Introduction

HYPERSPECTRAL image (HSI) processing is concerned with the analysis and interpretation of spectra acquired from a given scene, at some distance, by an airborne or satellite sensor [1]–[3]. The basic task underlying many HSI processing applications is to identify different materials based on their reflectance spectra [4]. Target detection is one such challenge, which aims to separate specific target pixels or subpixels from background pixels, based on their spectral signatures [5]–[9]. When the target spectral signature, which can be obtained from a standard spectral library (SPL) or from given training data, is available as prior information, the most frequently used target detection algorithms are: 1) spectral angle mapper (SAM) [10], which measures the similarity of the given target spectra and test pixel spectra by the spectral angle; 2) adaptive matched filter (AMF) [11], [12], which is based on the generalized likelihood ratio test, but with a simplified test statistic; 3) adaptive coherence estimator (ACE) [13], [14], which estimates the noise covariance structure using a sample covariance matrix; 4) constrained energy minimization (CEM) [15], which maximizes the response of the target spectral signature while suppressing the response of the unknown background signatures; and 5) orthogonal subspace projection (OSP)-based detectors [16]–[20], which achieves the elimination of undesired background signatures by projecting each pixel's spectral vector onto a subspace, which is orthogonal to the undesired signatures. However, due to the challenging spectral variability phenomenon in HSI processing, these algorithms are only optimal under some assumptions and they can only work well in some situations, e.g., the signal-processing-based AMF algorithms assume that the target and background covariance matrices are identical, which is sometimes far from being realistic [21], [22].

In recent years, many machine-learning algorithms have been demonstrated as being effective for hyperspectral target detection. For instance, some nonlinear versions of the detection algorithms have been proposed, inspired by the kernel method [23], which can implicitly exploit the nonlinear characteristics of data through the use of kernels. These include kernel-matched subspace detectors [24], [25], kernel OSP [26], and kernel-based regularized-angle spectral matching [27]. However, only a few researchers have suggested the use of a machine-learning-based supervised dimension reduction (DR) approach for target detection, although it has been commonly adopted in HSI classification [28]–[32]. As we know, the main objective of HSI classification is to automatically assign all pixels into land-cover classes, while the main objective of detection is to search for pixels of a specific target (although detection can also be viewed as a binary classification); however, in target detection, the number of pixels in the target class (positive samples) is often very limited, and the background class (negative samples) is the union of the different land-cover classes, which includes almost all the pixels in the image [8]. Therefore, the following two aspects need to be addressed.

Manuscript received November 14, 2011; revised January 30, 2013; accepted February 7, 2013. Date of publication March 26, 2013; date of current version December 12, 2013. This work was supported in part by the National Basic Research Program of China 973 Program under Grant 2011CB707105, the National Natural Science Foundation of China under Grant 41101336 and Grant 41061130553, and the Program for New Century Excellent Talents in University of China under Grant NCET-11-0396.

L. Zhang, L. Zhang, and X. Huang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zhanglefei@whu.edu.cn; zlp62@whu.edu.cn; huang_whu@163.com).

D. Tao is with the Centre for Quantum Computation and Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney 2007, Australia (e-mail: dacheng.tao@uts.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2246837

1) Many machine-learning-based supervised DR and classification algorithms work well only under the assumption that training and test data are drawn from the same feature space and the same distribution [33]–[36]. However, in hyperspectral target detection, the training data, especially the real-target signatures (positive samples), are quite limited, so it is impossible to recollect the required training data to rebuild the learning model when the test data distribution changes.
2) Most of the existing machine-learning-based supervised DR algorithms learn a linear projection matrix without a sparse constraint. Given the limited training data represented by a large number of spectral features containing random noise, these existing DR algorithms perform poorly when dealing with a large test dataset, and they struggle to interpret the explicit relationship between the original feature and the reduced feature, and end up with the over-fitting phenomenon given a small-size training set [37].

In order to overcome the above two aspects, two types of constraints should be considered in the current supervised DR framework, i.e., a data distribution constraint learned by numerous unlabeled samples, and a sparse constraint imposed on the linear projection matrix. Therefore, in this paper, we introduce a novel sparse transfer manifold embedding (STME) algorithm for HSI DR, which focuses on target detection. Based on a unified manifold embedding framework, we put forward a discriminative manifold embedding optimization using the supervised information. We then propose transfer regularization and sparse formulation on STME to address the two above-mentioned points. The main advantages of STME for hyperspectral detection lie in the following.

1) The primary optimization of STME is derived in a discriminative manifold embedding manner, which can encode given prior information by utilizing a pairwise manifold discriminative analysis to separate target samples from background ones.
2) A transfer-learning-based constraint is adopted in the STME model to propagate discriminative information (given prior information) from training samples to test samples by preserving the dominant structure of the distribution in the reduced feature space, so the learned DR model can be well satisfied with test samples of various classes.
3) An elastic net constraint, which combines the l1-norm and l2-norm penalties, is imposed to control the number of nonzero elements in the linear projection matrix. This constraint controls the complexity of the DR model, provides a good interpretation of the learned subspace, avoids the possible over-fitting phenomenon caused by limited training data, and makes the learned DR model robust to random noise in HSI.

The rest of this paper is structured as follows. Section II provides the proposed STME algorithm in detail, including the primary discriminative optimization and the two significant constraints, i.e., transfer regularization and sparse regularization. The target detection experiments on three challenging hyperspectral datasets are reported in Section III, followed by the conclusion in Section IV.

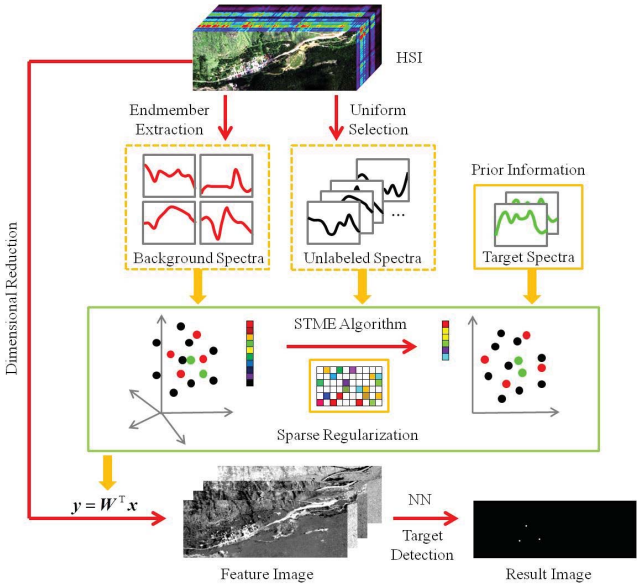

Fig. 1. Flowchart of the STME algorithm for hyperspectral target detection.

II. STME Algorithm

Fig. 1 shows the flowchart of the STME algorithm for hyperspectral target detection. The inputs of the STME algorithm can be divided into three parts.

1) The target spectra, or the labeled target samples, i.e., several spectral signature vectors of the target, which are known as prior information, are represented as $s_i \in R^L$ $(i = 1, 2, \ldots, N_1)$, where L is the number of spectral channels provided in HSI and N_1 denotes the number of given target samples.
2) The background spectra, or the labeled background samples, i.e., some spectral signature vectors of the background pixels, which can be obtained by the endmember extraction of HSI and then applying SAM to exclude the potential target spectra, are represented as $b_i \in R^L$ $(i = 1, 2, \ldots, N_2)$, and N_2 denotes the number of background samples.
3) The unlabeled spectra, or the unlabeled samples, i.e., the numerous spectral signature vectors of unlabeled samples, which are selected from all the pixels in HSI uniformly, are represented as $u_i \in R^L$ $(i = 1, 2, \ldots, N)$, in which N denotes the number of unlabeled samples. The full input data matrix is denoted by $X \in R^{L \times M}$, in which $M = N_1 + N_2 + N$ is the number of total input samples.

The STME algorithm is then performed, which learns a sparse linear projection matrix $W \in R^{L \times d}$ for DR, based on the discriminative information explored in the labeled samples and the dominant structure distribution knowledge provided by all of the input samples. For an arbitrary test pixel spectral vector $x \in R^L$, we compute the reduced feature representation in the sparse transfer manifold embedded feature space by

$$y = W^T x \tag{1}$$

Finally, the target detection result is obtained by sorting the nearest neighbors (NN) of the target feature vector in the sparse transfer manifold embedded feature space.

A. Discriminative Manifold Embedding

The manifold learning method is a nonlinear DR technology, which aims to explore the manifold embedded in the high-dimensional ambient space by preserving a particular local geometry property; e.g., locally linear embedding preserves the linear coefficients, which are used to reconstruct a given measurement by its k-NN [38], isometric feature mapping preserves the global geodesic distances of all the pairs of measurements [39], Laplacian eigenmaps preserve proximity relationships by manipulation of an undirected weighted graph [40], and so on. Based on the two generalized frameworks published in [41] and [42], different manifold embedding algorithms, including their linearization, can be unified to a framework, i.e., a graph embedding framework and a patch alignment framework, respectively. For a target-detection-driven DR, in order to maximally separate the target samples from the background ones, we require the target–target sample pairs to be as close as possible and the target–background sample pairs to be as far away as possible in the learned subspace. The optimization of such an idea can be derived in a manner similar to that of the patch alignment framework, but considering the two groups of training samples differently.

