Neural Process Lett (2016) 44:779–793
DOI 10.1007/s11063-015-9494-6

Discriminative Subspace Alignment for Unsupervised Visual Domain Adaptation

Hao Sun[1] · **Shuai Liu**[1] · **Shilin Zhou**[1]

Published online: 4 January 2016

Abstract We address the problem of unsupervised visual domain adaptation for transferring category models from one visual domain or image data set to another. We present a new unsupervised domain adaptation algorithm based on subspace alignment. The core idea of our approach is to reduce the discrepancy between the source domain and the target domain in a latent discriminative subspace. Specifically, we first generate pseudo-labels for the target data by applying spectral clustering to a cross-domain similarity matrix, which is built from sparse coefficients found in a low-dimensional latent space. This coarse alignment between the two domains exploits the assumption that the collection of data of different classes from both domains can be viewed as samples from a union of low-dimensional subspaces. Then, we create discriminative subspaces for both domains using partial least squares correlation. Finally, a mapping which aligns the discriminative source subspace into the target one is learned by minimizing a Bregman matrix divergence function. Experimental results on benchmark cross-domain visual object recognition data sets and cross-view scene classification data sets demonstrate that the proposed method outperforms the baselines and several state-of-the-art competing methods.

Keywords Unsupervised domain adaptation · Sparse subspace clustering · Partial least square correlation · Subspace alignment

1 Introduction

Traditional learning based image classification algorithms rely heavily on the assumption that data used for training and testing are drawn from the same distribution. However, many real-world applications challenge this assumption. For a typical object recognition task, the labeled training data are often obtained from well-established object database such

✉ Hao Sun
clhaosun@gmail.com

[1] College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410072, China

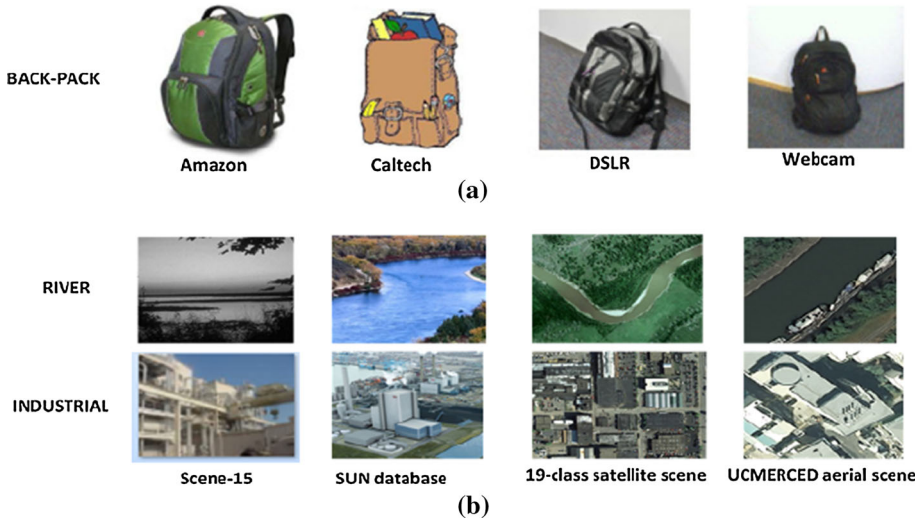

Fig. 1 Dataset bias for visual recognition. **a** Example images from BACK-PACK category in Caltech-256, Amazon, Webcam and DSLR [3]. **b** Example images from RIVER and INDUSTRIAL categories in two ground view scene datasets Scene-15 [4] and SUN database [2] and two overhead view scene datasets 19-class satellite scene dataset [5] and UCMERCED aerial scene dataset [6]

as Caltech-256 [1] while the testing data may be street daily shots acquired by mobile cameras. The same object model in both domains is subject to arbitrary shifts due to a combination of factors including different views, illuminations, object location and pose, resolutions and background clutter. For a more challenging cross view scene classification task, we want to transfer the semantic knowledge of scene models learned from rich annotated ground view images such as SUN database [2] to overhead view aerial or satellite scene images where the scarce of semantic annotations impedes the understanding of remote sensing images (see Fig. 1 for the dataset bias for the same set of object and scene categories).

Recent studies have demonstrated a significant degradation in the performance of state-of-the-art image classifiers under mismatched training and testing conditions [7]. Visual domain adaptation aims to address the problem of transferring object models or scene models from one visual domain to another. Depending on the availability of labeled training examples from the target domain, two scenarios are often differentiated: (i) the unsupervised setting where the training data consists of labeled source data and unlabeled target data and (ii) the semi-supervised setting where a large number of labels are available for the source domain and only a few labels are provided for the target domain. In this paper, we focus on the unsupervised case where the common practice of discriminative training is not applicable. Without target labels, it is not even clear how to define the right discriminative loss on the target domain.

In this letter, we introduce a new unsupervised visual domain adaptation algorithm based on subspace alignment. The contributions of our approach are twofold: (i) we propose to roughly align the two domains using sparse subspace clustering. Cross-domain sparse subspace clustering provides a natural way of passing down the label information from the source domain to the target domain and is robust to noise and outliers. (ii) we present a discriminative subspace alignment algorithm which maximizes the correlation between data and their

labels in the projected subspace and minimizes the data divergence by transforming source data to the target aligned source subspace.

The rest of this letter is organized as follows. Section 2 reviews related work and Sect. 3 introduces the motivation of our method. In Sect. 4, a detailed description of the discriminative sparse subspace clustering and alignment is presented. Section 5 reports the experimental results. Section 6 presents the discussion. Finally, Sect. 7 concludes the paper.

