Structure Preserving Transfer Learning for Unsupervised Hyperspectral Image Classification

Jianzhe Lin, *Student Member, IEEE*, Chen He, *Member, IEEE*, Z. Jane Wang, *Fellow, IEEE*, and Shuying Li

***Abstract*— Recent advances on remote sensing techniques allow easier access to imaging spectrometer data. Manually labeling and processing of such collected hyperspectral images (HSIs) with a vast quantities of samples and a large number of bands is labor and time consuming. To relieve these manual processes, machine learning based HSI processing methods have attracted increasing research attention. A major assumption in many machine learning problems is that the training and testing data are in the same feature space and follow the same distribution. However, this assumption doesn't always hold true in many real world problems, especially in certain HSI processing problems with extremely insufficient or even without training samples. In this letter, we present a transfer learning framework to address this unsupervised challenge (i.e., without training samples in the target domain), by making the following three main contributions: 1) to the best of our knowledge, this is the first time for transfer learning framework to be used for the classification of totally unknown target HSI data with no training samples; 2) the characteristics of HSI are learned on dual spaces to exploit its structure knowledge to better label HSI samples; and 3) two specific new scenarios suitable for transfer learning are investigated. Experimental results on several real world HSIs support the superiority of the proposed work.**

***Index Terms*— Hyperspectral image (HSI), remote sensing, transfer learning.**

I. Introduction

A PREVALENT trend for hyperspectral image (HSI) classification is to exploit spatial and spectral information of HSI data to the maximum. However, a major assumption for the dominant supervised process is that the training and testing data are in the same feature space and have the same distribution [1], [2]. However, in many real world problems, especially for the processing of the newly collected HSI data whose training samples are not ample, the above assumption may not hold and existing supervised methods would fail to work. Traditional unsupervised methods which barely consider the data distribution of HSI may yield unsatisfying classification performances. In this situation, we may transfer the knowledge obtained from another relevant data set with sufficient training samples (as the source domain), which might be with a different feature space or learning task, to assist the unsupervised HSI processing in the target domain.

Transfer learning makes use of the prior information from the relevant domain to learn new tasks on the objective domain [3]. The former and latter domain are named the source domain and the target domain, respectively. Depending on whether labeled training samples in the target domain are available or not, transfer learning methods are further divided into supervised and unsupervised categories. Here we mainly concentrate on unsupervised transfer learning which is of great interest in real world problems. A key challenge in transfer learning is how to set up the relevance between the two domains. Generally, the data in the two domains share the same task while follow different distributions. To be more specific for our HSI problem, even they represent the same categories of interest, the data in the two domains can be with very different spectra, due to factors such as various acquisition conditions and the use of different sensors in HSI.

To adapt the data from the source domain to the target domain, great efforts have been made by researchers. A representative idea is to transfer the representations of both source and target data into a common space. The most prevalent works can be found in [4] and [5], while they have some common drawbacks as follows. First, they generally do not take into consideration the intrinsic structures of data, including global and local features of HSI data. Second, the original data are treated equally and represented with the same principle [6], and the noise cannot be effectively eliminated during the representation process [7].

To address the above limitations, we propose a novel dual space unsupervised structure preserving transfer learning (DSTL) framework for HSI classification. To address the first concern, the Markov random field (MRF) is applied to exploit the intrinsic data structure to get the optimal labels for HSI data in the target domain [8]. For the second concern, we formulate the objective function by simultaneously considering the noise as well as the transformation matrix, and the proposed model can effectively handle data discrepancy between the source domain data and the target domain data. A flowchart for the proposed framework can be found in Fig. 1.

We summarize major contributions of this letter as follows.

1) We propose a novel transfer learning framework. We apply the joint low-rank and sparse constraints for

Manuscript received May 14, 2017; revised June 26, 2017; accepted June 29, 2017. This work was supported by the National Priorities Research Program through the Qatar National Research Fund (a member of Qatar Foundation) under Grant 7-684-1-127. *(Corresponding author: Shuying Li.)*

J. Lin and Z. J. Wang are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: jianzhelin@ece.ubc.ca; zjanew@ece.ubc.ca).

C. He is with School of Information Science and Technology, Northwest University, Xi'an 710069, China, and also with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: chenh@ece.ubc.ca).

