Neurocomputing ■ (■■■■) ■■■–■■■

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Unsupervised transfer learning for target detection from hyperspectral images

Bo Du [a,*], Liangpei Zhang [b], Dacheng Tao [c], Dengyi Zhang [a]

[a] School of Computer Science, Wuhan University, China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, China
[c] Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 23 December 2011
Received in revised form
16 July 2012
Accepted 24 August 2012

Keywords:
Hyperspectral images
Transfer learning
Target detection
Segmentation

ABSTRACT

Target detection has been of great interest in hyperspectral image analysis. Feature extraction from target samples and counterpart backgrounds consist the key to the problem. Traditional target detection methods depend on comparatively fixed feature for all the pixels under observation. For example, RX employs the same distance measurement for all the pixels. However, the best separation results usually come from certain targets and backgrounds. Theoretically, they are the purest targets and backgrounds pixels, or the constructive endmembers in the subspace model. So using those most representative pixels' feature to train a concentrated subspace is expected to enhance the separability between targets and backgrounds. Meanwhile, applying the discriminative information from these training data to the large testing data which are not in the same feature space and with different data distributions is a challenge. Here, the idea of transfer learning from interactive annotation technique in video is employed. Based on the transfer learning frame, several points are taken into consideration and the proposed method is named as an unsupervised transfer learning based target detection (UTLD) method. Firstly, the extreme target and background pixels are generated from robust outlier detection, providing the input for target samples and background samples in transfer learning. Secondly, pixels are calculated from the root points in a segmentation method with the purpose to preserve the most distribution feature of the backgrounds after reduced dimension. Thirdly, sparse constraint is imposed into the transfer learning procedure. With this constraint, a simpler and more concentrated subspace with clear physical meaning can be constructed. Extensive experiments reveal the performance is comparable to the state-of-art target detection methods.

1. Introduction

Automatic target detection from remote sensing images has been of great interest for years [1–3]. It is of particular importance in many domains, especially to military application. Spectral imaging reveals ground objects with fine resolution so as to explore the minor spectral difference between those visually undistinguished ones. Thus much attention has been paid to automatic target detection for hyperspectral images (HSI) [4–10].

The foremost key to targets detection from the HSI is the targets' features. Spectral features are most widely used feature in state-of-art target detection methods [4], where it is assumed that targets present a diagnostic difference from other background objects by means of spectral. Several kinds of methods are developed on this basis, such as linear mixture model based method and subspace based method. Linear mixture models construct the mixture composition of each pixel by endmembers, promising for detecting sub-pixel targets [5]. Subspace based methods present good performance on suppressing the pixels lying on background subspace and outburst those target signals [5,6]. The above two kinds of methods both employ physical model, requiring prior information about targets. Still other kind depends on the statistic model, with no prior knowledge about target, and it is called the unsupervised one (including anomaly detection). [7,8].

In these conventional methods, targets pixels and background pixels are manually chosen and then used to construct subspace based detector, where the target pixels and the background pixels are assumed to be separable. However, the number of the training pixels are usually limited and the correspondingly constructed subspace may over-fit the training pixels so as not to accurately detect the rest target pixels [1]. Can we find some way to preserve the discriminative information and avoid the over-fitting simultaneously? Transfer learning has shown its good performance to learn a subspace from limited samples [11,12], so it is thus introduced in target detection from hyperspectral images in this paper. Focusing on exploiting the training samples' discriminative information and learning a proper subspace from both training

* Corresponding author.
E-mail address: gunspace@163.com (B. Du).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.08.056

targets/background samples and those unlabeled samples, we have make several contributions for hyperspectral target detection in the paper:

1) A multivariate outlier analysis is used to automatically choose certain target pixels and background pixels as positive training and negative training samples, respectively. Meanwhile, existing target detection methods mainly depend on manually selecting pixels as training samples.
2) A segmentation method is employed to get the most representative and informative unlabeled samples. In this way, the wealthy continuous spatial feature in hyperspectral images can be fully considered. Existing methods usually randomly select the unlabeled samples or the all the samples in the image are used to learn a detector, like the constrained energy minimization [2] and adaptive matched subspace filter [5].
3) With the training labeled samples (including positive and negative samples) and unlabeled samples, a transfer learning based subspace construction method is formulated. Where a pairwise discriminative analysis is used to enhance the target-background pixels' separability. Existing target detection methods for hyperspectral images mainly depend on the labeled samples to construct subspace,.

The remainder of this paper is organized as follows. Section 2 presents unsupervised coarse target recognition by multivariate outlier detection method. Section 3 details the segmentation to get representative root pixels. The coarse target and background samples and unlabeled background root pixels are merged in sparse transfer learning for dimension reduction in Section 4. Section 5 discusses the extensive experiments by target detection to the dataset with reduced dimension. Section 6 concludes the paper.

2. Unsupervised target and background feature extraction

In this section, a multi-variable outlier analysis is employed to obtain the positive and negative samples [13], which are necessary for the transfer learning. Here, positive samples actually refer to the training targets pixels in the image dataset, and negative samples refer to the training non-target/background pixels from the image dataset.

The main idea behind the multi-variable outlier analysis is to iteratively figure out the mean and covariance matrix of the background pixels from image dataset to construct an optimal Mahalanobis distance based detector and finally extract the probable outliers.

Step 1: Randomly select one third of the pixels in the image dataset as the initial basic subset. Each pixel is virtually a vector containing b' components according to the b' bands.

Step 2: Compute the mean vector and the covariance matrix using the initial basic subset as follows:

$$\overline{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_i \tag{1}$$

$$\mathbf{C} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1b'} \\ \cdots & & & \\ \sigma_{b'1} & \sigma_{b'2} & \cdots & \sigma_{b'b'} \end{bmatrix} \tag{2}$$

where $\sigma_{ij} = 1/M\sum_k(\mathbf{x}_{ik}-\mathbf{m}_i)(\mathbf{x}_{jk}-\mathbf{m}_j)$, $k=1,\ldots,M$, M is the number of pixels in the subset, $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$ are the means of the ith band and the jth band, respectively. $\boldsymbol{x}_{ik}$ is the value in the ith band of the kth pixel.

Step 3: Compute the Mahalanobis distance of each pixel vector in the image using the mean vector and the covariance matrix constructed above:

$$d_i = \sqrt{(\mathbf{x}_i-\overline{\mathbf{x}})^T\mathbf{C}^{-1}(\mathbf{x}_i-\overline{\mathbf{x}})}, i=1,\ldots M \tag{3}$$

Step 4: Set the threshold η, and the pixels in the image with a distance under the threshold η would be set as the new basic subset. Reed and Yu have shown that RX statistics under the null hypothesis have a Chi-square distribution [14]. A Mahalanobis-based detector usually has a Chi-square distribution with p degrees of freedom [13]. The threshold is defined as the square root of $1-\alpha$ percentile of the Chi-square distribution with p degrees of freedom. p equals the band number b' Since the basic subset contains only part of the pixels in the hyperspectral imagery, the square root is multiplied by the inflation factor, which is the same factor as in [15]:

$$\eta = \chi_{p,\alpha}c_{Npr} \tag{4}$$

where $\chi_{p,\alpha}$ is the square root of $1-\alpha$ percentile of the Chi-square distribution with p degrees of freedom, and $c_{Npr}=c_{Np}+c_{hr}$ is the inflation factor.

$$c_{hr} = \max\{0,(h-r)/(h+r)\} \tag{5}$$

$$h = (n+p+1)/2 \tag{6}$$

$$c_{Np} = 1+\frac{p+1}{n-p}+\frac{1}{n-h-p} \tag{7}$$

where r is the size of the current basic subset, n is the total number of the pixels in the image.

Step 5: Iterate Step 2 to Step 4 until the basic subset no longer changes.

Step 6: Nominate the pixels excluded by the final basic subset as outliers. The finial basic subset constitutes the background pixels.

The above procedure is the basic idea of BACON. It has been proved that BACON includes those pixels with a distance smaller than η to the basic subset in each iteration, but the number of iterations is usually small [15]. Besides, it searches for the most reasonable mean vector and the covariance matrix to detect the anomaly pixels by updating the basic subsets in the iterations. The final mean tends to drift toward the real center of non-outlying background pixels [7].