2 Related Work

Techniques for building classifiers that are robust to mismatched distributions have been investigated under the names of domain adaptation, covariate shift, or transfer learning. Recently, considerable effort has been devoted to domain adaptation in computer vision and machine learning communities. Several reviews can be found in [8–11]. Existing visual domain adaptation methods either try to find a common feature space where the data divergence between the source domain and the target domain can be significantly reduced or explicitly learn a new classifier model which minimizes the generalization error in the target domain.

Techniques that modify the representation of the data attempt to adjust the distributions of either the source or the target data, or both, to ultimately obtain a well-aligned feature space. In particular, subspace based visual domain adaptation methods have demonstrated good performance. Si et al. [12] introduced the Bregman divergence based regularization to several popular subspace learning algorithms for cross-domain face recognition and text categorization. Tuia et al. [13] proposed manifold alignment of different modalities of remote sensing images. Pan et al. [14] introduced transfer component analysis, which tries to learn some transfer components across domains in a reproducing kernel Hilbert space using maximum mean discrepancy. In [15], Chang transforms the source data into an intermediate representation such that each transformed source sample can be linearly reconstructed by the target samples. In [16], Shao et al. present a low-rank transfer subspace learning technique which exploits the locality aware reconstruction in a similar way to manifold learning. In [17], Gopalan et al. generate intermediate representations in the form of subspaces along the geodesic path connecting the source subspace and the target subspace on the Grassmann manifold. In [3], Gong et al. propose a geodesic flow kernel which models incremental changes between the source and target domains. In both [3] and [17], a set of intermediate subspaces are used to model the domain shift. Baktashmotlagh et al. [18] propose to learn a projection of the data to a low-dimensional latent space where the distance between the empirical distributions of the source and target samples is minimized. Fernando et al. [19] propose to align PCA based source subspace and PCA target subspace directly. Their method seeks a domain invariant feature space by learning a mapping function which aligns the source subspace with the target one. The solution of the corresponding optimization problem can be obtained in a simple closed form, leading to an extremely fast algorithm.

Previous research [3,17–19] suggests that partial least squares (PLS) is preferred over other supervised dimensionality reduction techniques for subspace based domain adaptation when label information is available. PLS locates and emphasizes group structure in the data and is closely related with canonical correlation analysis (CCA) and linear discriminant analysis (LDA). The PLS family consists of PLS correlation (PLSC) (also sometimes called PLS-SVD), PLS regression (PLSR) and PLS path modeling (PLS-PM). In this letter, we use PLSC to model the correlation between data samples and their labels.

3 Motivation

As suggested by [20], a reduction of the data distribution divergence between the source domain and the target domain is required to adapt well. From a mutual information point of view, let $H(\chi_S)$ denotes the entropy of the source data and $H(\chi_S, \chi_T)$ denotes the cross entropy between the source data and the target data, the mutual information between the two domains can then be given as follows:

$$\begin{aligned} MI(\chi_S; \chi_T) &= H(\chi_S) + H(\chi_T) - H(\chi_S, \chi_T) \\ &= H(\chi_T) - D_{KL}(\chi_S || \chi_T) \end{aligned} \tag{1}$$

According to Eq. (1), if we want to maximize the mutual information between the source distribution and the target distribution, we need to simultaneously increase the target entropy and reduce the data divergence between the two domains. If we project data from all domains to a target subspace, it will increase the term $H(\chi_T)$ and hence the mutual information. It will further improve the classification performance if a discriminative target subspace is used. For cross-domain data discrepancy reduction in low-dimensional subspaces, subspace alignment [19] provides a simple but effective framework for unsupervised scenario. However, it projects all data into a PCA based target subspace, which is not optimal for classification tasks. We aim to create a discriminative target subspace when no labels are available from the target domain and project all data from both domains into the generated subspace to minimize the data divergence.

Despite the shift that has occurred, samples belonging to the same category from both domains can be well represented by a low-dimensional subspace of the high-dimensional ambient space. The collection of data from multiple classes lies in a union of low-dimensional subspace. Cross-domain subspace clustering provides a natural way of passing down the labels from the source data to the target data, which can then be exploited for discriminative target subspace creation. Sparse representation and low-rank approximation based subspace clustering methods [21,22] have gained attention in recent years as they can handle noise and outliers in data, and they do not need to know the dimensions and the number of subspaces a priori.

4 Discriminative Subspace Alignment

4.1 Cross-Domain Subspace Clustering

We adopt the latent space sparse subspace clustering algorithm [21] for cross-domain subspace clustering. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in R^{D \times N}$ be a collection of N samples drawn from a union of n linear subspaces $S_1 \cup S_2 \cup \cdots \cup S_n$ of dimensions $\{d_\ell\}_{\ell=1}^n$ in R^D. Let $\mathbf{X}_\ell \in R^{D \times N_\ell}$ be a submatrix of $\mathbf{X}$ of rank d_ℓ with $N_\ell > d_\ell$ points that lie in S_ℓ. It is easy to see that each point in $\mathbf{X}$ can be efficiently represented by a linear combination of at most d_ℓ other points in $\mathbf{X}$. That is, one can represent $\mathbf{x}_i$ as

$$\mathbf{x}_i = \mathbf{X}\mathbf{c}_i, c_{ii} = 0, \|\mathbf{c}_i\|_0 \leq d_\ell \tag{2}$$

where $\mathbf{c}_i = [c_{i1}, c_{i2}, \ldots, c_{iN}]^T \in R^N$ are the coefficients. For sparse subspace clustering, the following minimization problem is solved to obtain the coefficients:

$$\min \|\mathbf{c}\|_1 \, s.t. \mathbf{x}_i = \mathbf{X}\mathbf{c}_i, c_{ii} = 0 \tag{3}$$