S. Li is with the Aerospace Science and Technology Corporation, Xi'an 710100, China (e-mail: angle_lisy@163.com).

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

Digital Object Identifier 10.1109/LGRS.2017.2723763

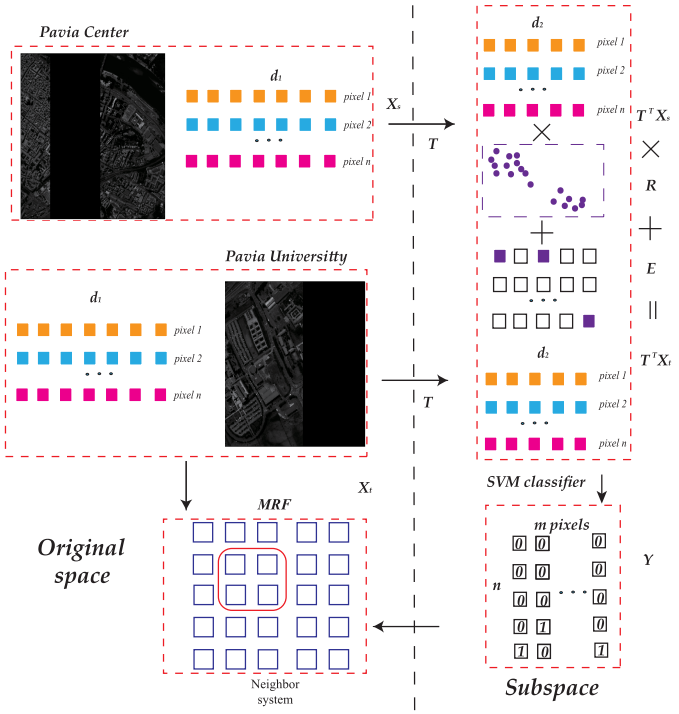

Fig. 1. Framework of the proposed method. The source data X_S are the HSI data from the PC domain whereas the target data X_t are also HSI from the PU domain. We first project the data on both domains into a new feature subspace and find the transformation matrix T. With T, we can obtain the initial classes of pixels. Further, the data on the target domain return to the original space to be optimized by MRF, which imposes the structure constraint.

reconstruction to preserve both local and global structures of the data on the new feature subspace that we specifically introduce [6], [9]. Data from both domains are well interlaced by these constraints. We also address the noise effect concern by using a sparse matrix to model the noise so that the noise information can be filtered.

2) We explore a new dual space structure preserving constraint. We get initial classification results on the subspace. On the original data space, we further restrict the results from the subspace by considering the structure of the target data through MRF to avoid the local minimum and obtain globally optimal labels.
3) We investigate two specific transfer subspace learning scenarios, to be more specific, the intrascenario transfer and the interscenario transfer. Experimental results show the effectiveness of the proposed framework for HSI classification on the target domain without labeled training samples.

The rest of this letter is organized as follows. Section II presents major components of the proposed framework in detail. Experiments and comparisons are reported in Section III to indicate the superiority of the proposed framework. We finally conclude this letter in Section IV.

II. Proposed Framework

In this section, we present the proposed method in detail, as illustrated in Fig. 1. The related definitions are given first.

A. Definitions

Let $X_s \in R^{d_1 \times n_s}$ be the data from the source domain and $X_t \in R^{d_1 \times n_t}$ be the data on the target domain, in which d_1 is the dimension of HSI data, n_s and n_d are the number of pixels in these two domains, respectively. Suppose the transformation matrix is $T \in R^{d_1 \times d_2}$, and the reconstruction matrix for the data on the source domain is $R \in n_s \times n_t$, the noise matrix is $E \in R^{d_2 \times n_t}$, in which d_2 is the dimension of data on the subspace. The label matrix is denoted by $Y_s = [y_1, y_2, y_3, \ldots, y_{n_s}] \in R^{m \times n_s}$ and $Y_t = [y_1, y_2, y_3, \ldots, y_{n_t}] \in R^{m \times n_t}$ where m is the number of classes. For a random sample x_i from class k, its corresponding label vector y_i has its kth element $y_i^k = 1$, and the rest are zeros.