The following is the way we propose to choose positive and negative samples from BACON results. The outliers excluded from the dataset may contain the spectral anomalies, mainly the noisy pixels or the rare ones of no interest. In order to eliminate the spectral anomaly pixels, all the excluded pixels are further investigated. Half of the pixels with larger norms in them are discarded. The remaining pixels are used as positive samples. As to the background pixels, they are clustered into several groups by k-means [16]. The number of the groups is equal to three times of the number of positive samples. In each group, the pixel with the largest average distance from all the positive samples is chosen. All the chosen background pixels comprise the final negative samples. With the positive and negative samples, pair-wise discriminative information can be constructed, which will be detailed in Section 4.

3. Construction of unlabeled pixels

In this section, a segmentation and a subsequent manifold analysis are carried out to obtain the unlabeled samples, which are necessary for transfer learning. The purpose is to fully exploit the spatial feature and the manifold feature in hyperspectral images to get the proper unlabeled samples for transfer learning.

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056

As is pointed out in [17,18], only high quality unlabeled samples can correctly represent distribution of the data, which are used in the transfer learning procedure, avoiding the over-fitting problem. The most common practice is randomly sampling a subset of the points as unlabeled sample. However, owing to the dominance of big classes in statistics [19], samples may be too redundant for big classes and too scarce to model the special geometry structure of small classes. For this problem, efficient graph-based Image Segmentation is used [20], excessively segmenting the image into the small regions. It can reduce the sample size to characterize homogeneous region and increase the number of samples from heterogeneous regions. We consider the pixels in each small homogeneous region have higher similarity and can share the same label. One representative pixel in each region can be chosen as unlabeled samples.

Efficient graph-based image segmentation works directly on the data points in feature space, without first performing a filtering step, and uses a variation on single linkage clustering. To perform traditional single linkage clustering, a minimum spanning tree of the data points is first generated, from which any edges with length greater than a given hard threshold are removed. The connected components become the clusters in the segmentation. Some method eliminates the need for a hard threshold, replacing it with a data-dependent term instead [19].

let $G=(V, E)$ be a (fully connected) graph, with m edges and n vertices. Each vertex corresponds to a pixel in the hyperspectral image. The final segmentation will be $S=(C_1,\ldots,C_r)$ where C_i is a cluster of data points. The algorithm is:

1. Sort $E=(e_1,\ldots,e_m)$ such that$|e_t|\leq\leq|e_{t'}|\forall t<t'$
2. Let $S^o=(\{x_1\},\ldots\{x_n\})$ in other words each initial cluster contains exactly one vertex.
3. For $t=1,\ldots,m$
   (a) Let x_i and x_j be the vertices connected by e_t
   (b) Let $C_{x_i}^{t-1}$ be the connected component containing point x_i on iteration $t-1$ and $l_i=\max_{mst} C_{x_i}^{t-1}$ be the longest edge in the minimum spanning tree of $C_{x_i}^{t-1}$. Likewise for l_j.
   (c) Merge $C_{x_i}^{t-1}$ and $C_{x_j}^{t-1}$ if
   $$|e_t|<\min\left\{l_i+\frac{k}{|C_{x_i}^{t-1}|},l_j+\frac{k}{|C_{x_j}^{t-1}|}\right\} \quad (8)$$
   where k is a constant.
4. $S=S^m$

In contrast to single linkage clustering which uses a constant k to set the threshold on edge length for merging two components in Eq. (8), efficient graph-based segmentation uses a variable threshold. This threshold effectively allows two components to be merged if the minimum edge connecting them has a length smaller than either of the components' minimum spanning trees, plus a term $\tau=k/|C_{x_i}^{t-1}|$. As defined here, τ is dependent on a constant k and the size of the component. Note that on the first iteration, $l_i=0$ and $l_j=0$, and $|C_{x_i}^0|=1$ and $|C_{x_j}^0|=1$, so k represents the longest edge which will be added to any cluster at any time, $k=l_{\max}$. Also, as the number of points in a component increases, the tolerance on added edge length for new edges becomes tighter and fewer mergers are performed, thus indirectly controlling region size. However, it is possible to use any non-negative function forτwhich reflects the goals of the segmentation system. In our experiments, k is defined as 0.8 since with this value enough unlabeled samples can be obtained already. A smaller k would present more unlabeled samples. In order to alleviate computation burden, 0.8 is used for most hyperspectral images.

The common assumption of semi-supervised learning method is that the labeled and unlabeled samples are lying on same sub-manifold [17],. Thus, relevance relationship between the existing labeled dataset and unlabeled candidates is applied to discard irrelevant unlabeled samples. To achieve this goal, the above mentioned segmentation procedure is done to obtain the segmented regions. In each region, the two vertices first merged are named roots of the region. One of the roots in each region is chosen as unlabeled samples. Then, we seek the t spectrally nearest neighbors for each unlabeled sample in the image. For each unlabeled sample, if labeled samples exist among its neighbors, the relevance is exit. Thus this unlabeled candidate should be selected in the final unlabeled set. Otherwise, the unlabeled samples would be discarded. Note that the number of the neighbors t controls this relevance criterion. In practice, we should set a larger t to relax the constraint on this selection procedure so as tolerate more variability of the spectrum. We fix t as 20 in the experiments. At last, we will obtain the rest unlabeled samples on the segmentation image, which would be input into transfer learning procedure to preserve the distribution property.

4. Transfer learning for target detection

In this section, both the labeled targets/backgrounds samples and unlabeled samples obtained in the above two sections are used to construct a subspace for target detection by transfer learning theory. Then all the pixels in the dataset are projected into the subspace, where the targets and backgrounds pixels are assumed to lie apart.

4.1. Discriminative manifold embedding

Many manifold learning methods were proposed, with the aim to construct local geometric properties preserved low-dimensional representation of the original input feature, e.g., locally linear embedding (LLE) preserves the linear coefficients which are used for reconstruct a given measurement by its k nearest neighbors [21], isometric feature mapping (ISOMAP) preserves global geodesic distances of all pairs of measurements [22], Laplacian eigenmaps (LE) preserves proximity relationships by manipulations on an undirected weighted graph [23], and so on. Among them (LDA) explores the pair-wise discriminative information to maximize the distance between dissimilar pair and minimize the distance between similar pair. For target detection, a subspace after dimensionality reduction should be able to maximize the distance between target samples and background samples and minimize the distance within target samples and background samples respectively. In other words, 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 the above pair-wise discriminative dimension reduction problem could be formulated in a manner similar to the patch alignment framework. In the low-dimensional feature space, for each sample of target class t_i, we expect that the Euclidean distances between the given sample and the other samples in the target class are as small as possible, while distances between the given sample and the other samples in the background class are as large as possible, i.e.,

$$\min_W \sum_{j=1}^{N_1} ||W^{\mathrm{T}}t_i-W^{\mathrm{T}}t_j||^2 \quad (9)$$

$$\max_W \sum_{j=1}^{N_2} ||W^{\mathrm{T}}t_i-W^{\mathrm{T}}b_j||^2 \quad (10)$$

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056

Combine (9) and (10) together and introduce a trade-off parameter c to control the influence of the two parts:

$$\min_W \sum_{j=1}^{N_1} ||W^T t_i - W^T t_j||^2 - c \cdot \sum_{j=1}^{N_2} ||W^T t_i - W^T b_j||^2 \tag{11}$$

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

$$X^i = [t_i, t_1^i, t_2^i, \cdots, t_{N_1}^i, b_1^i, b_2^i, \cdots, b_{N_2}^i] \in \mathrm{R}^{L \times (N_1+N_2+1)} \tag{12}$$

in which t_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 sample pairs (t_i, t_j^i) and (t_i, b_j^i), respectively. The corresponding low-dimensional feature matrix of patch X^i is given by:

$$Y^i = W^T X^i \in R^{d \times (N_1+N_2+1)} \tag{13}$$

A coefficient vector is defined as:

$$\delta = \left[\underbrace{1, \cdots, 1,}_{N_1} \quad \underbrace{-c, \cdots, -c}_{N_2} \right] \tag{14}$$