Following the objective of DR, in the low-dimensional feature space, for each sample of target class s_i, we expect that the Euclidean distances between the given sample and the other samples in the target class to be as small as possible, while distances between the given sample and the other samples in the background class are as large as possible

$$\min_W \sum_{j=1}^{N_1} \left\| W^T s_i - W^T s_j \right\|^2 \tag{2}$$

$$\max_W \sum_{j=1}^{N_2} \left\| W^T s_i - W^T b_j \right\|^2. \tag{3}$$

We can combine (2) and (3) together by introducing a trade-off parameter c to control the influence of the two parts

$$\min_W \sum_{j=1}^{N_1} \left\| W^T s_i - W^T s_j \right\|^2 - c \cdot \sum_{j=1}^{N_2} \left\| W^T s_i - W^T b_j \right\|^2. \tag{4}$$

Note that for some hyperspectral target detection situations, only the unique target spectra is available, i.e., $N_1 = 1$ in (2) and (4). Thus, the optimization (4) reduces to only maximize the distances between the given target sample and the other samples in the background class because the first part in optimization (4) equals zero. Then, we can simply set $c = 1$ at this condition.

Based on the patch alignment framework, the local patch of sample s_i is defined as

$$X^i = \left[s_i, s_1^i, s_2^i, \ldots, s_{N_1}^i, b_1^i, b_2^i, \ldots, b_{N_2}^i \right] \in R^{L \times (N_1+N_2+1)} \tag{5}$$

in which $s_j^i (j = 1, \ldots, N_1)$ is the jth sample in the target class and $b_j^i (j = 1, \ldots, N_2)$ is the jth sample in the background class, sorted by the Euclidean distance of the sample pairs (s_i, s_j^i) and (s_i, s_j^i), respectively. The corresponding low-dimensional feature matrix of patch X^i is given by

$$Y^i = W^T X^i \in R^{d \times (N_1+N_2+1)}. \tag{6}$$

In order to further simplify (4), we define a coefficient vector as

$$\delta = \left[\overbrace{1, \ldots, 1}^{N_1}, \overbrace{-c, \ldots, -c}^{N_2} \right]. \tag{7}$$

Then, (4) can be reduced, following the patch optimization of sample t_i

$$\begin{aligned} &\min_W \sum_{j=1}^{N_1+N_2} \delta_{(j)} \left\| W^T X_{(1)}^i - W^T X_{(j+1)}^i \right\|^2 \\ &= \min_W \operatorname{tr} \Bigg(W^T X^i \begin{bmatrix} -e_{N_1+N_2}^T \\ I_{N_1+N_2} \end{bmatrix} \\ &\qquad \times \operatorname{diag}(\delta) \left[-e_{N_1+N_2} I_{N_1+N_2} \right] X^{iT} W \Bigg) \\ &= \min_W \operatorname{tr} \left(W^T X^i G^i X^{iT} W \right). \end{aligned} \tag{8}$$

In (8), $\delta_{(j)}$ is the jth element in $\delta (j = 1, \ldots, N_1 + N_2)$, $X_{(j)}^i$ is the jth column in $X^i (j = 1, \ldots, N_1 + N_2 + 1)$, and

$$G^i = \begin{bmatrix} -e_{N_1+N_2}^T \\ I_{N_1+N_2} \end{bmatrix} \operatorname{diag}(\delta) \left[-e_{N_1+N_2} I_{N_1+N_2} \right] \tag{9}$$

in which $e_{N_1+N_2} = [1, \ldots, 1]^T$ and $I_{N_1+N_2}$ is an identity matrix.

The whole optimization of introduced discriminative manifold embedding is obtained by summing all the patch optimizations of the target samples $s_i (i = 1, \ldots, N_1)$. However, we cannot sum the patch optimizations given in (8) directly, because each patch X_i has its own coordinate system of the detailed samples by the definition in (5). Here, we use a selection matrix to align all the samples together into a consistent coordinate [43]. We assume that the coordinate of patch X^i is selected from the global coordinate, which is also the full input data matrix

$$\begin{aligned} X = \Big[&s_1, s_2, \ldots, s_{N_1}, b_1, b_2, \ldots, b_{N_2}, u_1, \\ &u_2, \ldots, u_N \Big] \in R^{L \times M}. \end{aligned} \tag{10}$$

Then, X^i can be rewritten as

$$X^i = X S^i \tag{11}$$

in which $S^i \in R^{M \times (N_1+N_2+1)}$ is defined by

$$S_{(a,b)}^i = \begin{cases} 1, & \text{if } a = \Phi^i\{b\} \\ 0, & \text{else} \end{cases} \tag{12}$$

where $\Phi^i = \left[i, i_1, \ldots, i_{(N_1+N_2)} \right]$ is the index vector for samples in patch X^i. We then sum the patch optimizations

of the target samples to obtain the whole optimization of discriminative manifold embedding

$$\begin{aligned}\min_W \sum_{i=1}^{N_1} \mathrm{tr}\left(W^T X S^i G^i S^{iT} X^T W\right) \\ = \min_W \mathrm{tr}\left(W^T X \left[\sum_{i=1}^{N_1}\left(S^i G^i S^{iT}\right)\right] X^T W\right) \\ = \min_W \mathrm{tr}\left(W^T X G X^T W\right) \end{aligned} \tag{13}$$

in which

$$G = \sum_{i=1}^{N_1}\left(S^i G^i S^{iT}\right). \tag{14}$$

B. Transfer Regularization

The optimization of discriminative manifold embedding learns a subspace which can maximally separate the target samples from background ones, based only on the prior given training information; however, it is a fact that only a few training samples can be obtained in HSI target detection, especially for the target samples. As a consequence, the dominant structure distribution will be changed in the learned subspace, which may be biased toward that spanned by limited training samples [44]–[46]. Transfer learning is a technology, which proposes to deal with how to preserve the discriminative information from training data to test data that are not in the same feature space and with different data distributions. In general, there are three main approaches to transfer learning [33]: 1) reweigh some supervised samples in the training domain for use in the test domain [47]; 2) find a feature representation that can reduce the difference between the training and test domains and the error of the classification models [48]; and 3) discover shared priors between the training and test domain models, which can benefit from transfer learning [49]. In this paper, we propose to use the dominant structure distribution provided by a large number of unlabeled samples in HSI as a regularization to find a subspace that can transfer the discriminative information from training data to test data and reduce the error of the subsequent classification.

In STME, to maintain the dominant structure of the distribution after DR, we use the subspace arrived at by principal component analysis (PCA) [50], based on the full input data matrix, to restrict the feature space learned by discriminative manifold embedding, which is only based on the given labeled data. PCA is a linear transformation used to find principal components in accordance with the maximum variance of a data matrix. Thus, the dominant structure of the distribution can be well preserved in the subspace after such a transformation. Here, we denote the PCA projection matrix by $P \in R^{L\times d}$. The transfer regularization is used to minimize the Euclidean distance between the full input data matrix in objective subspace and that in the subspace obtained by PCA

$$\min_W \left\| P^T X - W^T X \right\|^2. \tag{15}$$

Finally, to transfer the discriminative information learned from the training samples to the test samples, we connect the optimization of discriminative manifold embedding (13) and transfer regularization (15) to restrict their latent structure to be consistent with that in the ambient feature space by introducing a trade-off parameter β in combination

$$\min_W \mathrm{tr}\left(W^T X G X^T W\right) + \beta \left\| P^T X - W^T X \right\|^2. \tag{16}$$

C. Sparse Formulation

The solution of (16) provides a subspace in which each basis is a linear combination of all the original features; thus, it is often difficult to interpret the results [51]. Moreover, W is a dense matrix in which most elements are nonzero. However, due to the small sample size (SSS) and the high dimensionality of the spectral feature, which usually contains random noise in HSI target detection, it is necessary to control the model complexity according to the regularization theory [52]. Here, we introduce a sparse formulation to restrict the number of nonzero elements in the matrix W. The sparse formulation imposed subspace has the following advantages.

1) Sparsity provides a good interpretation of the DR model and, thus, reveals an explicit relationship between the output feature representation and the given variables.
2) Sparsity can make the output feature simpler and more succinct, so the subsequent processing becomes more efficient.
3) Sparsity decreases the variance brought about by possible over-fitting with the least increment of the bias, so the learned DR model can generalize better.