B. Constraints on the Subspace

We can formulate both the HSI on the source domain and the target domain on a new common feature subspace as

$$T^T X_t = T^T X_s R \tag{1}$$

and the solution of T and R in (1) is further formulated as

$$\min_{T,R} ||T^T X_t - T^T X_s R||. \tag{2}$$

Since the data in both domains lie in the same subspace, we can find the knowledge that can be effectively transferred based on (2). However, for HSI, which can span multiple subspaces, the transfer process is not unique. Moreover, the structure information cannot be exploited by 2. We therefore would like to project the data in both domains to an optimal common subspace on which the divergence among data distributions of both domains is minimized. With the minimum divergence, the data from the target domain can be reconstructed by the neighboring data on the source domain. To be more specific, the data from both domains are almost with identical distributions and lie on the same mainfold subspace. On this space, each sample with the same task can be represented by its neighbors. To achieve this goal, we understand that the reconstruction matrix R should be with blockwise structures. This structure can be achieved by low-rank restriction, and we can reformulate (2) as

$$\min_{T,R} \text{rank}(R), \quad \text{such that } T^T X_t = T^T X_s R. \tag{3}$$

This equation is nonconvex and cannot get the globally optimal solution. Moreover, the solution of the problem is NP-hard. However, (2) is equivalent to

$$\min_{T,R} ||R||_*, \quad \text{such that } T^T X_t = T^T X_s R \tag{4}$$

where $||R||_*$ means the nuclear norm of R, which is a convex function. However, only introducing (3) and (4) is far from satisfying. This restriction just models the relationship between the source and target domains considering the neighboring relations on the subspace. As we want to represent the target samples on the subspace with the fewest neighbors to preserve the local structure of data, we further introduce the sparse constraint into the reconstruction matrix R as follows:

$$\min_{T,R} \ ||R||_* + \alpha||R||_1, \quad \text{such that } T^T X_t = T^T X_s R. \tag{5}$$

Another variable E is also introduced to model the noise of the target HSI on the subspace. To alleviate the

noise influence, we impose the sparse constraint on E and (5) is further modified as

$$\min_{T,R,E} \ ||R||_* + \alpha||R||_1 + \beta||E||_1$$
$$\text{such that } T^T X_t = T^T X_s R + E. \tag{6}$$

Here α and β are both heuristically set as 0.1, as we observe in our preliminary study that classification performance of our method is relatively stable when the parameters are in a feasible range. We assume that training samples can be transformed into the strict binary label matrix Y_s as

$$\theta(T, X_s, Y) = ||T^T X_s - Y_s|| + ||T||_1. \tag{7}$$

By minimizing (7), we can obtain the optimal Transformation matrix T. And the final objective function on the subspace should be

$$\min_{T,R,E} \theta(T, X_s, Y_s) \ + ||R||_* + \alpha||R||_1 + \beta||E||_1$$
$$\text{such that } T^T X_s = T^T X_t R + E. \tag{8}$$

We optimize (8) by employing the inexact augmented Lagrange multiplier (IALM) algorithm, as shown in [6].

After finding the optimal subspace, we employ the support vector machine (SVM) [10] as the final classifier on this subspace to get the label Y_t for the data in the target domain.

C. Constraints on the Original Target Space

On the above subspace, we fail to consider the local structure of HSI in the target domain, although the data is best represented by exploiting both its global and local structures of the HSI in the source domain. Therefore, besides knowledge transfer, we could further optimize the classification through preserving the self structure knowledge of data in the target domain by exploring MRF. The MRF describes the following energy minimization problem:

$$N = N_d + \lambda N_s \tag{9}$$

where N_d is the data term representing the likelihood of the objective data, and N_s is the smoothness term reflecting the joint Gibbs distribution of the label field which satisfies the Markov property [11]. λ defines the weights of these terms (here we consider the two terms equally and set λ as 1). This function is first initialized by the initial results Y_t obtained from the subspace. By finding the minimum solution of the energy function N, the corresponding label field can be acquired.

In this energy function, we further define the smoothness term N_s by the ISING model to model the structure of neighboring pixels in HSI, and the data term N_d by the Gaussian mixture model(GMM) as HSI is generally formed by the mixture of several categories. To be more specific, N_s formulated by the ISING model is expressed as

$$N_s = \sum_{a,b \in C} V_c \ (Y_{t_a}, Y_{t_b})$$
$$\text{such that } \ V_{a,b}(Y_{t_a}, Y_{t_b}) = \begin{cases} \frac{-\rho}{n}, & \text{if } Y_{t_a} = Y_{t_b} \\ \frac{\rho}{n}, & \text{if } Y_{t_a} \neq Y_{t_b} \end{cases} \tag{10}$$