Let $\mathbf{P} \in \mathbf{R}^{d \times D}$ denotes a linear transformation that maps signals from the original space $\mathbf{R}^D$ to a latent output space of dimension d, latent space sparse subspace clustering learns the mapping and finds the sparse codes simultaneously by minimizing the following cost function:

$$[\mathbf{P}^*, \mathbf{C}^*] = \min_{\mathbf{P},\mathbf{C}} J(\mathbf{P}, \mathbf{C}, \mathbf{X}),\ s.t.\ \ \mathbf{P}\mathbf{P}^T = \mathbf{I},\ \ diag(\mathbf{C}) = 0 \tag{4}$$

$$\mathrm{J}(\mathbf{P}, \mathbf{C}, \mathbf{X}) = ||\mathbf{C}||_1 + \lambda_1 ||\mathbf{P}\mathbf{X} - \mathbf{P}\mathbf{X}\mathbf{C}||_F^2 + \lambda_2 ||\mathbf{X} - \mathbf{P}^{\mathbf{T}}\mathbf{P}\mathbf{Y}||_F^2 \tag{5}$$

where $\mathbf{X} = [\mathbf{X}_S; \mathbf{X}_T] \in \mathbf{R}^{D \times N}$ is the multi-class data from both domains, $\mathbf{C} \in \mathbf{R}^{N \times N}$ is the sparse coefficient matrix, λ_1 and λ_2 are non-negative constants that control sparsity and regularization. The first two terms promote sparsity of data in the latent subspace. The second term ensures that the projection preserves the main statistics of the data.

The problem can be efficiently solved by using the classical alternating direction method of multiplier. Once $\mathbf{C}$ is found, spectral clustering is applied on the affinity matrix $\mathbf{W} = |\mathbf{B}| + |\mathbf{C}|^{\mathbf{T}}$ to obtain the segmentation of the data. Each sample from the target domain can then be assigned a pseudo-label by majority voting of source sample labels in each cluster.

4.2 Discriminative Subspace Creation

Let $\mathbf{X}_T$ denotes a N_T by D data matrix from target domain where the rows are observations and the columns are variables and $\mathbf{Y}_T$ denotes the corresponding N_T by C pseudo-label matrix coded in the following way:

$$\mathbf{Y}_T = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \ldots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \ldots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_c} & \mathbf{0}_{n_c} & \ldots & \mathbf{1}_{n_c} \end{pmatrix}_{N_T \times C} \tag{6}$$

where $\mathbf{1}_{k \times 1}$ is a $k \times 1$ vector of all ones and likewise $\mathbf{0}_{k \times 1}$ is a $k \times 1$ vector of all zeros, C is the number of categories common to both domains. Each column of $\mathbf{X}_T$ is z-normalized (i.e. of zero mean and unit standard deviation). The correlation matrix is then computed as: $\mathbf{R}_T = \mathbf{Y}_T^{\mathbf{T}}\mathbf{X}_T$.The SVD of $\mathbf{R}_T$ decomposes it into three matrices: $\mathbf{R}_T = \mathbf{U}_T \Delta \mathbf{V}_T^{\mathbf{T}}$.The subspace basis for $\mathbf{X}_T$ is obtained as: $\mathbf{L}_{\mathbf{X}_T} = \mathbf{X}_T \mathbf{V}_T$. Similarly, we use PLSC to create a discriminative source subspace $\mathbf{L}_{\mathbf{X}_S}$.

4.3 Discriminative Subspace Alignment

A mapping function $\mathbf{M} \in \mathbf{R}^{d \times d}$ that transforms the source subspace $\mathbf{L}_{\mathbf{X}_S}$ into the target subspace $\mathbf{L}_{\mathbf{X}_T}$ is learned by minimizing the following Bregman matrix divergence:

$$F(\mathbf{M}) = ||\mathbf{L}_{\mathbf{X}_S}\mathbf{M} - \mathbf{L}_{\mathbf{X}_T}||_F^2 \tag{7}$$

where $||\cdot||_F^2$ is the Frobenius norm. Because the Frobenius norm is invariant to orthonormal operations, the objective function can be written as:

$$\begin{aligned} \mathbf{M}^* &= \arg\min_{\mathbf{M}} ||\mathbf{L}_{\mathbf{X}_S}\mathbf{M} - \mathbf{L}_{\mathbf{X}_T}||_F^2 \\ &= \arg\min_{\mathbf{M}} ||\mathbf{L}_{\mathbf{X}_S}^{-1}\mathbf{L}_{\mathbf{X}_S}\mathbf{M} - \mathbf{L}_{\mathbf{X}_S}^{-1}\mathbf{L}_{\mathbf{X}_T}||_F^2 \\ &= \arg\min_{\mathbf{M}} ||\mathbf{M} - \mathbf{L}_{\mathbf{X}_S}^{-1}\mathbf{L}_{\mathbf{X}_T}||_F^2 \end{aligned} \tag{8}$$

The optimal $\mathbf{M}^*$ is obtained as $\mathbf{M}^* = \mathbf{L}_{\mathbf{X}_S}^{-1}\mathbf{L}_{\mathbf{X}_T} = \mathbf{L}_{\mathbf{X}_S}^{\mathbf{T}}\mathbf{L}_{\mathbf{X}_T}$ and it transforms the source subspace coordinate system into the target subspace coordinate system by aligning the source basis vectors with the target ones. When the source and target domains are the same, and then $\mathbf{M}^*$ is the identity matrix. The pseudo-code of the proposed unsupervised discriminative subspace alignment algorithm is presented in Algorithm 1.

Algorithm 1 Discriminative Subspace Alignment

Input: Source data $\mathbf{X}_S$, source labels $\mathbf{Y}_S$, target data $\mathbf{X}_T$, sparsity level λ_1, regularization parameter λ_2.