For the sparse formulation, we need to control the number of nonzero elements of the projection matrix W, which can be characterized by the l^0-norm of the matrix W

$$\min_W F(W) + \|W\|_0 \tag{17}$$

in which $F(W)$ is the optimization of discriminative manifold embedding and transfer regularization defined in (16), and l^0-norm $\|W\|_0$ is simply the number of nonzero elements in W. However, it turns out to be an NP-hard problem and, thus, the computational complexity of solution searching is always exponential [53]. Therefore, the l^1-norm of the projection matrix, i.e., least absolution shrinkage and selection operator (LASSO) [54], is usually adopted as a relaxation of the l^0 formulation

$$\min_W F(W) + \|W\|_1. \tag{18}$$

Least angle regression (LARS) [55] has been proposed to greedily search for the optimal solution of the LASSO-penalized linear regression problem. In the signal-processing field, although LASSO has shown success in many situations, in order to produce a sparse representation that selects the subspace compactly expressing the input feature [56], the LASSO penalty still has the following disadvantages [57]: 1) the number of selected features is limited by the number of given samples, because of the nature of the convex optimization problem and 2) the performance of LASSO may not be optimal if there are high correlations between the input features. To overcome these drawbacks, the elastic net [57],

which combines the l^1-norm and l^2-norm, is introduced. By introducing the elastic net penalty as a sparse formulation to impose on (17), we have the full optimization of STME

$$\min_{W} \operatorname{tr}\left(W^T X G X^T W\right) + \beta \left\| P^T X - W^T X \right\|^2 + \varphi_1 \|W\|_1 + \varphi_2 \|W\|_2^2 \quad (19)$$

in which φ_1 and φ_2 are parameters to control the l^1-norm penalty and l^2-norm penalty in the sparse formulation.

D. Solution for STME

To find the optimal solution of STME in (19), we first show that the objective function is of a quadratic form, with the l^1-norm penalty based on some formula derivation (a detailed discussion of this is given in the Appendix). According to the analysis in the attached Appendix, the full optimization of STME can be transformed to

$$\min_{W^*} \left\| G^* - X^* W^* \right\|^2 + \varphi \left\| W^* \right\|_1 \quad (20)$$

in which the detailed definition of $G^* \in R^{(M+L)\times d}$, $W^* \in R^{L\times d}$, $X^* \in R^{(M+L)\times L}$, and $\varphi \in R$ are given in the Appendix (A7)–(A10). The optimization problem (20) can be efficiently solved by the LARS algorithm [55], [57].

Then, based on (A8), the optimal solution of STME is obtained by

$$W = \frac{W^*}{\sqrt{1+\varphi_2}}. \quad (21)$$

There are three parameters, β, φ_1, and φ_2, in the objective function of STME (19) and one parameter, c, in the coefficient vector (7). In this paper, in order to avoid cross-validation, we set these parameters according to their physical meanings. β reveals the weight between discriminative manifold embedding and the transfer regularization and, thus, can be decided by the number of input samples, i.e., we set it to N_1^2/M. φ_1 and φ_2 are the weights of sparse regularization, which are also decided by the given data. The value of φ_1/φ_2, in particular, is the weight of the grouping effect in the elastic net penalty, which should be large when the features are strongly correlated, and vice versa [37]. c is a parameter in discriminative manifold embedding that is used to control the minimization of the target class and background class; here, we set it to N_1/N_2. When there is only one unique target spectra available, we could simply set it as $c = 1$. It should be noted that without a full parameter cross-validation, we cannot guarantee the best performance of STME; however, the above-mentioned parameter setting procedure results in reasonable performance for target detection applications.

Finally, we provide the computational complexity of the proposed STME. Based on an analysis of the proposed procedure, the full input data matrix of STME is denoted by $X \in R^{L\times M}$. In discriminative manifold embedding, the time complexity of calculating the matrix G is $O\left(N_1 M^2\right)$. In transfer regularization, the time complexity of calculating the covariance matrix and eigen decomposition is $O\left(M^3\right)$. In the solution of STME, according to the LARS algorithm, the time complexity for LARS is $O(ML)$. Taking all the above parts into account, the total time complexity for STME is $O\left(ML + M^3\right)$. For the space cost, the maximum size of the matrices we need to process in the STME algorithm is a matrix G of size $M \times M$ in discriminative manifold embedding, a covariance matrix of size $L \times L$ in transfer regularization, and a matrix X^* of size $(M + L) \times M$ in the solution of STME. Therefore, the overall space complexity of STME is $O\left(M^2 + ML + L^2\right)$.

III. Experiments and Analysis

In this section, we first describe the three hyperspectral datasets used in the target detection experiments. We then demonstrate the effectiveness of the introduced sparse and transfer regularizations on the first dataset. Finally, we present the target detection results obtained from the three hyperspectral datasets, on both implanted targets and real-world targets.

A. Dataset Description

To demonstrate the capability of the proposed STME algorithm for detecting target pixels in a HSI, three different hyperspectral datasets were used for the target detection.

The first dataset was the target detection self-test dataset provided by the Rochester Institute of Technology [58], [59], which includes radiance and scaled reflectance images acquired by a HyMap airborne hyperspectral sensor at Cook City in MT, USA, on July 4, 2006. The image covers an area of 280 × 800 pixels, with 126 spectral channels in the VNIR–SWIR range. The ground spatial resolution of the dataset is about 3 m, and the spectral resolution is about 14 nm. The spectral channels around the wavelengths of 1320–1410 and 1800–1980 nm are the water-absorption bands and have been ignored in this experiment. This dataset is also equipped with the exact locations and SPL files of all the desired targets, so it is one of the standard datasets for hyperspectral detection algorithms.

The second dataset was an airborne his, which was acquired by the reflective optics system imaging spectrometer (ROSIS) at the urban test area in Pavia, northern Italy. The spectral coverage of the ROSIS-03 sensor ranges from 430 to 860 nm, and 102 spectral channels were analyzed after the removal of 13 noisy bands. The whole data set size is 1400 × 512 pixels, with a spatial resolution of 1.3 m per pixel. This image scene contains a number of ground-cover classes, and we aimed to detect the implanted target pixels with different implant fractions.

In the last experiment, detection analysis was performed on an airborne visible/infrared imaging spectrometer (AVIRIS) dataset from San Diego, CA, USA. The spatial resolution of image is 3.5 m per pixel. The image has 224 spectral channels in wavelengths ranging from 370 to 2510 nm, in which the wavelengths of 1350–1420 and 1810–1940 nm are the water-absorption bands and have been ignored in the experiment. This dataset is an urban scene in which there are three planes, including 58 target pixels as targets to be detected.

B. Effect of the Sparse and Transfer Regularizations

We first used the target detection self-test dataset to analyze the effect of the sparse and transfer regularizations. In this

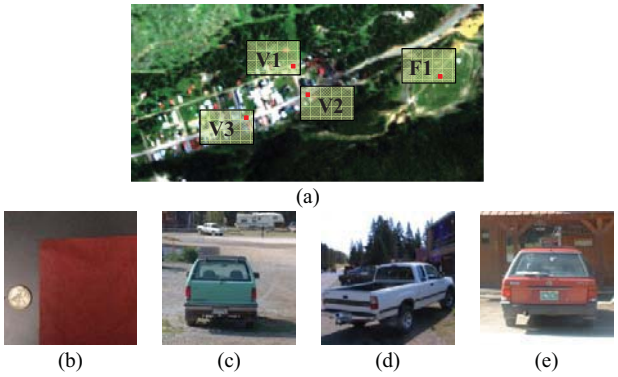

Fig. 2. (a) Hyperspectral dataset 1 (channels 14, 8, and 1 for RGB) and reference target locations. (b)–(e) Ground truth photos of targets F_1, V_1, V_2, and V_3, respectively.

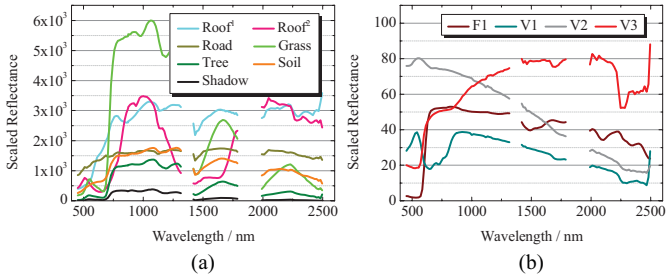

Fig. 3. (a) Background classes spectral signatures from HSI. (b) Target spectral signatures from the supplied SPL.

self-test dataset, there are seven real targets located in the image, comprising four fabric panels and three vehicles. The real locations of all the targets are provided by the dataset project [60]. In the experiment, we first took a fabric panel as target F1 to show the effect of the sparse and transfer regularizations in the STME algorithm. We then evaluated the target detection performance for all three vehicles, V_1, V_2, and V_3. In order to show the detection results more clearly and efficiently, we chose a subimage (dataset 1) of size 200×400 in the hyperspectral cube, which covers the real locations of F_1, V_1, V_2, and V_3. The locations of the four reference targets are highlighted in the hyperspectral dataset shown in Fig. 2(a), and the ground truth photos are shown in Fig. 2(b)–(e). The main background classes in the HSI are roof, road, soil, grass, tree and shadow, and the spectral signatures of both the background classes and targets are shown in Fig. 3(a) and (b). Note that target V_2 actually has two distinctly different spectral signatures, and we used the white spectral signature for analysis.