Then, (11) could be simplified to following patch optimization of sample t_i:

$$\begin{aligned} &\min_W \sum_{j=1}^{N_1+N_2} \delta_{(j)} ||W^T X^i{}_{(1)} - W^T X^i{}_{(j+1)}||^2 \\ &= \min_W \mathrm{tr}\left(W^T X^i \begin{bmatrix} -e_{N_1+N_2}^T \\ I_{N_1+N_2} \end{bmatrix} \mathrm{diag}\,(\delta)[-e_{N_1+N_2} I_{N_1+N_2}] X^{iT} W \right) \\ &= \min_W \mathrm{tr}(W^T X^i G^i X^{iT} W) \end{aligned} \tag{15}$$

In (15), $\delta_{(j)}$ is the jth element in δ ($j=1, \ldots, N_1+N_2$), $X^i{}_{(j)}$ is the jth column in X^i ($j=1, \ldots, N_1+N_2+1$), and:

$$G^i = \begin{bmatrix} -e_{N_1+N_2}^T \\ I_{N_1+N_2} \end{bmatrix} \mathrm{diag}\,(\delta)[-e_{N_1+N_2} I_{N_1+N_2}] \in \mathrm{R}^{(N_1+N_2+1)\times(N_1+N_2+1)} \tag{16}$$

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

Summing up the patch optimizations of target samples $t_i (i=1, \ldots, N_1)$ presents the whole optimization of introduced discriminative manifold embedding. Since each patch X_i has its own coordinate system of the detailed samples by the definition in (12), directly assembling the patch optimizations given in (15) is not reasonable. A selection matrix is employed to align the all samples together into a consistent coordinate [24,4]. Assuming the coordinate of patch X^i is selected from the global coordinate, which is also the full input data matrix:

$$X = [t_1, t_2, \cdots, t_{N_1}, b_1, b_2, \cdots, b_{N_2}, u_1, u_2, \cdots, u_N] \in R^{L \times M} \tag{17}$$

where u_k ($k=1, \ldots, N$) is the kth unlabeled sample obtained from the Section 2. Then, X_i can be rewritten as:

$$X^i = XS^i \tag{18}$$

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

$$S^i_{(a,b)} = \begin{cases} 1, & \text{if } a = F^i b \\ 0, & \text{else} \end{cases} \tag{19}$$

where $F^i = [i, i_1, \cdots, i_{(N_1+N_2)}]$ is the index vector for samples in patch X^i. Then sum the patch optimizations of target samples to obtain the whole optimization of discriminative manifold embedding:

$$\begin{aligned} &\min_W \sum_{i=1}^{N_1} \mathrm{tr}(W^T X S^i G^i S^{iT} X^T W) \\ &= \min_W \mathrm{tr}\left(W^T X \left[\sum_{i=1}^{N_1} (S^i G^i S^{iT}) \right] X^T W \right) \\ &= \min_W \mathrm{tr}(W^T X G X^T W) \end{aligned} \tag{20}$$

in which,

$$G = \sum_{i=1}^{N_1} (S^i G^i S^{iT}) \tag{21}$$

4.2. Transfer regularization

One problem needs special attention is that the target samples (or the positive samples) are very rare due to the unsupervised coarse outlier detection and the background samples (or the negative samples) also present a small number. With these samples, a subspace is learned which could maximally separate the target samples from background ones only based on the prior given training information. The resulting subspace would present a bias to that spanned by limited training samples and the dominant structure distribution also changed. Transfer learning is employed to preserve the discriminative information from training data to testing data which are not in the same feature space and with different data distributions. Several types of transfer learning methods are developed including: (1) re-weight some supervised samples in the training domain for the testing domain [25]; (2) construct a feature representation with reduced difference between the training and testing domains and smaller error of classification models [26]; (3) find shared priors between training and testing domains models for transfer learning [27]. In this paper, the dominant structure distribution provided by a large number of unlabeled samples is introduced as regularization to find a subspace which is supposed to transfer the discriminative information from training data to training data with reduced the error of subsequence classification. The unlabeled samples are obtained in Section 3.

Principal Component Analysis (PCA) is a linear transformation to find principal components in accordance with the maximum variance of a data matrix, so that the dominant structure of the distribution could be well preserved in the subspace after such transformation. PCA is applied to the unlabeled samples and the resulting subspace is imposed into the discriminative manifold embedding obtained from unsupervised targets and backgrounds samples acquirement [28]. In detail, denote the PCA projection matrix by $P \in R^{L \times d}$, and the transfer regularization is aimed at minimizing the Euclidean distance between the unlabeled samples data matrix in objective subspace and that in the subspace obtained by PCA, i.e.,

$$\min_W ||P^T X - W^T X||^2 \tag{22}$$

For the optimization problem formulation, a trade-off parameter β is introduced and the optimization of discriminative manifold embedding (20) and transfer regularization (22) is combined:

$$\min_W \mathrm{tr}(W^T X G X^T W) + \beta ||P^T X - W^T X||^2 \tag{23}$$

The solution of (23) provides a linear transformation matrix W in which each basis is a linear combination of all the original features, thus it is often difficult to interpret the results. Actual W may contain many zero elementaries since the sample number is low and the dimension of HSI is high. The sparse constraint is thus employed, which may be promising in two points: reveals an explicit relationship between the output feature representation and the given variables; and decreases the variance brought by possible over-fitting with the least increment of the bias to get more generalization to the model. In the sparse formulation, the key is to control the number of nonzero elements of the projection

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056

matrix W, characterized by the l^0-norm of the matrix W:

$$\min_W F(W)+||W||_0 \tag{24}$$

where $F(W)$ is the optimizations of discriminative manifold embedding and transfer regularization defined in (24), and l^0-norm $||W||_0$ is simply the number of nonzero elements in W. However, the l^1-norm of the projection matrix, i.e., lasso (least absolution shrinkage and selection operator), is usually used as a relaxation of the l^0 formulation to avoid an NP-hard problem [29]:

$$\min_W F(W)+||W||_1 \tag{25}$$

The convention solution to (18) is Least angle regression (LARS) [30] which searches the optimal solution of the lasso penalized linear regression problem [31]. However, the lasso penalty still has the following two restrictions [32]: (1) the number of selected features should be smaller than the number of given samples (2) the correlations between the input features should be low. The key is combining the l^1-norm and l^2-norm by the elastic net:

$$\min_W \mathrm{tr}(W^{\mathrm{T}}XGX^{\mathrm{T}}W)+\beta||P^{\mathrm{T}}X-W^{\mathrm{T}}X||^2+\varphi_1||W||_1+\varphi_2||W||_2 \tag{26}$$

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

There are three parameters, β, φ_1 and φ_2, in the objective function (26) and one parameter c in the coefficient vector (11). In this paper, in order to avoid cross-validation, we manually 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 around 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. c is a parameter in discriminative manifold embedding that is used to control the minimization of the target class and background class. We set it around N_2/N_1.