where C means the set of cliques in a specific neighborhood system, a and b mean a random center pixel and its neighboring pixel in the image, and Y_{t_a} and Y_{t_b} are their corresponding labels. n is the order of neighboring, which means the distance between a and b. In fact, this formulation means that the larger the distance between a specific pixel and the center is, the less effect it has. We treat pixels with different orders differently by assigning various weights to them. Such weights reflect the significance of neighboring pixels in the system. The data term N_d is used to penalize solutions which are not consistent with the prior knowledge. More specifically, we have

$$N_d = \Sigma_{a \in C} w_a(Y_{t_a})$$
$$\text{such that } w_a(Y_{t_a}) = f(a \mid \theta_{Y_{t_a}}) \tag{11}$$

where $w_a(Y_{t_a})$ shows the cost of assigning label Y_{t_a} to pixel a. $f(\cdot)$ means the probability density function of the GMM distribution, and $\theta_{Y_{t_a}}$ denotes the parameter set that we need to construct this GMM model. We optimize the problem in (9) by the electromagnetic method [8] and obtain the final labels for HSI data on target domain based on its structure in its original data space. A brief pseudocode for the whole process is further concluded in Algorithm 1.

Algorithm 1 Structure Preserving Transfer Learning for HSI Classification

Input: $X_s, X_t, Y_s, \alpha, \beta, \lambda$
Output: T, R, E, Y_t
1: $T, R, E \leftarrow$ INITIALIZATION($X_s, X_t, Y_s, \alpha, \beta$)
2: Find the subspace by T, R, E
3: On the subspace, get the initial label Y_t by SVM classifier
4: On the original space, $Y_t \leftarrow$ OPTIMIZATION(Y_t, λ)
5: **function** INITIALIZATION($X_s, X_t, Y, \alpha, \beta$)
6: calculate $\arg\min_{T,R,E} \theta \ (T, X_s, Y_s) + ||R||_* + \alpha||R||_1 + \beta||E||_1$ in (8) by IALM
7: **end function**
8: **function** OPTIMIZATION(Y_t, λ)
9: $N_s = \sum_{a,b \in C} V_c \ (Y_{t_a}, Y_{t_b})$, $N_d = \Sigma_{a \in C} w_a(Y_{t_a})$,
10: $Y_t = \arg\min_{Y_t} N(Y_t) = N_d(Y_t) + \lambda N_s(Y_t)$
11: **end function**

III. Experiments and Results

In this section, we will show the experimental results of the proposed method when compared with several existing transfer learning methods. To verify the performance, we apply the proposed DSTL to hyperspectral classification. Although the main classifier SVM in the proposed framework is theoretically a supervised method, in our problem, the training samples come from the source domain and the testing samples come from the target domain. In this way, we can still claim the classification of HSI in the target domain as unsupervised. We compare the proposed framework with several approaches, including the recently proposed unsupervised classification methods k-means++ [12], [13] (which was shown to yield the best performance for unsupervised HSI classification) and two transfer learning based classification methods, direct non-subspace transfer learning (DTL) [14] and Subspace transfer learning (STL) [6]. It is worth noting that it is the first time

TABLE I

CLASSIFICATION ACCURACY RESULTS ON *Salinas* (BEST PERFORMANCE IS EMPHASIZED BY BOLDFACE.)

Salinas	k-means++ [13]	DTL [14]	STL [6]	DSTL
L1, L2 → L3, L4	0.9285	0.2941	0.9388	**0.9528**
L1, L3 → L2, L4	0.9833	0.9903	0.9995	**0.9998**
L1, L4 → L2, L3	0.9983	0.7081	0.9395	**0.9986**
L2, L3 → L1, L4	0.9822	0.9944	0.9963	**0.9992**
L2, L4 → L1, L3	0.9985	0.6089	1.0000	**1.0000**
L3, L4 → L1, L2	09776	0.9530	0.9533	**0.9933**

TABLE II

CLASSIFICATION ACCURACY RESULTS ON *PU* AND *PC* (BEST PERFORMANCE IS EMPHASIZED BY BOLDFACE.)

Pavia	PC → PU	PU → PC
k-means++ [13]	0.4231	0.5433
NMF [13]	0.5497	0.5623
DTL [14]	0.1822	0.1712
STL [6]	0.6288	0.6677
DSTL	**0.6537**	**0.6971**

that DTL and STL are applied here, as transfer learning has not been explored for our specific problem yet.