To evaluate the performance of the detection results for a wide range of specific targets, the target implant method [61], [62] was used. Here, we first consider target F1 as an example. All the implanted targets were located at the 20 implanted locations, as well as the real location, as shown in Fig. 4. A synthetic subpixel target spectra at the specified fraction f was generated by fractionally implanting the given target spectra in a located pixel of the background. We generated three test datasets with implant fractions of $f = 0.8$, 0.7, and 0.6, respectively. In each of the test datasets, we implanted

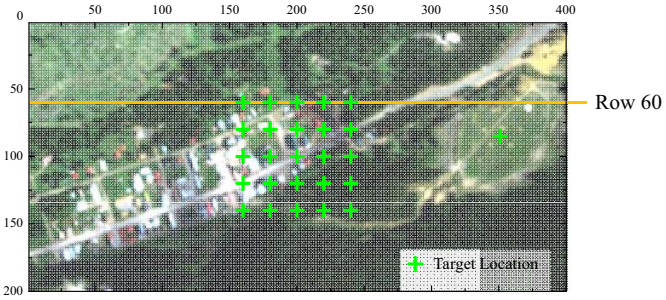

Fig. 4. Implanted target locations of F1 in dataset 1.

targets with a fixed implant fraction. Moreover, in order to consider the effect of random noise in the HSI [63]–[65], we added additive white Gaussian noise to all the test datasets. Different variance zero-mean Gaussian noise was added to each band, with the signal-to-noise ratio (SNR) given in (22) being randomly selected from 10 to 20 dB. The mean SNR value was 15 dB.

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\left(\sigma^2_{\mathrm{signal}}/\sigma^2_{\mathrm{noise}}\right) \tag{22}$$

in which $\sigma^2_{\mathrm{signal}}$ and $\sigma^2_{\mathrm{noise}}$ are variances of the signal and noise components, respectively.

We used the methods of manifold embedding (ME), transfer ME (TME), and STME for supervised dimensionality reduction, to verify the effect of the introduced sparse and transfer regularizations. In these machine learning methods, the prior target spectra is given as a unique positive sample, so c is set to 1, as mentioned above. We used the vertex component analysis (VCA) [66] algorithm to extract background endmembers as negative samples. In the VCA algorithm, we set the number of endmembers to 15, and compared each of the extracted endmember spectra to the given target spectra by SAM in order to guarantee that all the extracted spectra were background samples. If the SAM value between an endmember spectra and given target spectra was greater than 0.98, we deleted this endmember spectra from the negative samples. In the transfer regularization, the training samples coupled with another 800 unlabeled samples uniformly selected from all the pixels were used. For the sparse regularization parameter settings, we set the values to $\varphi_1 = 0.1$ and $\varphi_2 = 0.03$. Based on the Euclidean distance between the prior target spectral vector and an arbitrary test spectral vector in the reduced feature space, the threshold-based test statistic of the reciprocal of the Euclidean distance was adopted to derive the target detection result.

Fig. 5(a)–(c) show the detection test statistic transect plots of row 60 using the three methods on the test datasets with different implant fractions. As shown in Fig. 4, there are five implanted targets in this row. From Fig. 5(c), it can be seen that our proposed method suppressed the background pixels to a steady range, while the target pixels have a much higher statistical value, and can be easily separated. However, as shown in Fig. 5(a) and (b), the methods without sparse and transfer regularizations output the background pixels with greater fluctuations, in which some background pixels present high statistical values that would result in false alarms in

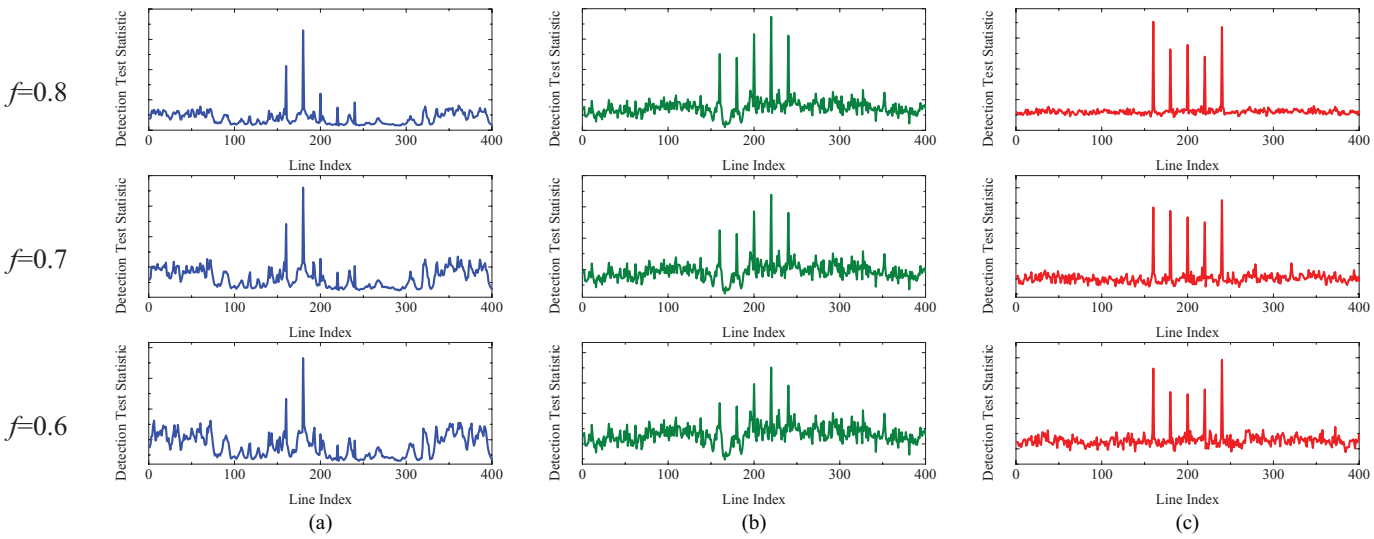

Fig. 5. Detection test statistic transect plots of row 60. (a) ME. (b) TME. (c) STME.

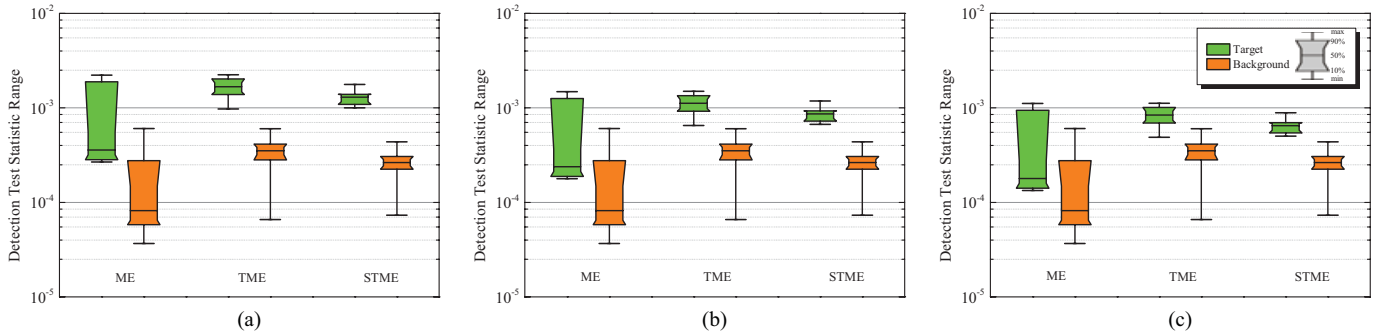

Fig. 6. Separability analysis of F_1, when (a) $f = 0.8$, (b) $f = 0.7$, and (c) $f = 0.6$.

the subsequent threshold segmentation. These target detection test statistic results suggest that the introduced sparse and transfer regularizations can avoid over-fitting when only very few training samples are provided.

To further investigate the separability between all the target and background pixels in the reduced feature space, we plotted the output target detection test statistic results range of the three methods in the test datasets with different implant fractions. In each of the subfigures of Fig. 6, there are three groups of boxes for ME, TME, and STME, respectively. Each group has a green box representing all the implanted target pixels, and an orange box representing the range of all the background pixels in the image. Each box provides the detailed value distribution of the detection test statistic results, as shown in the legend. The gap between the green box and the orange box in each group refers to the separability between the target and the background pixels in all the datasets. From Fig. 6, we can see that in different cases of implant fractions, STME shows the best separability between the target and background pixels, which reveals that the introduced sparse and transfer regularizations can preserve the discriminative information from training data to test data.

In order to investigate the effect of parameter d on the target detection performance, we made a detailed comparison of the detection results using ME, TME and STME with increasing values of d. The detection results were measured by the false alarm rate (FAR) under 100% detection. Fig. 7 shows the FARs under the three algorithms with different implant fractions. As can be seen in Fig. 7, STME performed better than the other algorithms for all of the d values. We also observed the same phenomenon in the following three implanted target detection experiments, i.e., targets V1, V2 and V3.

The above-mentioned analysis, from the target detection test statistic results, separability analysis, and subspace dimensionality comparison, proves the significant improvement brought about by the introduced sparse and transfer regularizations.