Step 1: $\mathbf{Y}_T^* \leftarrow cross\ \ domain\ \ clustering([\mathbf{X}_S;\mathbf{X}_T],\mathbf{Y}_S,\lambda_1,\lambda_2)$;

Step 2: $\mathbf{L}_{\mathbf{X}_S} \leftarrow PLSC(\mathbf{X}_S,\mathbf{Y}_S)$ and $\mathbf{L}_{\mathbf{X}_T} \leftarrow PLSC(\mathbf{X}_T,\mathbf{Y}_T^*)$;

Step 3: Transform source data by $\tilde{\mathbf{X}}_S = \mathbf{X}_S\mathbf{L}_{\mathbf{X}_S}\mathbf{L}_{\mathbf{X}_S}{}^{\mathbf{T}}\mathbf{L}_{\mathbf{X}_T}$;

Step 4: Transform target data by $\tilde{\mathbf{X}}_T = \mathbf{X}_T\mathbf{L}_{\mathbf{X}_T}$;

Step 5: Predict $\mathbf{Y}_T$ by $NN-classifier(\tilde{\mathbf{X}}_S,\mathbf{Y}_S,\tilde{\mathbf{X}}_T)$.

Output: Predicted target labels $\mathbf{Y}_T$.

5 Experimental Results

We evaluate our methods in the context of cross-domain visual object recognition and cross-view scene classification. All baseline methods and other competing unsupervised domain adaptation methods project the original high-dimensional feature to a new feature space where a nearest neighbor (NN) classifier is trained on the labeled source data and tested on the unlabeled target data. NN is chosen as the base classifier as it does not require tuning cross-validation parameters. Under our experimental setup, it is difficult to tune the optimal parameters using cross-validation since labeled and unlabeled data are sampled from different distributions.

5.1 Cross-Domain Visual Object Recognition

In order to evaluate the effectiveness of our method for transferring object models from one visual domain to another, we use the benchmark Office+Caltech-10 [3] dataset for cross-domain visual object recognition. The dataset includes four visual domains: Amazon (A: images downloaded from online merchants), Webcam (W: low-resolution images by a web camera), DSLR (D: high-resolution images by a digital SLR camera) and Caltech-256 dataset (C). It consists of 2533 images of 10 classes common to all four domains: BACK-PACK, TOURING-BIKE, CALCULATOR, HEAD-PHONES, COMPUTER-KEYBOARD, LAPTOP-101, COMPUTER-MONITOR, COMPUTER-MOUSE, COFFEE-MUG, and VIDEO-PROJECTOR. There are 8 to 151 samples per category per domain. Figure 2 shows the differences among these domains with example images from the category of MONITOR.

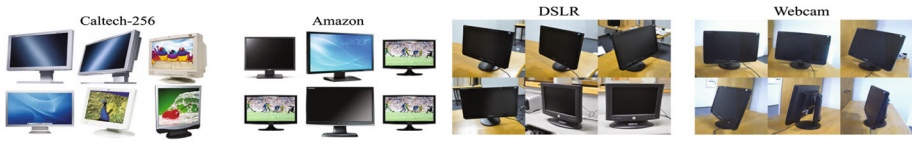

Fig. 2 Example images from the MONITOR category in Caltech-256, Amazon, DSLR, and Webcam

By randomly selecting two different domains as the source domain and target domain respectively, there exist 12 different cross domain object recognition problems. In our experiments, we use the same image representation (SURF features encoded with an 800 words dictionary) and protocol for generating the source and target samples with the literature [3,17,19]. We set $\lambda_1 = \lambda_2 = 0.5$ for latent space sparse subspace clustering (LS3C) [22]. We compare our methods to three baselines and several state-of-the-art methods including SGF (sampling geodesic flow) [17], GFK (geodesic flow kernel) [3], transfer component analysis (TCA) [14], subspace alignment (SA) [19], and low-rank transfer subspace learning (LTSL) [16]. The accuracy of pseudo-label generated by LS3C is also reported.

- Baseline-NA (no adaptation): where we use the original feature representation after z-normalization.
- Baseline-S: where we project both source and target data into PCA based source subspace.
- Baseline-T: where we project both source and target data into PCA based target subspace.

The subspace dimensionality of Baseline-S and Baseline-T is determined by MLE based domain intrinsic dimensionality estimation [23,24]. For each method, we report the best accuracy for each case from the corresponding paper in order to avoid any implementation differences (SGF is based on the implementation of LTSL paper). The results are presented in Table 1. The best performing methods in each column are in bold font and the second best group is in italics and underlined. Figure 3 shows the cross-domain similarity matrix built from LS3C for the case D $\rightarrow$ C. It can be seen that despite the shift that has occurred, the cross-domain data from the same category have a strong similarity.

The best performing methods (differences up to one stand ard error) in each column are in bold font and the second best group is in italics and underlined.

5.2 Cross-View Scene Classification

We have collected a cross-view scene dataset from two ground level scene datasets: SUN database [2] (Source domain 1, S1) and Scene-15 [4] (Source domain 2, S2), and three overhead remote sensing scene datasets: Banja Luka dataset [25] (Target domain 1, T1), UCMERCED dataset [6] (Target domain 2, T2), 19-class satellite scene dataset [5] (Target domain 3, T3). The dataset consists of 2768 images of four common categories (field/agriculture, forest/trees, river/water and industrial). Figure 4 shows an example of the dataset (one image per category per dataset). Table 2 gives the statistics of image numbers in the dataset.