A. Data Sets

Three publicly available HSIs are tested to illustrate the superiority of the proposed method. We conduct the experiments on two types of data: different categories on the same HSI, Salina scene; and the data from two different HSIs, Pavia Center (PC) and Pavia University (PU). A common characteristic of these two type of data is that they are with the same dimension. The detailed descriptions are as follows.

The *Salinas scene* was collected by the Airborne Visible Infrared Imaging Spectrometer sensor on Salinas Valley, California. The image is characterized with 224 spectral bands, which comprises 512 × 217 samples. The image contains 16 classes of interest. We select four different categories of Lettuce romaines (categories 11–14), which are difficult to distinguish on Salinas to conduct our experiment. These four categories are abbreviated as L1, L2, L3, and L4, respectively.

The PC and PU are two scenes acquired by the Reflective Optics System Imaging Spectrometer sensor during a flight campaign over Pavia, nothern Italy. They have seven same classes of interest, and we choose them to complete the experiment. The number of spectral bands is 102 for PC and 103 for PU (we select 102 bands to conduct the experiment). PC is a 1096×1096 pixels image, and PU is with 610×610 pixels, but some of the samples in both images contain no information and have to be discarded before the analysis. The geometric resolution is 1.3 m. Both images include nine classes each. We abbreviate the two data set as PC and PU, respectively.

B. Experiments on Salinas

We first conduct the experiments on *Salinas*, and report the results in Table I. In Table I, we can observe that the proposed DSTL consistently yields the best performance and DTL seems the weakest. The classification between L3 and L4 as well as L2 and L3 are most challenging, as L2 has similar characteristics to that of L3 and L4.

The significance of structure preserving can be supported by comparing STL with DTL. Without constructing subspace learning, barely transferring the knowledge from the source to the target cannot provide good performance. Take the classification of L3 and L4 as an example. The classification between L3 and L4 is 29.41% which is quite poor. The performance is improved to 93.88% by introducing subspace learning and exploiting the structure information of HSI on the source domain. The results are further improved by structure preserving of the HSI in the target domain. When MRF is employed, the accuracy boosts to 95.28%, which is promising.

By comparing the proposed DSTL with the k-means method, we can find that with the assistance of transfer learning and knowledge learned from the source domain, the classification accuracy of HSI on the target domain is improved. In most time, the improvement is 1%–2%, and the final accuracy is near 100%.

C. Experiments on Pavia University and Center

Second, we conduct our experiment on the PU and PC data. The overall results can be found in Table II. The proposed framework is compared with two traditional unsupervised methods including k-means, nonnegative matrix factorization (NMF), as well as DTL and STL. We compare the overall accuracy of each method on both data. For the transfer learning based methods (DTL, STL, and DSTL), the labeled data from the source domain are 100 samples for each class on PU (when pavia university is the source domain), and 200 samples on PC (when PC is the source domain). The influence of the number of labeled data will be further discussed in Section III-D.

As shown in Table II, the classification performance on PC is always better than that of PU since the former one is with more complex features.

By comparing the unsupervised method k-means and NMF with the proposed DSTL, we can note that introduction of transfer learning and the assistance of the knowledge from the source domain data can boost the performance of HSI classification. Although the two data are with quite different characteristics, the prior knowledge can still be exploited. Barely imposing the k-means or NMF is far from satisfying which barely considers the features of data and without manual labeling.

The comparison between DTL, STL, and DSTL illustrates the significance of each component of the proposed method. Without subspace learning, the performance of DTL is the worst. The accuracy of DTL is about 18% which is almost the random accuracy. Such a poor accuracy also suggests that the two data are unrelated. By introducing the subspace learning process, the relationship between the two data is established and the accuracy increases to more than 60%. By applying the restriction of structure preserving on the target domain, the accuracies of classification on both data further increase by 3% which are about 70%. This accuracy is almost the same as that of some supervised classification methods.

TABLE III
CLASSIFICATION ACCURACY RESULTS WHEN USING DIFFERENT NUMBERS OF LABELED DATA FROM THE SOURCE DOMAIN (BEST PERFORMANCE IS EMPHASIZED BY BOLDFACE.)