4.3. The solution to the problem

The objective function (26) is of a quadratic form with l^1-norm penalty based on some formula derivation by some mathematical derivation. We rewrite the full optimization of STME to:

$$\begin{aligned}&\min_W \mathrm{tr}(W^{\mathrm{T}}XGX^{\mathrm{T}}W)+\beta[(P^{\mathrm{T}}X-W^{\mathrm{T}}X)(P^{\mathrm{T}}X-W^{\mathrm{T}}X)^{\mathrm{T}}]\\&\quad+\varphi_1||W||_1+\varphi_2||W||_2=\min_W \mathrm{tr}[W^{\mathrm{T}}X(G+\beta\cdot I)X^{\mathrm{T}}W-P^{\mathrm{T}}XX^{\mathrm{T}}W\\&\quad-W^{\mathrm{T}}X(P^{\mathrm{T}}X)^{\mathrm{T}}]+\varphi_1||W||_1+\varphi_2||W||_2\end{aligned} \tag{27}$$

here we use Ω for simplicity:

$$\Omega=(G+\beta I) \tag{28}$$

Then, (27) is reduced to:

$$\min_W \mathrm{tr}[W^{\mathrm{T}}X\Omega X^{\mathrm{T}}W-P^{\mathrm{T}}XX^{\mathrm{T}}W-W^{\mathrm{T}}X(P^{\mathrm{T}}X)^{\mathrm{T}}]+\varphi_1||W||_1+\varphi_2||W||_2 \tag{29}$$

Consider Ω is symmetric, so we have:

$$\Omega=VDV^{\mathrm{T}} \tag{30}$$

in which V and D are the eigenvector and eigenvalue matrix of Ω, respectively. Then, the first part of (29) can be rewritten as:

$$\begin{aligned}&\mathrm{tr}[W^{\mathrm{T}}X(VDV^{\mathrm{T}})X^{\mathrm{T}}W-P^{\mathrm{T}}XX^{\mathrm{T}}W-W^{\mathrm{T}}X(P^{\mathrm{T}}X)^{\mathrm{T}}]\\&\quad=\mathrm{tr}[W^{\mathrm{T}}X(VD^{1/2})(D^{1/2}V^{\mathrm{T}})X^{\mathrm{T}}W-P^{\mathrm{T}}X(VD^{1/2})(VD^{1/2})^{-1}X^{\mathrm{T}}W\\&\quad-W^{\mathrm{T}}X(VD^{1/2})(VD^{1/2})^{-1}(P^{\mathrm{T}}X)^{\mathrm{T}}]\\&\quad=||(VD^{1/2})^{-1}(P^{\mathrm{T}}X)^{\mathrm{T}}-(D^{1/2}V^{\mathrm{T}})X^{\mathrm{T}}W||^2-\mathrm{tr}[(P^{\mathrm{T}}X)\Omega^{-1}(P^{\mathrm{T}}X)^{\mathrm{T}}]\end{aligned} \tag{31}$$

The second part in (31) is a constant item and can be ignored for optimization.

Then we can further rewrite (29) as:

$$\begin{aligned}&\min_W ||(VD^{1/2})^{-1}(P^{\mathrm{T}}X)^{\mathrm{T}}-(D^{1/2}V^{\mathrm{T}})X^{\mathrm{T}}W||^2+\varphi_1||W||_1+\varphi_2||W||_2\\&\quad=\min_{W^*}||G^*-X^*W^*||^2+\varphi||W^*||_1\end{aligned} \tag{32}$$

in which,

$$G^*=\begin{bmatrix}(VD^{1/2})^{-1}(P^{\mathrm{T}}X)^{\mathrm{T}}\\0^{L\times d}\end{bmatrix} \tag{33}$$

$$W^*=\sqrt{1+\varphi_2}\cdot W \tag{34}$$

$$X^*=\frac{1}{\sqrt{1+\varphi_2}}\cdot\begin{bmatrix}(D^{1/2}V^{\mathrm{T}})X^{\mathrm{T}}\\\sqrt{\varphi_2}\cdot I^{L\times L}\end{bmatrix} \tag{35}$$

$$\varphi=\frac{\varphi_1}{1+\varphi_2} \tag{36}$$

In other words, the full optimization of STME could be transformed to:

$$\min_{W^*}||G^*-X^*W^*||^2+\varphi||W^*||_1 \tag{37}$$

which can be efficiently solved by the LARS algorithm [12,13].

Then, based on (34), the optimal solution of STME is obtained by:

$$W=\frac{W^*}{\sqrt{1+\varphi_2}} \tag{38}$$

Finally, with the reduced dimension data, simple k-nn classification and a thresholding procedure is employed to distinguish targets from backgrounds. The whole flowchart of the proposed method is illustrated as Fig. 1.

5. Experiments and analysis

In this section, two real datasets are used to evaluate our proposed UTLD, and the effect of transfer learning and unlabeled samples selection to the target detection performance is investigated.

5.1. Data description

The first dataset is the HYDICE hyperspectral dataset obtained from an aircraft platform. This dataset covers an urban area and has a spectral resolution of 10 nm and a spatial resolution of 1 m. The image scene contains vegetation area, a construction area and several roads where some vehicles exist. The whole dataset has a size of 307 × 307 as shown in Fig. 2(a). However, the only definite ground truth is that in the upper-right of the scene; the anomaly targets are the cars and roofs embedded in different backgrounds, so we only use the dataset covering this area in our experiments. The sub-image used in the experiments and the anomaly targets' positions are shown in Fig. 2(b) and (c), respectively. The low SNR and water vapor absorption bands were also eliminated so that 160 bands remained.

The other dataset is the HYMAP hyperspectral dataset obtained from the blind test website [33], which has become a community standard dataset. It consists of a self-test image and a blind-test image. Due to the available of ground truth, the self-test image is used here. And the dataset about self-test image includes library of target spectral reflectance, regions of interest, and ground photos. It was acquired by the airborne HyMap hyperspectral sensor on the Cooke City, MT, US, in 2006. The spatial resolution is about 3 m

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056

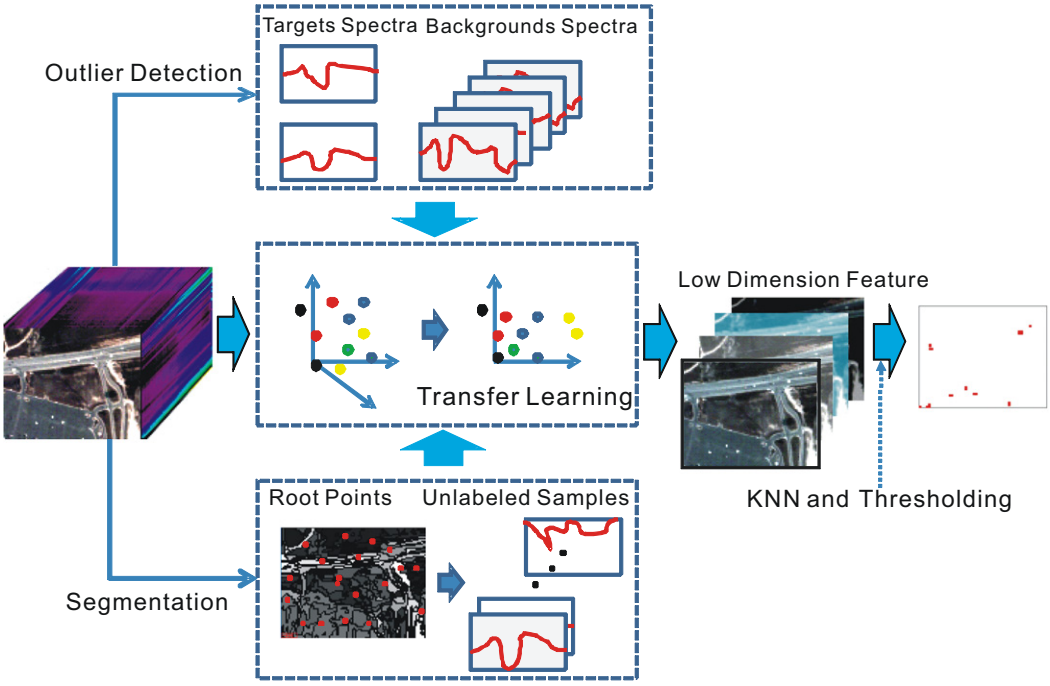

Fig. 1. Flowchart of UTLD.