For each image in the dataset, the histogram of oriented edges (HOG) feature is extracted (stacking 2×2 neighboring descriptor of 8×8 pixels cell). HOG descriptors have been quantized into 300 visual words by k-means. With local-constraint linear coding (LLC), three level spatial histograms are computed on grids of 1×1, 2×2 and 4×4. Each image is finally represented by a 6300 dimensional z-normalized vector. The subspace dimensionality is determined by MLE based domain intrinsic dimensionality estimation using the target data for SGF, GFK, TCA, SA (PCA, PCA), and Baseline-T and using the source data for Baseline-S. The sampling rate of SGF is set to [0.2, 0.4, 0.6, 0.8]. For SGF, GFK and SA, we use the implementations provided by the authors.

For each source-target DA problem, 20 images from each category in the source domain are randomly selected as the training set and all the images in the target domain as the testing set, the classification accuracies of all the above methods over 20 random trials using a NN classifier is summarized in Table 3.

The best performing methods (differences up to one standard error) in each column are in bold font and the second best group is in italics and underlined.

Table 1 Cross-domain visual object recognition accuracy (%) with unsupervised adaptation using NN classifier

Method	A → C	A → D	A → W	C → A	C → D	C → W	D → A	D → C	D → W	W → A	W → C	W → D	AVG
Baseline-NA	22.8 ± 0.6	22.4 ± 0.8	23.3 ± 1.3	21.5 ± 0.7	21.7 ± 1.3	20.0 ± 1.4	26.9 ± 0.6	24.8 ± 0.5	53.0 ± 0.8	20.8 ± 0.9	16.4 ± 0.6	40.5 ± 1.2	26.2
Baseline-S	30.9 ± 0.4	34.6 ± 0.6	35.1 ± 0.8	38.0 ± 1.0	37.4 ± 0.9	33.5 ± 1.0	29.8 ± 0.6	29.6 ± 0.4	74.0 ± 0.8	35.5 ± 1.2	31.3 ± 0.9	71.8 ± 0.8	40.1
Baseline-T	33.3 ± 0.7	34.7 ± 0.6	36.8 ± 0.6	40.5 ± 0.9	36.4 ± 1.1	34.4 ± 0.9	33.0 ± 0.3	31.2 ± 0.6	78.4 ± 0.5	36.0 ± 1.4	31.9 ± 0.7	72.9 ± 1.0	41.8
SGF [17]	35.6 ± 0.4	34.9 ± 0.9	34.4 ± 0.9	36.9 ± 0.5	35.2 ± 1.0	33.9 ± 1.2	32.6 ± 0.5	30.0 ± 0.2	74.9 ± 0.6	31.3 ± 0.7	27.3 ± 0.5	70.7 ± 0.9	39.8
GFK (PCA,PCA) [3]	35.6 ± 0.4	35.2 ± 0.9	34.4 ± 0.9	36.9 ± 0.4	35.2 ± 1.0	33.7 ± 1.1	32.5 ± 0.5	29.8 ± 0.3	74.9 ± 0.6	31.1 ± 0.8	27.2 ± 0.5	70.6 ± 0.9	39.8
GFK (PLSR,PCA) [3]	37.9 ± 0.4	35.1 ± 0.8	35.7 ± 0.9	40.4 ± 0.7	41.1 ± 1.3	35.8 ± 1.0	36.2 ± 0.4	32.7 ± 0.4	79.1 ± 0.7	35.5 ± 0.7	29.3 ± 0.4	71.2 ± 0.9	42.5
TCA [14]	28.8 ± 0.6	30.4 ± 1.0	30.3 ± 0.6	34.7 ± 0.8	34.7 ± 0.8	28.8 ± 0.9	27.5 ± 0.5	28.8 ± 0.3	70.9 ± 0.9	34.1 ± 0.6	30.5 ± 0.4	64.4 ± 1.1	37.0
SA (PCA,PCA) [19]	35.3 ± 0.4	37.6 ± 0.8	38.6 ± 0.6	39.0 ± 0.5	39.6 ± 0.6	36.8 ± 1.0	38.0 ± 0.6	32.4 ± 0.5	83.6 ± 0.7	37.4 ± 0.6	32.3 ± 0.5	80.3 ± 1.0	44.2
SA (PLSR,PCA) [19]	37.1 ± 0.4	38.0 ± 0.7	41.4 ± 1.0	43.4 ± 0.6	41.2 ± 0.8	42.6 ± 0.9	36.8 ± 0.4	35.9 ± 0.5	76.4 ± 0.6	38.4 ± 0.6	33.2 ± 0.7	61.9 ± 0.8	43.9
LTSL-PCA [16]	35.8 ± 0.3	25.0 ± 0.8	24.4 ± 0.8	38.9 ± 0.5	24.5 ± 0.9	23.3 ± 0.8	34.0 ± 0.3	29.4 ± 0.2	70.0 ± 0.4	32.4 ± 0.4	29.9 ± 0.1	73.7 ± 0.6	36.8
LTSL-LDA [16]	38.6 ± 0.3	*38.3* ± 1.1	*38.8* ± 1.3	**50.4** ± 0.4	**53.7** ± 0.9	**47.0** ± 1.0	**40.2** ± 0.6	35.3 ± 0.3	72.8 ± 0.7	**44.1** ± 0.3	**37.4** ± 0.2	79.8 ± 0.4	*48.0*
LS3C [22]	**46.6** ± 0.2	27.9 ± 0.6	28.7 ± 0.4	40.0 ± 0.6	25.7 ± 0.8	26.9 ± 0.9	*40.1* ± 0.3	**43.3** ± 0.3	*84.7* ± 0.6	34.6 ± 0.4	*36.9* ± 0.5	*81.7* ± 0.4	43.1
Our method	*40.8* ± 0.4	**41.1** ± 0.5	**40.4** ± 0.4	*48.5* ± 0.3	*45.5* ± 0.6	*43.7* ± 1.0	39.6 ± 0.1	*37.3* ± 0.3	**91.1** ± 0.5	*36.5* ± 0.2	33.4 ± 0.3	**84.9** ± 0.6	**48.6**