C. Target Detection Performance

In this part of the experimental section, we provide the target detection results for both implanted and real targets in three hyperspectral datasets. For the implanted target detection, we first focused on detecting the implanted targets of vehicles V_1, V_2, and V_3 in the HyMap dataset. We then looked at the detection of the simulated metal pixels in the ROSIS dataset. Receiver operating characteristic (ROC) curves [67] were used to evaluate the final detection performance. After that, in order to evaluate the performance of the proposed algorithm in a real-target scene, we have provided the target detection results of the three vehicles using a full self-test dataset and its supplied SPL files, the performance of which

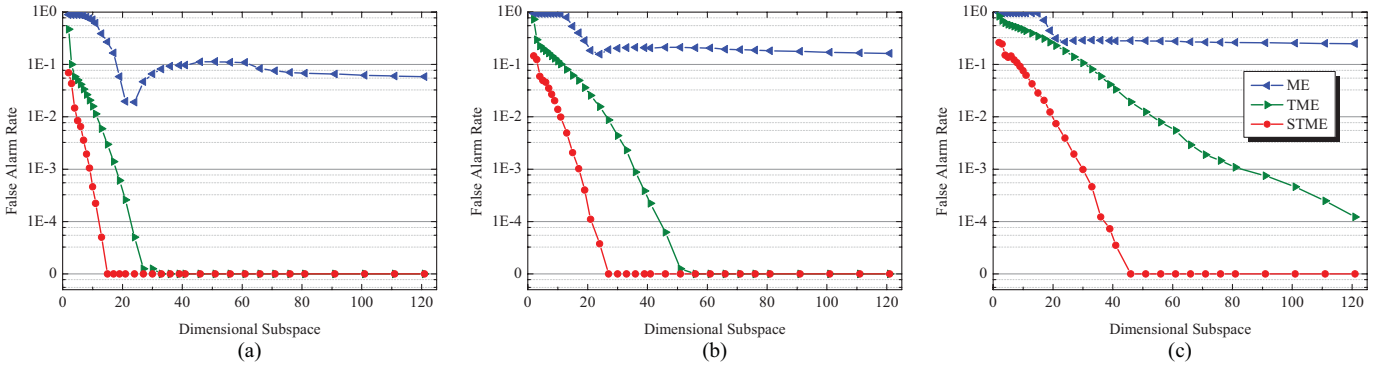

Fig. 7. Relationship of the d-dimensional subspace and FAR of F_1, when (a) $f = 0.8$, (b) $f = 0.7$, and (c) $f = 0.6$.

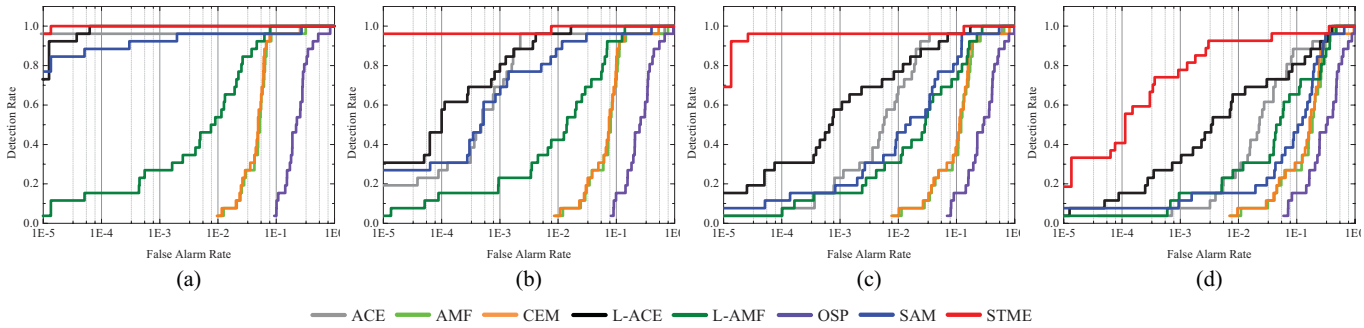

Fig. 8. ROC curves of V_1, when (a) $f = 0.7$, (b) $f = 0.6$, (c) $f = 0.5$, and (d) $f = 0.4$.

is measured by the FARs when desired targets are present. For the AVIRIS dataset, we aimed to detect the pixels of the aircraft in the real HSI. In both the implanted target and real-world target experiments, the following algorithms were employed for comparison: 1) ACE; 2) AMF; 3) CEM; 4) local ACE (L-ACE); 5) local AMF (L-AMF), which employs a local image to estimate background covariance in the ACE and AMF algorithms; 6) OSP; 7) SAM; and 8) the proposed STME. In all the detectors, we used the same given target spectra as the input *a priori* target spectra. In L-ACE and L-AMF, we adopted the double concentric window [9], in which the size of the small inner window was set according to the desired target size: for single (sub) pixel detection, the inner window was set to 3. The larger outer window was set from 15 to 25 to report the best performance. In OSP, we used the HySime algorithm [68] to estimate the background subspace, and the VCA algorithm to extract the background endmembers. Again, SAM was applied to check if target spectra were included within the background endmembers.

1) Experiments With Implanted Targets in the HyMap Dataset: Here, we report on the detection results of the implanted targets of vehicles V_1, V_2, and V_3 in dataset 1. For each target vehicle, we implanted 21 targets based on target spectra, with implant fractions of $f = 0.7, 0.6, 0.5$, and 0.4. The implanted locations included 20 implanted locations that were exactly the same as those shown in Fig. 4, and the real-target locations, as given in Fig. 2. Moreover, we also added the same level of random noise as in the test datasets.

Fig. 8 gives the ROC curves of the detection results, with different implant fractions, for target V_1 using the above-mentioned algorithms. In this paper, we obtained the ROC curves by computing the detection probability versus FAR. The algorithm with the best performance is indicated by a curve nearest to the upper left, which indicates the highest detection probability under the same FARs. The ROC curves of STME demonstrate a superior performance when compared to the other hyperspectral detection algorithms. The main reasons for this superior performance in detection are: 1) because of the essential attribute of discriminative ME, the proposed algorithm learns a subspace in which the target group and background group can be separated by the maximum possible amount, and 2) the introduced sparse and transfer regularizations help to obtain a more general and powerful DR model for the test samples. This is the key point in machine-learning-based hyperspectral detection with SSS and high-dimensional features. The experimental results also suggest that the proposed algorithm is effective for subpixel target detection. Generally, ACE and AMF always give a satisfactory performance in subpixel detection because both of the detectors are designed to detect the target spectral signature in the subspace, based on a linear mixing model [8]. The local versions of ACE and AMF, in particular, can follow the spatial variability of the background statistics in the scene to achieve a better detection performance. In this experiment, the proposed algorithm showed a clear advantage when compared with these other methods. The ROC curves in Fig. 8(a)–(d) reveal that the proposed STME algorithm outperforms the conventional hyperspectral detection methods

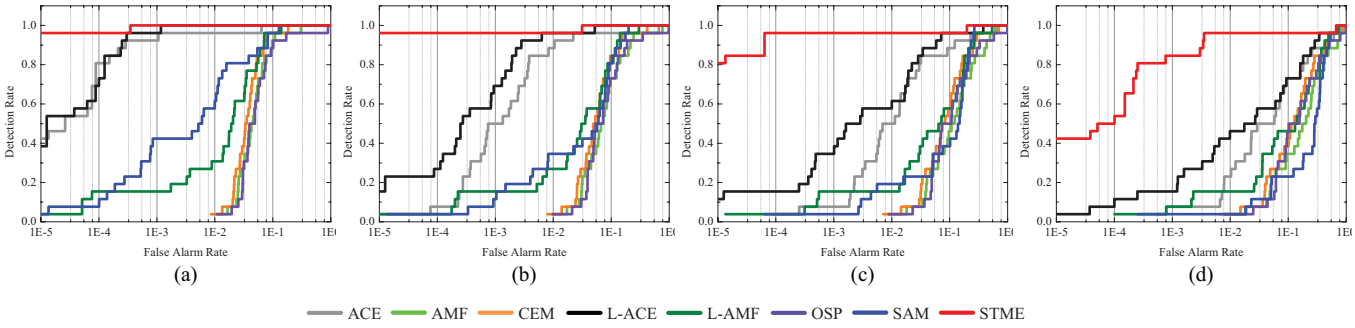

Fig. 9. ROC curves of V_2, when (a) $f = 0.7$, (b) $f = 0.6$, (c) $f = 0.5$, and (d) $f = 0.4$.

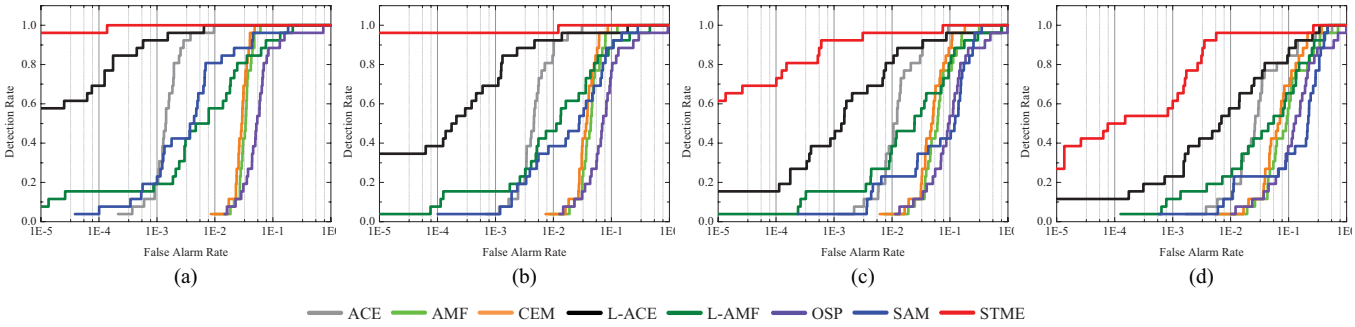

Fig. 10. ROC curves of V_3, when (a) $f = 0.7$, (b) $f = 0.6$, (c) $f = 0.5$, and (d) $f = 0.4$.