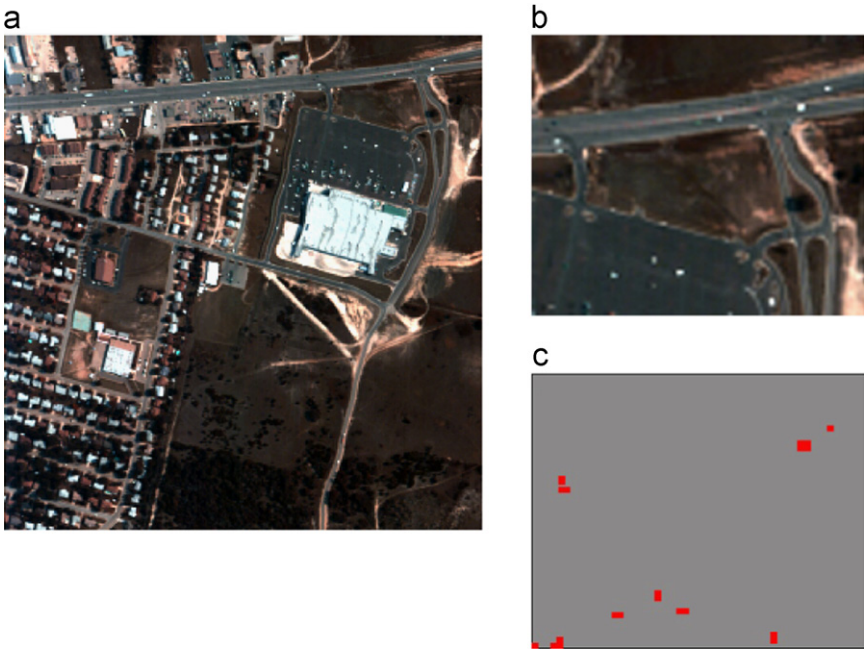

Fig. 2. HYDICE Hyperspectral Dataset.

and has a spectral bands number of 126 covering the VNIR-SWIR range (453 nm- 2496 nm). The ground scene covers the small town, and the near town forest and grass areas. Basically speaking, the area contains many kinds of ground objects, so the according image is complex. Several different kinds of targets are deployed, including fabric panels and vehicles. The ROIs datasets present their positioned distribution. As to our experiment, a subset image scene is chosen in order to focus our detection on it. The subset is shown in Fig. 3. It covers two kinds of panels with only F1 and F4 panels used as targets in our experiments, named panel 1, panel 2 and panel 3 respectively. Due to the spectral similarity, these panels are considered as the same kind of targets. Besides, due to the spectral mixture and target signal weakness in some boundary pixel of the panels, not all the ROI pixels are used as target detection reference. The chosen target pixels for detection reference are shown in Fig. 3(c), in which the fill factors are not too weak to present the target's feature. The detailed information about the targets' composition is presented in Table 1. The full pixels refer to the pixels composed of only target's signal, also named pure targets pixels. The sub-pixel pixels refer to the pixels composed of both target and background signals. Sub-pixel targets also termed as mixed pixels targets, which are the main difficulty in target detection from hyperspectral images [1,2].

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056

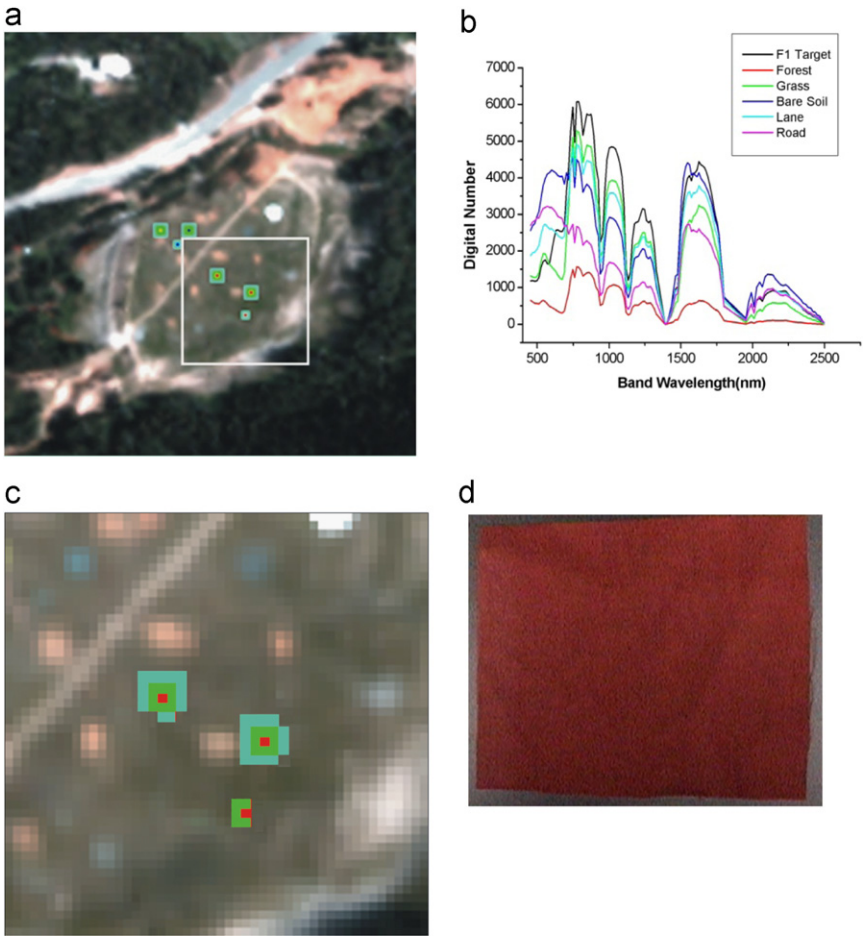

Fig. 3. HYMAP Hyperspectral Dataset.

Table 1
Composition of Targets Panels in HYMAP Dataset.

Targets name	Composed pixels' type	
	Full pixels	Sub-pixel pixels
Panel 1	10	13
Panel 2	10	14
Panel 3	1	5

5.2. Detection results and analysis

Several state-of-art unsupervised target detection methods are used as comparative methods, including local-RX [7], global-RX [14], CBAD (Cluster Based Anomaly Detector) [10]. Besides, several supervised ones are also employed so as to evaluate the performance of transfer learning with predefined positive and negative samples, including ACE (Adaptive Cosine Estimator) [2] and AMSD (Adaptive Matched Subspace Detector) [5], the classic unstructured and structured detector respectively [2]. For local-RX, three different window sizes are used: 13×13, 17×17, and 21×21, and only the best performance is chosen. In CBAD, the clustering information is chosen manually to provide best detection results. ACE only needs the targets' spectra, and the targets pixels distinguished by multiple variation outlier detection method are used. AMSD uses the same targets information and its background subspace is constructed by the eigen-vectors of the image's correlation matrix or the endmembers extracted from the image.

From the target detection method, a detection probability map is obtained, several detection results on HYDICE dataset is shown in Fig. 4, which are not so distinguishable from each other. Then, the last step in target detection methods is to segment the high probability value pixels as targets with an adaptive or subjective threshold [2]. Some recent work has been done on threshold learning method [34], which would be our future work. In order to be immune to the segmentation threshold, the receiver operating characteristic (ROC) curves are usually employed [2], since it provides a threshold-free performance comparison by means of continuous curves in the detection probability/false alarm domain [2]. Probability of detection in ROC curves refers to the ratio between the number of detected target pixels and the number of all the target pixels in the image. False alarms rate equals to the percentage of the detected non-targets pixels in the whole image. So the ROC curve lying nearer to the top-left in the coordinate space proves betters detection performance. The detection results by the above mentioned methods are shown in Fig. 5. Several conclusions can be drawn from those pictures. (1) UTLD does best among all the methods and achieves best detection performance soon in the low range of false alarm rate axis; (2) AMSD and ACE lie upwards the other unsupervised detection methods, so that targets information by our first procedure can be employed to enhance the detection performance; (3) Both CBAD and local-RX present better detection performance than global-RX, suggesting that a local feature may be more appreciate for these targets and the reason may be that their size are so small that only against a small compact of background pixels can those weak targets present a sufficient difference.

Since the detailed composition of targets is available for HYMAP dataset, further investigation is done so as to analyze its performance on those difficult targets pixels. Under the same false

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056

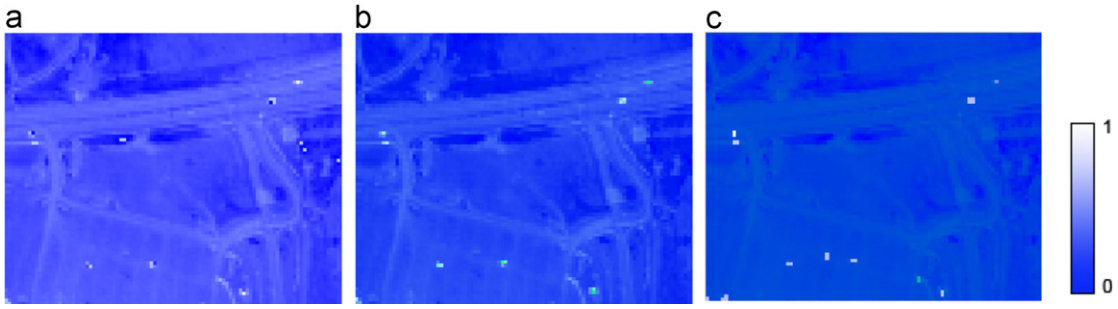

Fig. 4. Detection results before segmentation.