A Amazon, *C* Caltech, *D* DSLR, *W* Webcam

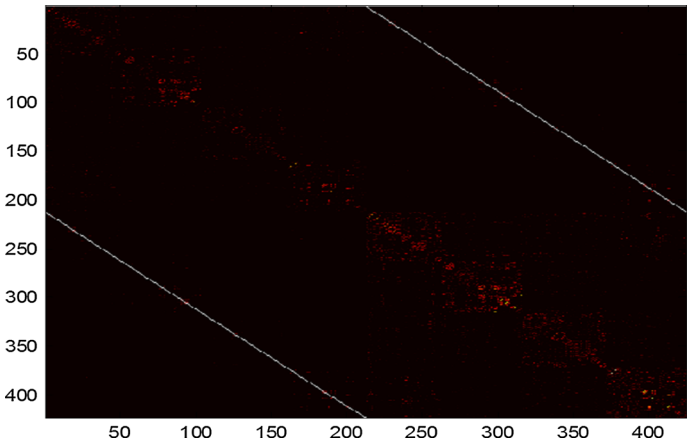

Fig. 3 Cross-domain similarity matrix built from LS3C for the case D → C

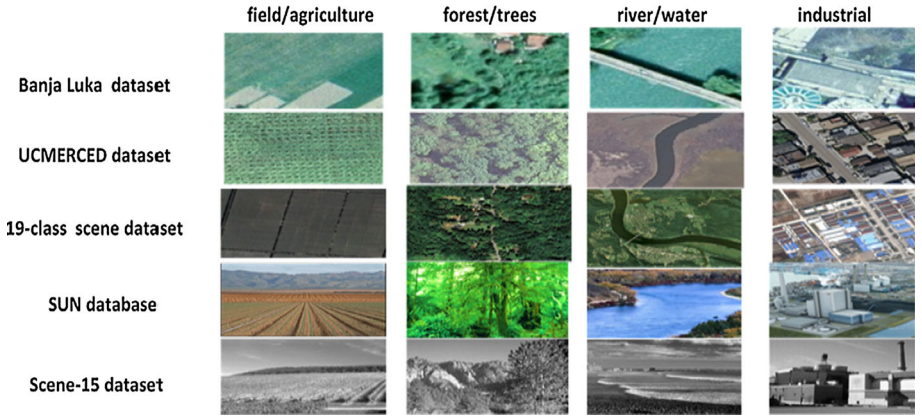

Fig. 4 Example images from the cross-view scene dataset

Table 2 Cross-view scene dataset statistics

Dataset	Field/agriculture	Forest/trees	River/water	Industrial
SUN (S1)	84	62	125	41
Scene-15 (S2)	410	328	360	311
Banja Luka (T1)	178	105	77	75
UCMERCED (T2)	100	100	100	100
19-class Scene (T3)	50	53	56	53

SUN database (S1) and Scene-15 (S2) are the source domains while Banja Luka (T1), UCMERCED (T2), and 19-class Scene (T3) datasets are the target domains

We have also evaluated the performance of our method in the context of transferring scene category models from aerial scenes to satellite scenes. We have collected 1377 images of nine common categories from UCMERCED aerial scene dataset [6] and 19-class satellite scene dataset [5]. Figure 5 shows the images from each category. We adopted the same parameter

Table 3 Cross-view scene classification accuracy (%) with unsupervised adaptation using NN classifier

Method	S1 → T1	S1 → T2	S1 → T3	S2 → T1	S2 → T2	S2 → T3	AVG
Baseline-NA	30.0 ± 1.4	29.7 ± 1.8	37.8 ± 1.6	31.6 ± 0.5	40.5 ± 1.1	38.5 ± 0.5	34.7
Baseline-S	30.9 ± 0.9	40.8 ± 1.8	47.6 ± 1.1	33.1 ± 0.8	51.4 ± 1.2	48.7 ± 0.8	42.1
Baseline-T	34.9 ± 0.7	49.1 ± 2.0	46.4 ± 1.2	32.0 ± 0.7	50.2 ± 0.9	53.4 ± 1.0	44.3
SGF [17]	27.8 ± 0.1	45.8 ± 0.3	43.2 ± 1.3	34.8 ± 0.2	47.8 ± 0.6	48.3 ± 0.2	41.3
GFK (PCA,PCA) [3]	34.3 ± 0.1	46.6 ± 2.9	45.1 ± 1.1	29.5 ± 0.5	48.3 ± 0.8	51.9 ± 0.6	40.5
TCA [14]	29.6 ± 0.6	44.8 ± 2.4	43.8 ± 0.2	28.9 ± 0.7	46.8 ± 1.4	49.8 ± 0.8	40.6
SA (PCA, PCA) [19]	*34.4* ± 0.6	45.4 ± 1.1	47.2 ± 1.0	31.9 ± 0.8	*50.6* ± 0.9	*53.8* ± 0.6	43.9
SA (PLSR, PCA) [19]	30.8 ± 0.7	*61.0* ± 1.2	*56.6* ± 0.5	*40.5* ± 0.7	50.0 ± 0.7	51.8 ± 0.9	*48.5*
LS3C [22]	32.1 ± 0.3	52.5 ± 1.0	46.9 ± 0.8	34.4 ± 0.6	49.9 ± 1.2	50.2 ± 0.6	44.3
Our method	**35.5** ± 0.4	**63.4** ± 0.9	**57.2** ± 1.4	**44.6** ± 0.6	**52.0** ± 0.6	**54.6** ± 0.7	**51.2**

S1 SUN database, *S2* Scene-15, *T1* Banja Luka, *T2* UCMERCED, *T3* 19-class Scene