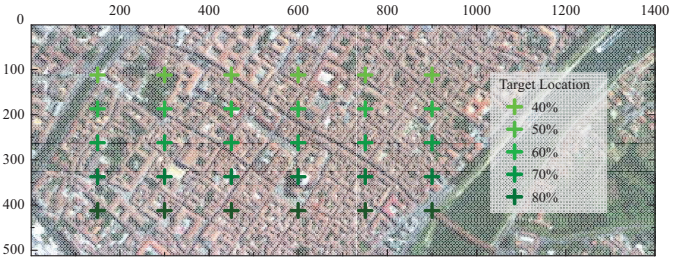

Fig. 11. Implanted target locations in the ROSIS dataset.

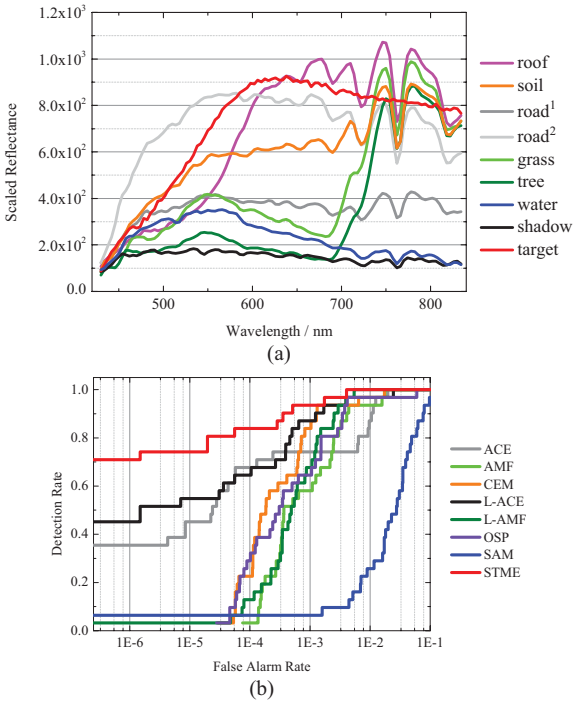

Fig. 12. (a) Target spectral signatures and background classes spectral signatures in the ROSIS dataset. (b) ROC curves of the detection results.

effectively and stably in the challenge of different implant fractions. We also evaluated the detection performance of targets V_2 and V_3, and the corresponding ROC curves are recorded in Figs. 9 and 10. The detection results are similar to the above analysis of V_1, and the proposed STME algorithm shows the best performance in the ROC curves of all the subfigures of the different implant fractions.

2) Experiments With Implanted Targets in the ROSIS Dataset: Here, we provide the detection results for the ROSIS dataset, which has a larger image size and more complicated background land-cover classes, as shown in Figs. 11 and 12.

The target of interest was a metal spectra acquired from the Johns Hopkins University (JHU) spectral library (see also the data in the ENVI spectral library: JHU library, Copper Metal 0692UUUCOP). For target implantation, we rescaled the target spectra to the image range and resampled it according to the HSI wavelength. We then implanted 30 simulated pixels with different fractions from $f = 0.8$ to $f = 0.4$, the reference locations of which are given in Fig. 11. As regards the parameter setting, we used the same parameter combination as in the HyMap dataset. The detection ROC curves of all the algorithms in Fig. 12(b) indicate that the proposed STME shows an excellent and robust detection performance.

3) Experiments With Real Targets in the Self-Test Dataset: The real-target detection performances of all the algorithms were evaluated using the full self-test dataset, with the same parameter setting strategy as before. In this experiment, the

TABLE I
FARs OF ALL THE ALGORITHMS WITH THE SELF-TEST DATASET

	V_1	V_2	V_3	Time (s)
ACE	2.57×10^{-2}	2.46×10^{-2}	2.47×10^{-1}	5.78
AMF	2.79×10^{-2}	8.00×10^{-2}	1.83×10^{-1}	2.71
CEM	2.14×10^{-2}	1.50×10^{-2}	1.99×10^{-1}	16.19
L-ACE	1.18×10^{-2}	1.23×10^{-2}	6.51×10^{-2}	2301.33
L-AMF	3.62×10^{-2}	6.87×10^{-2}	2.45×10^{-1}	2815.59
OSP	4.75×10^{-2}	1.02×10^{-1}	1.06×10^{-1}	4.48
SAM	3.65×10^{-2}	1.01×10^{-2}	9.45×10^{-2}	4.42
STME	$\mathbf{5.03 \times 10^{-3}}$	$\mathbf{7.86 \times 10^{-3}}$	$\mathbf{4.85 \times 10^{-2}}$	7.51

Fig. 13. (a) RGB composites of the AVIRIS dataset (channels 55, 33, and 19 for RGB). (b) Target locations in the AVIRIS dataset.

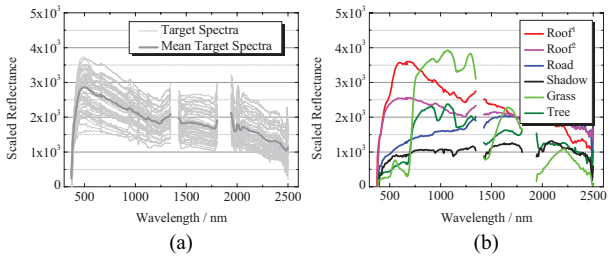

Fig. 14. (a) Target spectral signatures. (b) Background classes spectral signatures in the AVIRIS dataset.

project-supplied SPL files of V_1, V_2, and V_3 were used as the prior target spectra, while the ground truth provided with the self-test dataset was used to compute the target detection FAR, which is defined as the number of pixels having a test statistic value equal to or greater than the highest target pixel value divided by the total number of pixels in HSI.

The detailed detection performances are given in Table I, where we also show the average processing times for all the algorithms. It is evident from this table that the proposed STME algorithm gave the outstanding performance with all three vehicles in the real-target detection test. It can also be seen that the computational complexity of the STME algorithm is a little higher than ACE and AMF, but is much lower than the local versions of these detectors.

4) Experiment With the AVIRIS Dataset: Fig. 13(a) shows the RGB composites of the AVIRIS dataset which we adopted for the real-world target detection experiment. The detailed target locations are provided in Fig. 13(b). In this experiment, we used the mean value of the target pixels' spectra in the HSI as the prior target spectra for detection. Spectral signatures of both background classes and targets are shown in Fig. 14(a) and (b). This dataset is challenging for target detection because of the

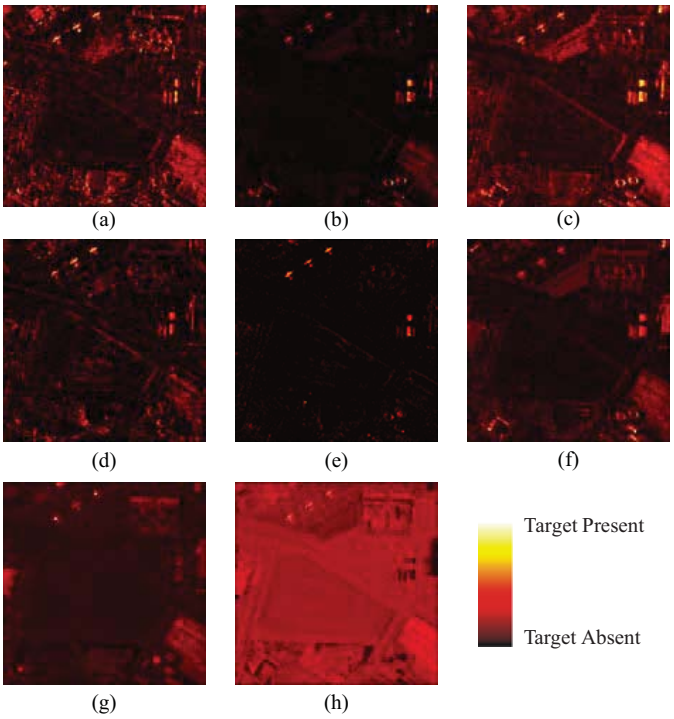

Fig. 15. 2-D plots of the detection test statistic results for the AVIRIS dataset. (a) ACE. (b) AMF. (c) CEM. (d) L-ACE. (e) L-AMF. (f) OSP. (g) SAM. (h) STME.

high spectral variability of the target and background classes, as shown in Fig. 14(a) and (b). There is no single spectral curve response representative of roof, and some pixels of roof are spectrally similar to the target pixels, so these roof pixels have a high probability of being misclassified as false alarms. For STME, we also selected 15 background endmembers from HSI as the negative samples, and another 400 unlabeled samples were used for transfer regularization. The detailed parameter settings were the same as we used in the HyMap dataset experiment. For all the detectors, for comparison, we used the same parameter settings as above. Note that the size of the typical target in this dataset is about 5×5 pixels; thus, in the L-ACE and L-AMF algorithms, the local background statistics estimation was obtained by setting the size of the small inner window to 11. Again, the larger outer window was set from 21 to 31 to report the best performance.