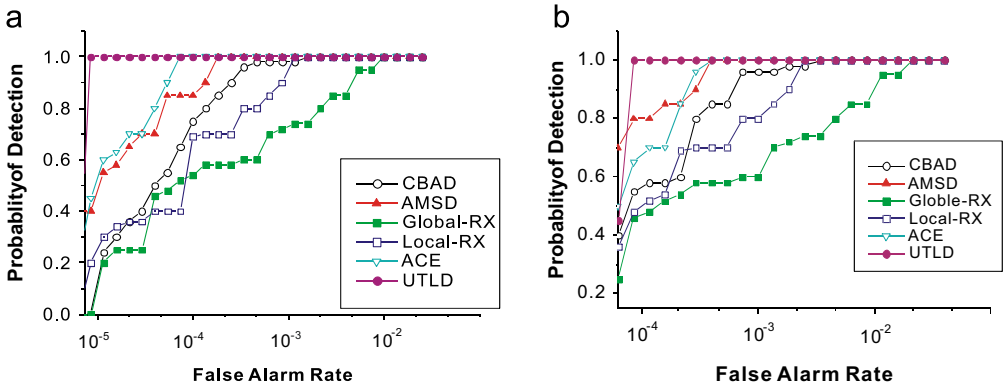

Fig. 5. Detection results by ROC curves.

alarm rate, the detected targets pixels numbers by these methods are shown in Fig. 6. In each picture, the detection results of full target pixels and sub-pixel target pixels are shown respectively. It is revealed that all the methods perform similarly on the full pixels. But as to the sub-pixel targets, the results are obviously different. (1) Mahalanobis distance based methods, including global-RX, local-RX and CBAD, do badly, presenting low number of detected ones; (2) Structured background method AMSD performs slightly better than unstructured background method ACE, suggesting that with more information from background samples, or the constructed background subspace, the separability between positive target samples and negative background samples can be retained. In other words, more information about background helps avoiding over-fitting problem; (3) UTLD outperforms the other state-of-the-art methods on the points that it keeps effective on all those sub-pixels targets, which is the main difficulty in target detection from hyperspectral remote sensing images [2].

The reasons for improved performance of UTLD compared with other state-of-the-art detection methods are listed here: (1) Firstly, the multivariate outlier analysis does perform well in distinguishing those targets of interest, since in most application, targets of interest in hyperspectral images are those man-made objects, which are spectrally distinguishing from most backgrounds objects. Furthermore, a subsequent manifold analysis is useful in choosing the proper backgrounds samples. (2) Secondly, the pair-wise discriminative constrains are imposed on the subspace learning procedure. So each target sample are constrained to be separated from the background samples. While in conventional methods, all the target samples are usually averaged to be a target spectrum in the detector construction. In other words, The target samples' discriminative information are more fully used in UTLD. (3) Limited number of labeled training samples usually cause small sample size (SSS) problem, especially for the high dimension feature extraction [35]. By sparsity based transfer learning, the SSS problem can be alleviated. By exploiting the unlabeled samples, the over-fitting can also be avoided. (4) In machine learning community, it is common that the unlabeled samples are randomly selected for dimension reduction or subspace learning methods. As to hyperspectral images, different land objects distribute on different area with continue spatial pattern, so selecting unlabeled samples from the image by considering majority land objects distribution pattern is necessary to get informative representative labeled samples.

The former two points are obvious, since outlier detection has proved effectual in separating those probable man-made targets [13] and traditional way of manually selecting samples is subjective. Besides, discriminative information based methods have successfully applied in image processing [36], including hyperspectral target detection [37]. So we just further investigate the latter two points with following experiments.

5.3. *Effect of transfer learning in hyperspectral target detection*

In order to evaluate the performance of sparsity based transfer learning introduced in our method, further experiments are done on the our proposed method, and one variant version without transfer learning and sparse constraint. The variant is named unsupervised manifold embedding detection (UMED) method, which just takes the training targets and backgrounds samples into consideration to learn a proper subspace, like the liner discriminative analysis [37]. So only the first item in (26) is used [38]. We firstly investigate the HYDICE dataset. Figs. 7 and 8 present the 3D plot of the detection results, which show that UTLD constrains the background to a steady value range while UMED presents more fluctuate backgrounds. Meanwhile, more targets are highlighted in UTLD than UMED. To further exploit the performance quantitatively, the targets value and background values in the two methods are both calculated in Fig. 9. In Fig. 9, all the values of the targets and those of backgrounds of UTLD can be easily separated, while there exists large overlap between targets and backgrounds of UMED. The reason is that UMED just use the training targets and backgrounds samples, which may over-fit for the rest samples. In other words, the transfer learning

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056

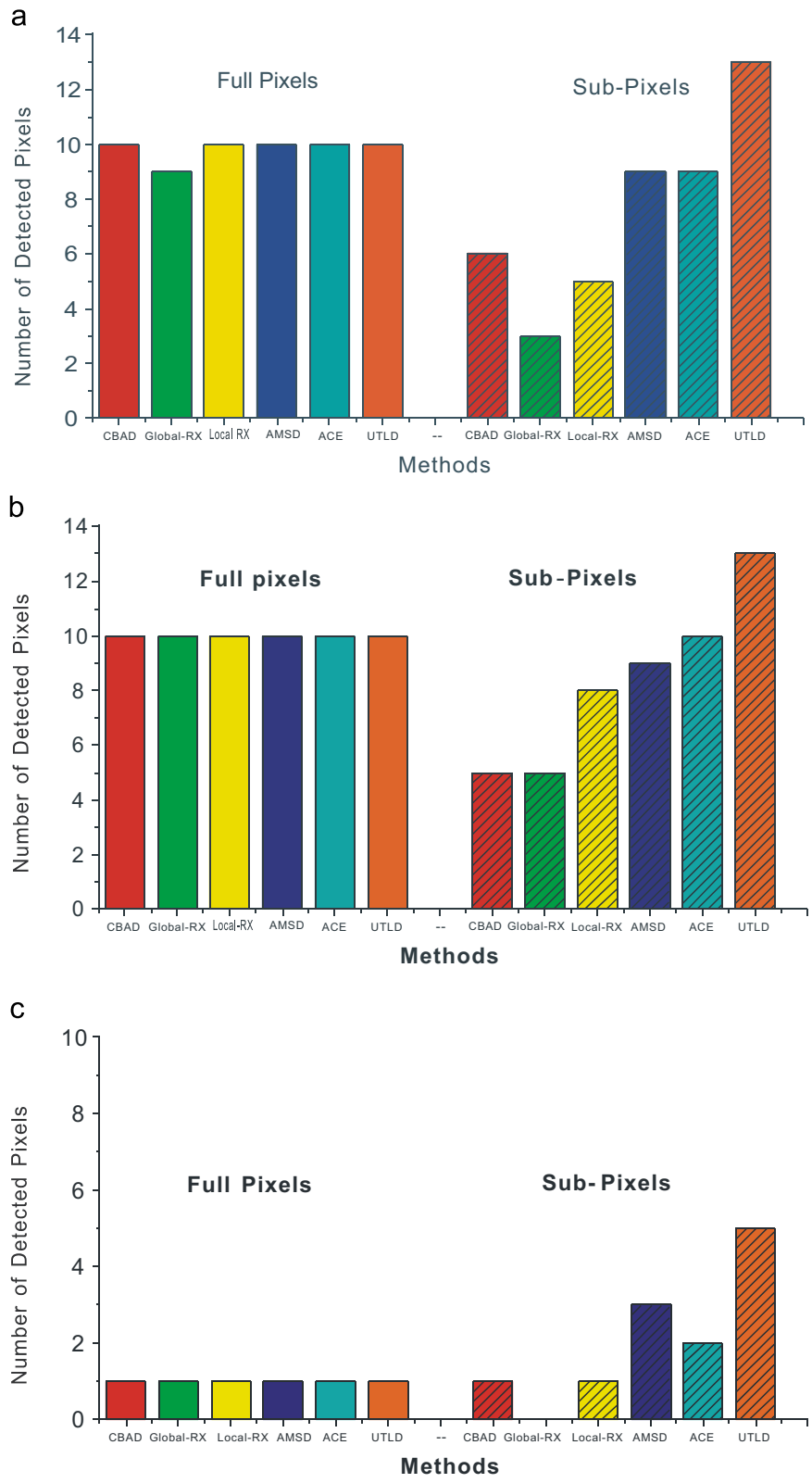

Fig. 6. Performances on different panels in HYMAP dataset.