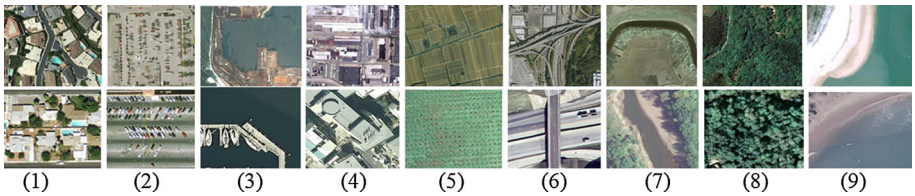

Fig. 5 Nine common categories of satellite scenes from 19-class satellite scene dataset (*top row*) and aerial scenes from UCMERCED dataset (*bottom row*). **1** Residential, **2** parking lot, **3** port/harbor, **4** industry/building, **5** farmland/ agriculture, **6** viaduct/ overpass, **7** river, **8** forest, **9** beach

settings with the cross-view scene classification experiments and the results are reported in Table 4. We have carried out the experiments on a machine with 2.80GHZ Intel CPU and 2.98GB RAM based on matlab implementation of our algorithm. The classification time of different methods is reported in Table 5.

The best performing methods (differences up to one standard error) in each column are in bold font and the second best group is in italics and underlined.

6 Discussions

The proposed method adopts a coarse-to-fine adaptation strategy based on subspace analysis. It is a general framework for unsupervised domain adaptation, and can be easily applied to other signal processing tasks. For cross-domain visual object recognition application, it can be seen from Table 1 that the proposed method outperforms the baselines and other competing methods on average. Our algorithm achieves the best performance for four problems (A $\rightarrow$ D, A $\rightarrow$ W, D $\rightarrow$ W and W $\rightarrow$ D) and the second best performance for six problems out of twelve cases. It should be noted that a coarse alignment using cross domain latent subspace clustering improves by 7 % in classification accuracy over no adaptation and outperforms SGF, GFK, TCA and LTSL-PCA. For cross-view scene classification application, it can be seen that the proposed method consistently outperforms the baselines and other competing methods in all six problems. SA achieves the second best performance. The results confirm the effectiveness of projecting source data into a discriminative target subspace. For aerial-to-satellite scene classification application, our method improves by 6 % in classification accuracy over the second best performing methods. The computational complexity of our algorithm mainly comes from two parts. The first part is the cross-domain latent subspace clustering, which has a complexity of $O((K''wn^2 + D''wn^2))$, where K and M are related to the iteration and feature dimensionality, n is number of samples. The second part is the SVD decomposition of PLSC, which has a complexity of $O(L^3)$, where L is number of categories.

Compared with manifold based methods such as SGF and GFK and kernel space based method such as TCA, the subspace alignment strategy is very simple in theory and can be solved in a closed form, leading to an extremely fast algorithm. The major improvements of our method over the original subspace alignment work are twofold: i) a cross-domain latent subspace clustering step is used to pass down the labels from source data to target data; ii) PLSC is adopted to model the correlation between data samples and their labels in both domains. It should be noted that our method relies on the assumption that the target samples can be represented by a sparse set of source samples in latent subspace. When the assumption is violated, it is not likely to perform well. Another weakness of our method is that the sparse

Table 4 Aerial-to-satellite scene classification accuracy (%) with unsupervised adaptation using NN classifier (UCMERCED dataset is the source domain and 19-class satellite scene dataset is the target domain)

Method	Residential	Industry	Farmland	River	Forest	Parking	Viaduct	Beach	Port	AVG
Baseline-NA	**75.6** ± 1.8	5.9 ± 0.9	9.9 ± 0.9	3.7 ± 0.8	**92.3** ± 0.9	34.7 ± 2.3	1.3 ± 0.4	10.9 ± 1.2	1.8 ± 0.5	26.0
Baseline-S	70.4 ± 2.8	6.6 ± 0.9	9.3 ± 0.8	2.9 ± 0.7	90.6 ± 1.3	40.2 ± 2.4	2.3 ± 0.7	12.1 ± 1.1	2.8 ± 0.5	26.4
Baseline-T	*73.2* ± 2.1	5.8 ± 0.8	9.0 ± 0.6	3.7 ± 1.1	89.7 ± 1.5	36.6 ± 2.3	0.9 ± 0.4	11.2 ± 1.2	2.4 ± 0.4	25.7
SGF [17]	41.6 ± 2.2	15.4 ± 1.1	3.7 ± 0.4	32.8 ± 2.2	45.8 ± 1.4	26.3 ± 1.3	13.2 ± 1.9	18.9 ± 1.2	9.6 ± 0.9	23.2
GFK (PCA,PCA) [3]	42.7 ± 3.3	35.7 ± 3.1	10.7 ± 0.8	46.8 ± 3.4	72.5 ± 2.7	20.8 ± 1.1	39.8 ± 3.8	19.4 ± 1.2	33.9 ± 1.7	36.3
TCA [14]	54.5 ± 2.0	45.1 ± 2.3	5.2 ± 0.7	43.7 ± 3.2	13.3 ± 0.4	9.3 ± 0.9	37.3 ± 2.5	16.4 ± 0.9	26.2 ± 2.9	27.1
SA (PCA,PCA) [19]	31.1 ± 2.2	42.5 ± 1.7	23.8 ± 1.0	47.5 ± 2.7	63.6 ± 3.0	21.8 ± 1.2	54.5 ± 2.1	*27.3* ± 1.5	*45.5* ± 2.5	40.2
SA (PLSR,PCA) [19]	16.9 ± 2.8	*57.2* ± 2.8	*32.9* ± 1.4	*47.9* ± 2.0	62.7 ± 2.9	**42.1** ± 1.7	*71.8* ± 1.7	20.5 ± 0.8	22.5 ± 1.9	42.1
LS3C [22]	27.8 ± 1.3	**62.3** ± 0.8	28.0 ± 0.6	44.6 ± 1.1	49.1 ± 0.7	*42.0* ± 1.9	**79.3** ± 0.5	18.0 ± 1.3	26.4 ± 0.9	*42.6*
Our method	19.8 ± 1.9	48.7 ± 1.3	**41.4** ± 1.4	**56.3** ± 1.1	*91.6* ± 0.9	35.5 ± 1.0	35.5 ± 2.6	**63.7** ± 1.7	**51.9** ± 1.7	**48.8**