The 2-D plots of the detection test statistic results of all the comparison algorithms are shown in Fig. 15(a)–(h). From these figures, we can see that the proposed STME provided a distinguishable statistic map for further threshold segmentation. It could be seen that the other comparison algorithms also output high statistical values for the target pixels, but a few of roof pixels in the image also output high values, which would cause false alarms.

Based on the output statistic maps in Fig. 15(a)–(h), the ROC curves and the FARs under 100% detection for all the comparison algorithms are shown in Fig. 16(a) and (b). Furthermore, Fig. 17(a)–(h) show the 95% confidence regions [67] drawn around each estimated ROC curve using $N_T = 58$. It can be observed from Figs. 16 and 17 that the STME superiority in AVIRIS experiment is less significant

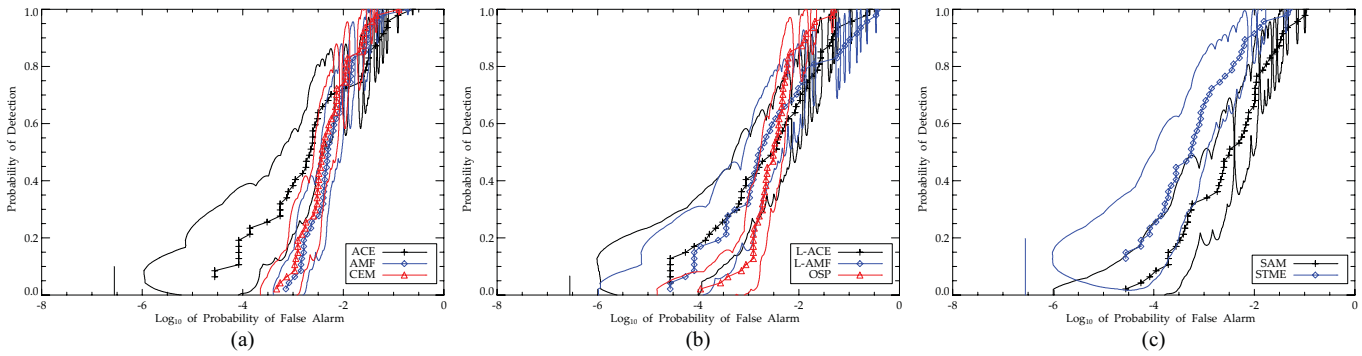

Fig. 17. Confidence intervals around ROC curves. (a) ACE, AMF, and CEM. (b) L-ACE, L-AMF, and OSP. (c) SAM and STME.

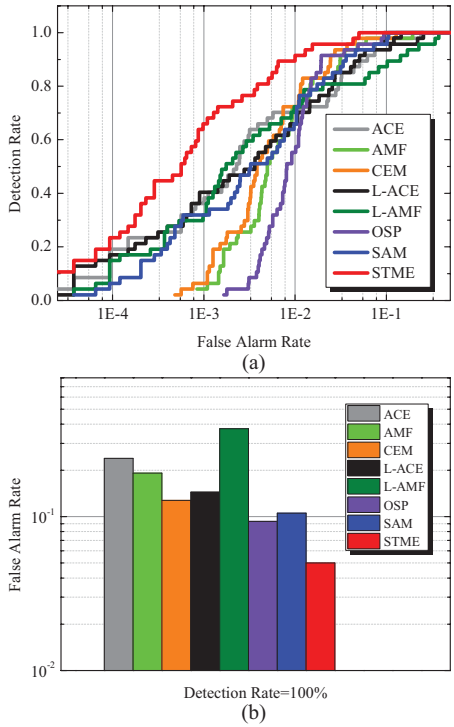

Fig. 16. (a) ROC curves, and (b) false alarm rate under 100% detection of the AVIRIS dataset.

than the implanted experiments. Just as discussed in a famous hyperspectral target detection paper [69], we can never evaluate that if a detection algorithm is always superiority with respect to other algorithms. Nevertheless, the proposed STME algorithm has showed the effective performance on all of the tested hyperspectral datasets.

IV. Conclusion

In this paper, the STME algorithm was presented as an effective and efficient feature extraction method for hyperspectral target detection. This algorithm utilized discriminative manifold embedding to learn a subspace, which can maximally separate the target samples from the background ones, based on the prior supervised information. In order to transfer the discriminative information from limited training samples to unlimited test data, we introduced transfer regularization into the optimization by preserving the dominant structure distribution of the samples. Meanwhile, by imposing an elastic net penalty on the DR projection matrix, the learned sparse representation of the transfer ME model can avoid the possible over-fitting in SSS learning.

Experiments on hyperspectral detection with three well-known datasets confirmed the performance of our algorithm. In comparison with the results of ME and TME, the proposed STME proved the significant improvement brought about by sparse and transfer regularizations. This can be seen in the detection test statistic results, separability analysis, and subspace dimensionality comparison. Moreover, the ROC curves and FAR reports also validate the performance of STME in both implanted and real-world hyperspectral target detection. Although the proposed approach works well, it does have some points that need to be addressed. Similarly to most machine learning methods, it is difficult to provide some criteria for estimating the various parameters, which weakens the robustness of the STME. Thus, in the future, we will focus on determining the optimal trade-off parameters automatically, which makes the proposed procedure applicable in practice conveniently.

Appendix

The full optimization of STME can be rewritten as

$$\begin{aligned}&\min_W \operatorname{tr}\left(W^T X G X^T W\right)\\&+\beta\operatorname{tr}\left[\left(P^T X - W^T X\right)\cdot\left(P^T X - W^T X\right)^T\right]\\&+\varphi_1\|W\|_1+\varphi_2\|W\|_2^2\\&=\min_W \operatorname{tr}\Big[W^T X\left(G+\beta\cdot I\right)X^T W-\beta P^T X X^T W\\&\qquad-\beta W^T X\left(P^T X\right)^T\Big]+\varphi_1\|W\|_1+\varphi_2\|W\|_2^2.\end{aligned}\tag{A1}$$

Here, we use Ω for simplicity

$$\Omega=\left(G+\beta\cdot I\right)/\beta.\tag{A2}$$

Then, (A1) is reduced to

$$\begin{aligned}&\min_W \operatorname{tr}\Big[\beta W^T X\Omega X^T W-\beta P^T X X^T W\\&\qquad-\beta W^T X\left(P^T X\right)^T\Big]+\varphi_1\|W\|_1+\varphi_2\|W\|_2^2.\end{aligned}\tag{A3}$$

We consider that Ω is symmetric, so we have

$$\Omega = V D V^T \tag{A4}$$

in which V and D are the eigenvector and eigenvalue matrices of Ω, respectively. Then, the first part of (A3) can be rewritten as

$$\begin{aligned}
&\beta \mathrm{tr}\Big[W^T X \left(V D V^T \right) X^T W \\
&\qquad - P^T X X^T W - W^T X \left(P^T X \right)^T \Big] \\
&= \beta \mathrm{tr}\Big[W^T X \left(V D^{1/2} \right) \left(D^{1/2} V^T \right) X^T W \\
&\qquad - P^T X \left(V D^{1/2} \right) \left(V D^{1/2} \right)^{-1} X^T W \\
&\qquad - W^T X \left(V D^{1/2} \right) \left(V D^{1/2} \right)^{-1} \left(P^T X \right)^T \Big] \\
&= \beta \left\| \left(V D^{1/2} \right)^{-1} \left(P^T X \right)^T - \left(D^{1/2} V^T \right) X^T W \right\|^2 \\
&\qquad - \beta \mathrm{tr}\left[\left(P^T X \right) \Omega^{-1} \left(P^T X \right)^T \right]
\end{aligned} \tag{A5}$$

The second part in (A5) is a constant item and can be ignored for the optimization. Then, we can further rewrite (A3) as

$$\begin{aligned}
&\min_W \beta \left\| \left(V D^{1/2} \right)^{-1} \left(P^T X \right)^T - \left(D^{1/2} V^T \right) X^T W \right\|^2 \\
&\quad + \varphi_1 \|W\|_1 + \varphi_2 \|W\|_2^2 \\
&= \min_{W^*} \left\| G^* - X^* W^* \right\|^2 + \varphi \left\| W^* \right\|_1
\end{aligned} \tag{A6}$$

in which

$$G^* = \begin{bmatrix} \beta^{1/2} \left(V D^{1/2} \right)^{-1} \left(P^T X \right)^T \\ 0^{L \times d} \end{bmatrix} \tag{A7}$$

$$W^* = \sqrt{1 + \varphi_2} \cdot W \tag{A8}$$

$$X^* = \frac{\beta^{1/2}}{\sqrt{1 + \varphi_2}} \cdot \begin{bmatrix} \left(D^{1/2} V^T \right) X^T \\ \sqrt{\varphi_2} \cdot I^{L \times L} \end{bmatrix} \tag{A9}$$

$$\varphi = \frac{\varphi_1}{1 + \varphi_2}. \tag{A10}$$

ACKNOWLEDGMENT

The authors would like to thank the handling editor and anonymous reviewers for their insightful and comments. The authors would also like to thank D. W. Messinger and J. P. Kerekes from the Rochester Institute of Technology for providing the hyperspectral target detection datasets. Dr. Lefei Zhang would also like to dedicate this paper to his mother, in memory of her love forever.