Data	samples/class	50	100	150	200	250	300	350	400
PC → PU	STL	0.4133	0.5204	0.5964	**0.6288**	0.6104	0.6042	0.5923	0.5523
PC→ PU	DSTL	0.3925	0.4963	0.6355	**0.6537**	0.6522	0.6436	0.6448	0.6343
PU→ PC	STL	0.5876	**0.6677**	0.6632	0.6567	0.6388	0.6201	0.6064	0.6206
PU → PC	DSTL	0.6322	**0.6971**	0.6823	0.6422	0.6653	0.6731	0.6304	0.6452

D. Influence of Labeled Data

The labeled data in the source domain always could influence the classification accuracy in the target domain. We take the transfer between PU and PC as an example to further discuss this issue in Table III. In Table III, three observations can be noticed. The first is that the performance of transfer learning is not always better with more labeled samples from the source domain. The accuracy might increase at first with more labeled data but fall later. This phenomenon is likely caused by overfitting. With too many labeled data, the learned subspaces are more likely to be compatible with the distribution of the data in the source domain. When we classify the data from the target domain, the accuracy might decrease. We can note that the best number of labeled data for PC to PU transfer learning is 200 samples/class, and it is 100 samples/class for PU to PC transfer learning.

The second observation is that the process of structure preserving on the target domain is not always beneficial for classification. When the initial classification result Y_t on the subspace is poor, the final accuracy might decrease after this structure preserving process. One example can be found by comparing STL and DSTL of PC → PU. When the labeled data are no more than 100 samples/class, the accuracy of DSTL is lower than that of STL. This phenomenon is likely caused by the characteristic of the structure preserving process. When the accuracy of classification is not high and the error happens too frequently, the error may spread during the structure preserving process. However, with higher accuracy, this phenomenon can be overcome.

At last, we have to point out that the best number of labeled data from the source domain is not stable and varies according to the volumes of the source data and target data. We cannot generally conclude which domain influences more. More data from the source domain can provide more knowledge to transfer and more data from the target domain requires more robust subspace. This property will be further exploited in our future work.

IV. CONCLUSION

In this letter, a novel method named the DSTL is proposed for unsupervised HSI classification. The main idea is to transfer the knowledge of HSI in the source domain to the target domain, to perform the classification with no prior information and thus address the time and labor-consuming HSI labeling problem.

The proposed method consists of two major parts. The first is to transfer the data on both domains to a specific subspace, on which we can obtain the initial classification results for the target HSI by exploiting the data structure. The second is to optimize the initial results on the original target data space based on its structure, by applying the MRF approach.

As an unsupervised HSI classification method, the proposed DSTL is robust and effective, as supported by the experimental results. Extensive comparisons also demonstrate the superiority over the competitors.

However, limitations also exist. Regarding the CPU time cost and computational complexity, the proposed DSTL is not as efficient as some regular unsupervised classification methods due to its subspace learning process, since such a process may take tens of seconds. Our future work is to further address this complexity problem.

REFERENCES

[1] M. Kan, J. Wu, S. Shan, and X. Chen, "Domain adaptation for face recognition: Targetize source domain bridged by common subspace," *Int. J. Comput. Vis.*, vol. 109, no. 1, pp. 94–109, 2014.
[2] Q. Wang, J. Fang, and Y. Yuan, "Adaptive road detection via context-aware label transfer," *Neurocomputing*, vol. 158, pp. 174–183, Jun. 2015.
[3] M. Shao, D. Kit, and Y. Fu, "Generalized transfer subspace learning through low-rank constraint," *Int. J. Comput. Vis.*, vol. 109, nos. 1–2, pp. 74–93, 2014.
[4] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2066–2073.
[5] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 999–1006.
[6] Y. Xu, X. Fang, J. Wu, X. Li, and D. Zhang, "Discriminative transfer subspace learning via low-rank and sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 850–863, Feb. 2016.
[7] J. Liu, Y. Chen, J. Zhang, and Z. Xu, "Enhancing low-rank subspace clustering by manifold regularization," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4022–4030, Sep. 2014.
[8] Y. Yuan, J. Lin, and Q. Wang, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2966–2976, Dec. 2016.
[9] Z. Ding and Y. Fu, "Robust transfer metric learning for image classification," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 660–670, Feb. 2017.
[10] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector-Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[11] Q. Lu, X. Huang, J. Li, and L. Zhang, "A novel mrf-based multifeature fusion for classification of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 515–519, Apr. 2016.
[12] B. Bharath, C. Nicolas, and L. Sébastien, "Unsupervised classifier selection approach for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2016, pp. 5111–5114.
[13] W. Zhu *et al.*, "Unsupervised classification in hyperspectral imagery with nonlocal total variation and primal-dual hybrid gradient algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2786–2797, May 2017.
[14] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.