Table 5 Aerial-to-satellite scene classification time of different methods (UCMERCED dataset is the source domain and 19-class satellite scene dataset is the target domain)

Method	Baseline-NA	Baseline-S	Baseline-T	SGF	GFK (PCA,PCA)	TCA	SA (PCA,PCA)	SA (PLSR,PCA)	LS3C	Our method
Time (s)	16.0	24.0	22.0	109.4	120.9	8.8	22.8	23.4	169.5	180.0

representation step for cross domain subspace clustering is computationally demanding. In the future, we will try to use some accelerated sparse approximation tools.

7 Conclusions

Inspired by the recent success of sparse subspace clustering and subspace based visual domain adaptation, we propose a novel discriminative sparse subspace clustering and alignment framework for unsupervised scenario. We aim to create a discriminative target subspace when no labels are available from the target domain and project all data from both domains into the generated subspace to minimize the data divergence. Our method consists of three major components: cross domain latent space sparse subspace clustering, discriminative subspace creation and subspace alignment. Experimental results on benchmark cross-domain visual object recognition datasets and cross-view scene datasets demonstrate the effectiveness of the proposed method.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China under Grant 61303186 and by the Ph.D. Programs Foundation of Ministry of Education of China under Grant 20124307120013.

References

1. Griffin G, Holub A, Perona P (2007) Caltech-256 object category dataset (technical report), Caltech
2. Xiao J, Ehinger K, Hays J, Torralba A, Oliva A (2014) SUN database: exploring a large collection of scene categories. Int J Comput Vis 108:1–8
3. Gong B, Grauman K, Sha F (2014) Learning kernels for unsupervised domain adaptation with applications to visual object recognition. Int J Comput Vis 109:3–27
4. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of IEEE conference on computer vision and pattern recognition, vol 2, pp 2169–2178
5. Dai D, Yang W (2011) Satellite image classification via two-layer sparse coding with biased image representation. IEEE Geosci Remote Sens Lett 8(1):173–176
6. Yang Y, Newsam S (2010) Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the ACM international conference on Advances in geographic information systems, ACM, New York, pp 270–279
7. Torralba A, Efros A (2011) Unbiased look at dataset bias. In: Proceedings of IEEE conference on computer vision and pattern recognition, vol 2, pp 1521–1528
8. Margolis A (2011) A literature review of domain adaptation with unlabeled data (technical report), University of Washington, Washington
9. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
10. Shao L, Zhu F, Li X (2014) Transfer learning for visual categorization: a survey. IEEE TNNLS 26:1019–1034
11. Patel VM, Gopalan R, Li R, Chellappa R (2015) Visual domain adaptation: an overview of recent advances. In: IEEE signal processing magazine
12. Si S, Tao D, Geng B (2010) Bregman divergence-based regularization for transfer subspace learning. IEEE Trans Knowl Data Eng 22(7):929–942
13. Tuia D, Volpi M, Trolliet M, Camps-Valls G (2014) Semisupervised manifold alignment of multimodal remote sensing images. IEEE Trans Geosci Remote Sens 52(12):7708–7720
14. Pan SJ, Tsang IW, Kwok JT, Yang Q (2011) Domain adaptation via transfer component analysis. IEEE Trans Neural Netw 22(2):199–210
15. Chang SF (2012) Robust visual domain adaptation with low-rank reconstruction. In: Proceedings of IEEE conference on computer vision and pattern recognition, vol 2, pp 1–8
16. Shao M, Kit D, Fu Y (2014) Generalized transfer subspace learning through low-rank constraint. Int J Comput Vis 109:74–93

17. Gopalan R, Li R, Chellappa R (2011) Domain adaptation for object recognition: an unsupervised approach. In: Proceedings of IEEE international conference on computer vision, pp 999–1006
18. Baktashmotlagh M, Harandi MT, Lovell BC, Salzmann M (2013) Unsupervised domain adaptation by domain invariant projection. In: Proceedings of IEEE international conference on computer vision, pp 769–776
19. Fernando B, Habrard A, Sebban M, Tuytelaars T (2013) Unsupervised visual domain adaptation using subspace alignment. In: Proceedings of IEEE international conference on computer vision, pp 2960–2967
20. Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F, Wortman J (2010) A theory of learning from different domains. Mach Learn 79:151–175
21. Elhamifar E, Vidal R (2013) Sparse subspace clustering: algorithm, theory, and applications. IEEE Trans Pattern Anal Mach Intell 35(11):2765–2781
22. Patel VM, Nguyen HV, Vidal R (2013) Latent space sparse subspace clustering. In: Proceedings of IEEE international conference on computer vision, pp 225–232
23. Levina E, Bickel PJ (2004) Maximum likelihood estimation of intrinsic dimension. In: Proceedings of the NIPS, pp 1–8
24. Van der Maaten L, Hinton G (2008) Visualizing data using t-sne. J Mach Learn Res 9(11):1–8
25. Risojevic V, Babic Z (2011) Aerial image classification using structural texture similarity. In: Proceedings of the IEEE international symposium on signal processing and information technology, pp 190–195