REFERENCES

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

[2] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 110–122, Sep. 2009.

[3] A. Plaza, J. Plaza, A. Paz, and S. Sánchez, "Parallel hyperspectral image and signal processing," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 119–126, May 2011.

[4] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, USA: Academic, 2003.

[5] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001.

[6] J. P. Kerekes and J. E. Baum, "Spectral imaging system analytical model for subpixel object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 5, pp. 1088–1101, May 2002.

[7] M. S. Stefanou and J. P. Kerekes, "Image-derived prediction of spectral image utility for target detection applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1827–1833, Apr. 2010.

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

[9] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.

[10] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York, USA: Springer-Verlag, 2006.

[11] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10, no. 6, pp. 853–863, Nov. 1974.

[12] 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, Jan. 1992.

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

[14] 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, Sep. 1999.

[15] Q. Du, H. Ren, and C.-I. Chang, "A comparative study for orthogonal subspace projection and constrained energy minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1525–1529, Jun. 2003.

[16] C.-I. Chang, and D. C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1144–1159, May 2000.

[17] C.-I. Chang, "Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.

[18] Q. Du and I. Kopriva, "Automated target detection and discrimination using constrained kurtosis maximization," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 38–42, Jan. 2008.

[19] S. Matteoli, N. Acito, M. Diani, and G. Corsini, "An automatic approach to adaptive local background estimation and suppression in hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 790–800, Feb. 2011.

[20] A. Schaum, "Hyperspectral anomaly detection beyond RX," *Proc. SPIE*, vol. 6565, p. 656502, May 2007.

[21] W. Sakla, A. Chan, J. Ji, and A. Sakla, "An SVDD-based algorithm for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 384–388, Mar. 2011.

[22] 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.

[23] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.

[24] 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.

[25] H. Kwon and N. M. Nasrabadi, "Kernel spectral matched filter for hyperspectral imagery," *Int. J. Comput. Vis.*, vol. 71, no. 2, pp. 127–141, Feb. 2007.

[26] 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.

[27] Y. Gu, C. Wang, B. Liu, and Y. Zhang, "A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 469–473, Jul. 2010.

[28] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.

[29] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[30] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.

[31] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[32] X. Huang, L. Zhang, and P. Li, "Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 260–264, Apr. 2007.

[33] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[34] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, Jul. 2006.

[35] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[36] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.

[37] T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: A unified framework for sparse dimension reduction," *Data Min. Knowl. Discovery*, vol. 22, no. 3, pp. 340–371, Jul. 2010.

[38] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 22, pp. 2323–2326, Dec. 2000.

[39] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 22, pp. 2319–2323, Dec. 2000.

[40] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Mar. 2003.

[41] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[42] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.

[43] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, Jan. 2002.

[44] X. Tian, D. Tao, and Y. Rui, "Sparse transfer learning for interactive video search reranking," *ACM Trans. Multimedia Computing, Commun., Appl.*, vol. 8, no. 3, p. 26, Jul. 2012.

[45] D. Tao, X. Tang, X. Li, and Y. Rui, "Direct kernel biased discriminant analysis: A new content-based image retrieval relevance feedback algorithm," *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 716–727, Aug. 2006.

[46] S. Si, D. Tao, and B. Geng, "Bregman divergence-based regularization for transfer subspace learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 7, pp. 929–942, Jul. 2010.

[47] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. Int. Conf. Mach. Learn.*, 2007, pp. 193–200.

[48] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: Transfer learning from unlabeled data," in *Proc. Int. Conf. Mach. Learn.*, 2007, pp. 759–766.

[49] Y. Luo, D. Tao, B. Geng, C. Xu, and S. J. Maybank, "Manifold Regularized Multi-task Learning for Semi-supervised Multi-label Image Classification," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 523–536, Feb. 2013.

[50] I. T. Jolliffe, *Principal Component Analysis*. New York, USA: Springer-Verlag, 2002.

[51] D. Cai, X. He, and J. Han, "Spectral regression: A unified approach for sparse subspace learning," in *Proc. 7th IEEE Int. Conf. Data Mining*, Oct. 2007, pp. 73–82.

[52] A. Neumaier, "Solving ill-conditioned and singular linear systems: A tutorial on regularization," *SIAM Rev.*, vol. 40, no. 3, pp. 636–666, Sep. 1998.

[53] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[54] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[55] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Ann. Stat.*, vol. 32, no. 2, pp. 407–499, 2004.

[56] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2008.

[57] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Statist. Soc. B*, vol. 67, no. 2, pp. 301–320, Apr. 2005.

[58] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a web-based application to evaluate target finding algorithms," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 2. Jul. 2008, pp. 915–918.

[59] S. Matteoli, E. J. Ientilucci, and J. P. Kerekes, "Operational and performance considerations of radiative-transfer modeling in hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1343–1355, Apr. 2011.

[60] *Target Detection Blind Test*. (2009) [Online]. Available: http://dirsapps.cis.rit.edu/blindtest/

[61] M. S. Stefanou and J. P. Kerekes, "A method for assessing spectral image utility," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 6, pp. 1698–1706, Jun. 2009.

[62] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly detection in hyperspectral images based on an adaptive support vector method," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 646–650, Jul. 2011.

[63] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Mar. 2011.

[64] M. L. Uss, B. Vozel, V. V. Lukin, and K. Chehdi, "Local signal-dependent noise variance estimation from hyperspectral textural images," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 3, pp. 469–486, Jun. 2011.

[65] N. Acito, M. Diani, and G. Corsini, "Signal-dependent noise modeling and model parameter estimation in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2957–2971, Aug. 2011.

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

[67] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, Apr. 2008.

[68] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[69] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Is there a best hyperspectral detection algorithm?" *Proc. SPIE*, vol. 7334, p. 733402, Jun. 2009.

Lefei Zhang (S'11) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2008, where he is currently pursuing the Ph.D. degree in the State Key Laboratory of Information Engineering in surveying, mapping, and remote sensing.

His current research interests include hyperspectral data analysis, high-resolution image processing, and pattern recognition in remote sensing images.

Dr. Zhang is a reviewer of several journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE SIGNAL PROCESSING LETTERS, *Information Sciences, Pattern Recognition*, and *Pattern Analysis and Applications*.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project from 2011 to 2016 appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 260 research papers. He holds five patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the IEE, Executive member (Board of Governor) of the China National Committee of International Geosphere-Biosphere Programme, Executive member of the China Society of Image and Graphics, etc. He regularly serves as a Cochair of the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and Geoinformatics Symposiums. He serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-spatial Information Science*, and the *Journal of Remote Sensing*. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Dacheng Tao (M'07–SM'12) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, the M.Phil. degree from The Chinese University of Hong Kong, Hong Kong, and the Ph.D. degree from the University of London, London, U.K.

He is a Professor of computer science with the Centre for Quantum Computation and Information Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. He mainly applies statistics and mathematics for data analysis problems in data mining, computer vision, machine learning, multimedia, and video surveillance. He has authored and co-authored more than 100 scientific articles at top venues, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the *Neural Information Processing Systems*, the International Conference on Machine Learning, The Conference on Uncertainty in Artificial Intelligence, International Conference on Artificial Intelligence and Statistics, International Conference on Data Mining, International Joint Conference on Artificial Intelligence, Association for the Advancement of Artificial Intelligence, IEEE International Conference on Computer Vision and Pattern Recognition, European Conference on Computer Vision, ACM Transactions on Knowledge Discovery from Data, Multimedia, and International Conference on Knowledge Discovery and Data Mining.

Prof. Tao was the recipient of the Best Theory/Algorithm Paper Runner Up Award in IEEE ICDM'07.

Xin Huang received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2009.

He is currently an Associate Professor with the LIESMARS, Wuhan University. He has published more than 30 peer-reviewed articles in international journals, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the *Photogrammetric Engineering and Remote Sensing*, and the *International Journal of Remote Sensing*. His current research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition, and remote sensing applications.

Dr. Huang has served as a Reviewer for most of the international journals for remote sensing. He was the recipient of the Top-Ten Academic Star of Wuhan University in 2009. In 2010, he received the Boeing Award for the Best Paper in Image Analysis and interpretation from the American Society for Photogrammetry and Remote Sensing. In 2011, he was the recipient of the New Century Excellent Talents in University from the Ministry of Education of China. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society as the Best Reviewer of IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. In 2012, he was the recipient of the National Excellent Doctoral Dissertation Award of China.