strategy employs the whole dataset's distribution to avoid overfitting in learning a discriminative subspace. For HYMAP dataset, we also do the same comparison experiments and the results separability analyisi is shown in Fig. 10, which again present a better separability by introducing transfer learning in subspace construction for target detection.

5.4. Effect of unlabeled samples selection

(Figs. 11 and 12) As we have mentioned before, we use a graph-based segmentation procedure to get the proper unlabeled samples, aimed to fully exploit the spatial information in hyperspectral images. So we carry out comparative experiments with the

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056

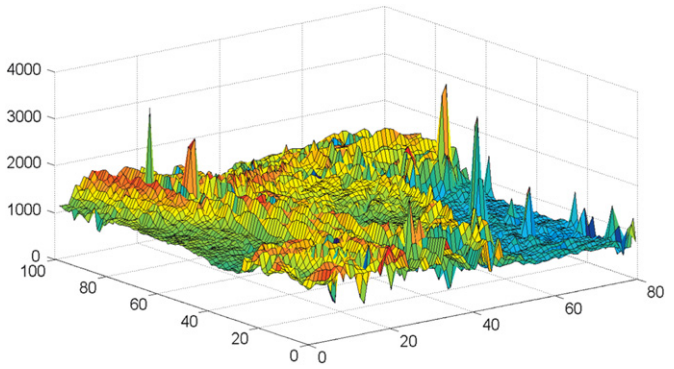

Fig. 7. Detection results of UMED.

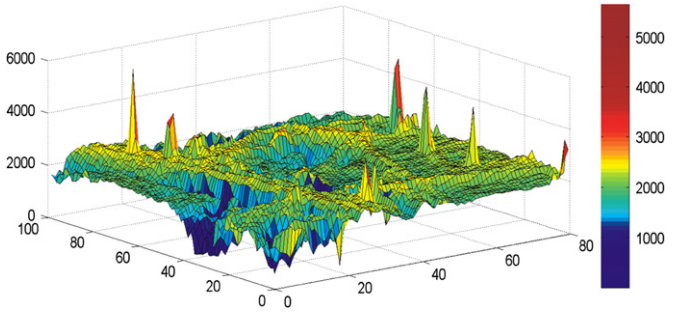

Fig. 8. Detection results of UTLD.

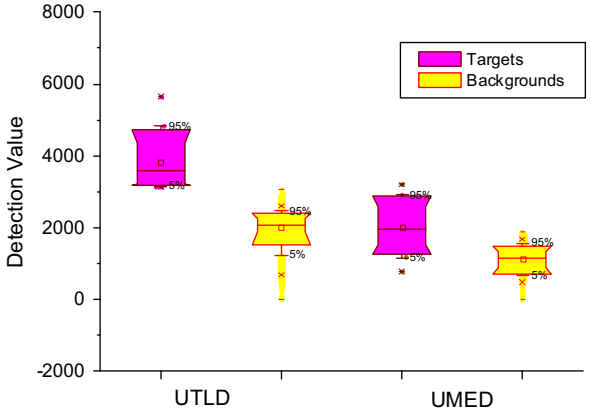

Fig. 9. Separability analysis of HYDICE dataset.

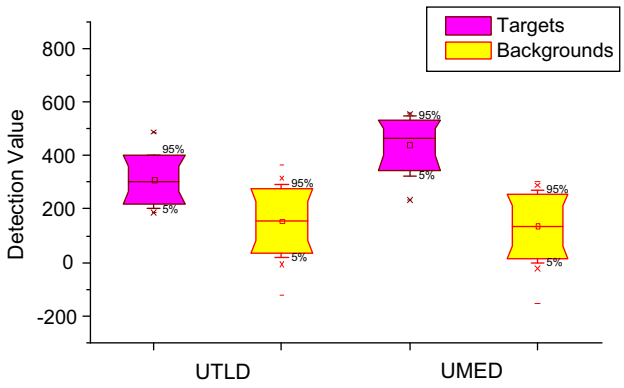

Fig. 10. Separability analysis of HYMAP dataset.

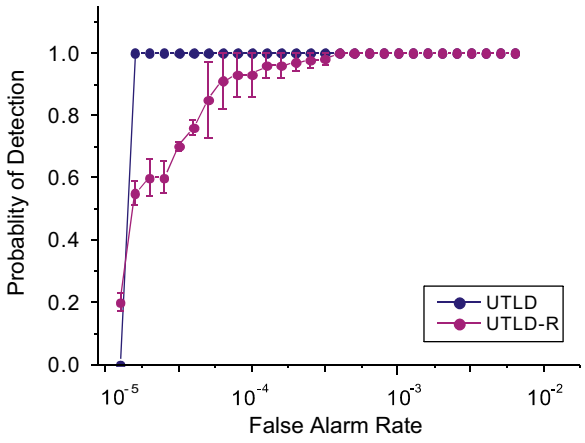

Fig. 11. ROC curves of UTLD and UTLD-R on HYDICE dataset.

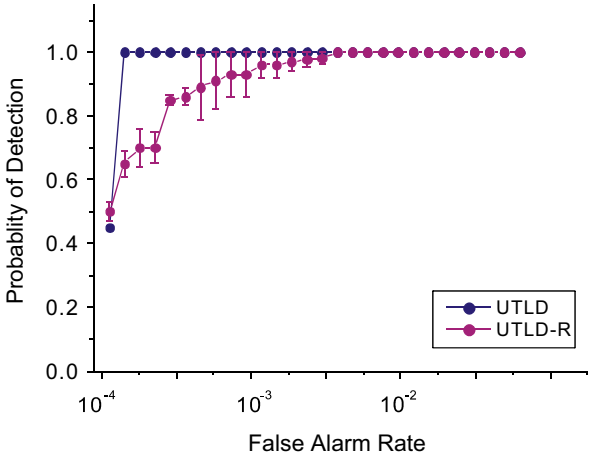

Fig. 12. ROC curves of UTLD and UTLD-R on HYMAP dataset.

conventional way of randomly selecting unlabeled samples for UTLD, named UTLD-R. We randomly select the same number of unlabeled samples with UTLD and then use the same way to learn the subspace for detection. We repeat the selection of unlabeled samples 30 times, since the unlabeled samples may be different each time and get different performances. For our UTLD, all the targets/backgrounds samples and unlabeled samples are fixed for a definite dataset, because the first two stages of UTLD focus on finding the most suitable labeled and unlabeled for the subspace learning. So UTLD's detection results are fixed. The ROC curves of UTLD and UTLD-R are shown in Fig. 10. In Fig. 10, the variances of detection performance of UTLD-R caused by different unlabeled samples are also shown. It is revealed that UTLD outperforms UTLD-R, especially at the low false alarm rate. The reason is that by elaborately selecting unlabeled samples, the learned subspace in UTLD takes more representative samples into consideration, so that the separability between targets and most backgrounds samples can be maximized. It concludes that combining the spatial distribution and the spectral feature is beneficial to get informative unlabeled samples for target detection from hyperspectral images.

6. Conclusion

In this paper, an unsupervised target detection method based on transfer learning is proposed. The more detectable targets are first separated from the backgrounds by its spectral difference with the surrounding, where a robust outlier detection method is employed. A segmentation method is used to get unlabeled samples. Both the detection and segmentation results provide

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056

samples for transfer learning, which construct a low-dimensional subspace to better separate targets from backgrounds. Experiments results with real-world hyperspectral remote sensing images show that this automatic targets detection perform better than the state-of-the-art methods, especially on those sub-pixel targets. Our future research focus is how to determine the optimal parameters automatically, so as to apply the proposed method to achieve an even better performance.

References

[1] C.-I. Chang, Hyperspectral Imaging: Spectral Detection and Classification, Kluwer, New York, 2003.
[2] D. Manolakis, G. Shaw, Detection algorithms for hyperspectral imaging applications, IEEE Signal Process. Mag. 19 (1) (2002) 29–43.
[3] L. Zhang, B. Du, Y. Zhong, Hybrid Detectors Based on Selective Endmembers, IEEE Trans. Geosci. Remote Sensing 48 (6) (2010) 2633–2646.
[4] C.-I. Chang, Hyperspectral Data Exploitation: Theory and Applications, John Wiley & Sons, New York, 2007.
[5] D. Manolakis, C. Siracusa, G. Shaw, Hyperspectral subpixel target detection using the linear mixing model, IEEE Trans. Geosci. Remote Sensing 39 (7) (2001) 1392–1409.
[6] C.-I. Chang, D.C. Heinz, Constrained subpixel target detection for remotely sensed imagery, IEEE Trans. Geosci. Remote Sensing 38 (3) (2000) 1144–1159.
[7] S. Catterall, Anomaly detection based on the statistics of hyperspectral imagery, in: Proceedings of the SPIE Conference on Imagery Spectroscopy X, 5546, 2004, pp. 171–178.
[8] B. Du, L. Zhang, Random selection based anomaly detector for hyperspectral imagery, IEEE Trans. Geosci. Remote Sensing 49 (5) (2010) 1578–1589.
[9] L. Zhang, B. Du, Y. Zhong, Hybrid detectors based on selective endmembers, IEEE Trans. Geosci. Remote Sensing 48 (2010) 2633–2646.
[10] M.J. Carlotto, A Cluster-based, Approach for detecting man-made objects and changes in imagery, IEEE Trans. Geosci. Remote Sensing 43 (2) (2005) 374–387.
[11] X. Tian, D. Tao, Y. Rui, Sparse transfer learning for interactive video search reranking, ACM Trans. Multimedia Comput. Commun. Appl., (2011).
[12] S. Si, D. Tao, B. Geng, Bregman divergence-based regularization for transfer subspace learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 929–942.
[13] T.E. Smetek, K.W. Bauer, A comparison of multivariate outlier detection methods for finding hyperspectral anomalies, Mil. Oper. Res. 13 (4) (2008) 19–44.
[14] I.S. Reed, X. Yu, Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution, IEEE Trans. Acoust. Speech Signal Process. 38 (10) (1990) 1760–1770.
[15] N. Billor, A.S. Hadi, P.F. Velleman, BACON: Blocked adaptive computationally efficient outlier nominators, Comput. Stat. Data Anal., 34, 279–298.
[16] M.U. Munir, M.Y. Javed, S.A Khan, A hierarchical *k*-means clustering based fingerprint quality classification, Neurocomputing 85 (15) (2012) 62–67.
[17] A. Singh, R. Nowak, X. Zhu, Unlabeled data: Now it helps, now it doesn't, Adv. Neural Inf. Process. Syst. 21 (2008) 1513–1520.
[18] B. Kulis, S. Basu, I. Dhillon, R. Mooney, Semi-supervised graph clustering: a kernel approach, Mach. Learn. 74 (2009) 1–22.
[19] M.H. Hansen, An evaluation of model-dependent and probability-sampling inferences in sample surveys, J. Am. Stat. Assoc. 23 (1983) 776–793.
[20] P.F. Felzenszwalb, D.P. Huttenlocher, Efficient graph-based image segmentation, Int. J. Comput. Vision 59 (2004) 167–181.
[21] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (22) (2000) 2323–2326.
[22] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (22) (2000) 2319–2323.
[23] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[24] Z. Zhang, H. Zha, Principal manifolds and nonlinear dimension reduction via local tangent space alignment, SIAM J. Sci. Comput. 26 (1) (2004) 313–338.
[25] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowledge Data Eng. 22 (10) (2010) 1345–1359.
[26] R. Raina, A. Battle, H. Lee, B. Packer, A.Y. Ng, Self-taught learning: transfer learning from unlabeled data, Int. Conf. Mach. Learn. (2007) 759–766.
[27] T. Evgeniou, M. Pontil, Regularized multi-task learning, in: Proceedings of the International Conference on Knowledge Discovery and Data Mining Seattle, Washington, 2004, pp. 109–117.
[28] B.C. Moore, Principal component analysis in linear systems: controllability, observability, and model reduction, IEEE Trans. Autom. Control 26 (1) (1981) 17–32.
[29] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization, Proc. Natl. Acad. Sci. USA 100 (5) (2003) 2197–2202.
[30] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2) (2004) 407–499.
[31] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face Recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell 31 (2) (2008) 210–227.
[32] T. Zhou, D. Tao, X. Wu, Manifold elastic net: a unified framework for sparse dimension reduction, Data Mining Knowledge Disc. 22 (3) (2010) 340–371.
[33] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, S. Hager, Development of a web-based application to evaluate target finding algorithms, IEEE Int. Geosci. Remote Sensing Sympos. (2008) 915–918.
[34] Y. Pang, J. Deng, Y. Yuan, Incremental threshold learning for classifier selection, Neurocomputing 89 (15) (2012) 89–95.
[35] Y. Pang, Y. Yuan, X. Li, Effective feature extraction in high-dimensional space, EEE Trans. Syst. Man Cybern. Part B: Cybern. 38 (2008) 1652–1656.
[36] Y. Pang, Y. Yuan, K. Wang, Learning optimal spatial filters by discriminant analysis for brain–computer-interface, Neurocomputing 77 (1) (2011) 20–27.
[37] Q. Du, H. Ren, Real-time constrained linear discriminant analysis to target detection and classification in hyperspectral imagery, Pattern Recognit. 36 (2003) 1–12.
[38] Y. Pang, Y. Yuan, Outlier-resisting graph embedding, Neurocomputing 73 (2010) 968–974.

Bo Du received the B.S. degree in engineering from Wuhan University, Wuhan, China, in 2005, the Ph.D. degree in Photogrammetry and Remote Sensing from State Key Lab of Information Engineering in Surveying, Mapping and Remote sensing, Wuhan University, Wuhan, China, in 2010.

He is currently a lecture with the School of Computer, Wuhan University, Wuhan, China. His major research interests include pattern recognition, hyperspectral image processing, and signal processing.

Liangpei Zhang 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 of Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in Photogrammetry and Remote Sensing from Wuhan University, China, in 1998.

He is currently with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, as the head of the Remote Sensing Division. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education, China. He has more than 200 research papers and 5 patents. His research interests include hyperspectral remote sensing, high resolution remote sensing, image processing and artificial intelligence.

Dacheng Tao 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 coauthored more than 100 scientific articles at top venues, including IEEE T-PAMI, T-KDE, T-IP, NIPS, ICML, UAI, AISTATS, ICDM, IJCAI, AAAI, CVPR, ECCV, ACM T-KDD, Multimedia, and KDD.

Prof. Tao was the recipient of the best theory/algorithm paper runner up award in IEEE ICDM'07.

Dengyi Zhang received the B.Eng. degree from the Wuhan Technical University of Surveying and mapping, Wuhan, China, in 1986, the Master of Computer degree from the University of Wuhan, Wuhan, China, in 1988. He is a Professor of computer science with the School of Computer, Wuhan University, Wuhan, China. His research interests include data mining, computer vision, machine learning, and their applications, such as forest fire surveillance.

Please cite this article as: B. Du, et al., Unsupervised transfer learning for target detection from hyperspectral images, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2012.08.056