Spectral and Spatial Proximity-Based Manifold Alignment for Multitemporal Hyperspectral Image Classification

Hsiuhan Lexie Yang, *Member, IEEE*, and Melba M. Crawford, *Fellow, IEEE*

Abstract—Multitemporal hyperspectral images provide valuable information for a wide range of applications related to supervised classification, including long-term environmental monitoring and land cover change detection. However, the required ground reference data are time-consuming and expensive to acquire, motivating researchers to investigate options for reusing limited training data for classification of other temporal images. Current studies that address high dimensionality and nonstationarity inherent in temporal hyperspectral data for classification are limited for the case where significant spectral drift exists between images. In this paper, we adapt and extend two manifold alignment (MA) methods for classification of multitemporal hyperspectral images in a common manifold space, assuming that the local geometries of two temporal spectral images are similar. The first method exploits a locally based manifold configuration of a source image (considered to be the "prior" manifold), and the second approach links local manifolds of two images using bridging pairs. In addition to exploiting manifolds estimated with spectral information for MA, we also demonstrate how spatial information can be incorporated into the MA methods. When evaluated using three Hyperion data sets, the proposed methods outperform four baseline approaches and two state-of-the-art domain adaptation methods. The advantages of the proposed MA methods are more evident when significant spectral drift exists between two temporal images. In addition to the promising classification results, the proposed methods establish a domain adaptation framework for analysis of temporal hyperspectral data based on data geometry.

Index Terms—Classification, domain adaptation, manifold alignment (MA), multitemporal hyperspectral, transfer learning.

I. INTRODUCTION

HYPERSPECTRAL sensors acquire data over narrow contiguous bands of the electromagnetic spectrum. The detailed spectral information provides new opportunities to identify objects that exhibit distinguishable spectral characteristics associated with absorption and reflectance features. In remote sensing, hyperspectral imagery contributes enhanced capability for land cover classification, as high spectral resolution enables discrimination of classes that are difficult to characterize with multispectral images. Airborne hyperspectral sensors are now flown by government agencies for civilian and military applications, and commercial hyperspectral data acquisitions are commonplace, although expensive. With advances in remote sensing technology, spaceborne hyperspectral sensors provide capability to acquire a set of images that have both high spectral and temporal resolution. These data are suitable for monitoring and analyzing environmental changes with subtle spectral characteristics. The EO-1 mission, with the first operationally successful hyperspectral instrument, was launched in 2001 and paved the way for multitemporal hyperspectral sensing from space. Future launches of hyperspectral imagers are planned by several countries.

To fully exploit multitemporal hyperspectral data, we must address two issues: 1) high dimensionality and 2) nonstationarity within temporal sequences. To deal with the high-dimensionality issue, a dimension-reduction process is required prior to analysis, with the goal of extracting the most relevant information to characterize data in a lower dimensional space. For classification, one of the benefits of exploring the low-dimensional space is that fewer training samples are required to obtain a reliable classifier [1], [2]. While improved discrimination is possible by using hyperspectral data, the impact of nonstationary phenomena is particularly significant due to the narrow spectral bands. Spectral signatures are subject to change over time due to natural (e.g., seasonal phenology of vegetation or environmental conditions) and disruptive (e.g., fire, anthropogenic) impacts. In a two-image case, a source image $\mathbf{X}^s \in \mathbb{R}^p$ and a target image $\mathbf{X}^t \in \mathbb{R}^p$, for example, temporal nonstationarity can result in differences in the optimally reduced dimension space, resulting from feature selection or extraction. It is not appropriate to directly reuse a pretrained classifier that is obtained with features $\mathbf{H}^s$, given by a feature mapping $h^s : \mathbf{X}^s \to \mathbf{H}^s$ of $\mathbf{X}^s$, to classify $\mathbf{X}^t$, whose feature space is defined by a different parameterized function $h^t : \mathbf{X}^t \to \mathbf{H}^t$. For this reason, additional ground reference data are needed for classification of $\mathbf{X}^t$. However, it is expensive and time-consuming to obtain training samples for each image in practical applications, particularly if the classes are not spatially consistent across the series of images.

Several approaches have been proposed to tackle these issues. Natural temporal variation often results in smoothly varying trends over time, making updating schemes viable [3]–[5]. Inspired by the semisupervised learning philosophy,

these approaches gradually incorporate unlabeled samples from the new temporally related images into a pretrained classifier or feature extractors, so that they can be generalized to the new image. For these cases, well-posed initial conditions are essential for the updating scheme. Because significant spectral differences can result from dynamic class-dependent processes within a temporal series of images, these updating schemes may be ineffective [5], [6]. In addition, disparities in the probability distributions of $\mathbf{X}^s$ and $\mathbf{X}^t$ may limit convergence of those methods, as also seen in semi-supervised learning [7], [8]. More recently, active learning has been investigated in the context of domain adaptation in the remote sensing arena [9]. Here, the most informative samples from $\mathbf{X}^t$ are identified and labeled manually to help correctly classify $\mathbf{X}^t$.

In addition to the aforementioned two issues associated with multitemporal hyperspectral image analysis, we also need to consider nonlinearity within hyperspectral data. The nonlinearity results from nonlinear scattering associated with the bidirectional reflectance distribution function [10], nonlinearly attenuated radiance in water bodies [11] and other sources [12]. Recently, nonlinear manifold learning has been proposed as a successful dimension-reduction method to explore intrinsic information concealed in high-dimensional data and to capture nonlinearity within hyperspectral data, whereby data are projected onto a lower dimensional feature space that retains critical geometric structures of the data. Although manifold learning has been applied successfully in single-image hyperspectral data classification [13]–[16] to address the high-dimensionality issue, research related to manifold learning for multitemporal classification studies is limited [4]. As data-driven methodologies, manifold learning methods often yield distance manifolds whose correspondence is difficult to model, particularly for environments that vary over space and time. One may attempt to develop a joint manifold representation by exploiting spectrally or spatially "nearby" samples across two data sets. However, spectral drift resulting from nonstationarity within a temporal sequence can lead to changes in the probability density functions of the spectral signatures of the classes. Therefore, simply searching for neighbors within a cluster (high-density region) of spectral data may be not a reliable means of deriving joint data manifolds. These characteristic are major barriers to successful multitemporal image classification using manifolds. Assuming that the class types are the same and the intrinsic local structures of the classes contained in multiple images are similar, a proper cluster describing the same class may be obtained by aligning the local geometric structures of the spectral data. Global spectral discrepancies associated with class-dependent signature drift across a series of images may be reduced by grouping spectral neighbors within each image and mapping these clusters into an adequate common latent space supported by these local structures.

In this paper, we propose two manifold alignment (MA) techniques that involve aligning underlying local manifolds of temporally sequential data sets. The first approach exploits a local-based manifold of a source image, considered to be the optimal "prior" manifold that cannot be intuitively reused by another image. The second method links local manifolds of two images using bridging pairs by considering spectral and spatial relationships between two temporal images. The proposed methods extend graph-based semi-supervised learning and explore MA [17]–[19] for the multitemporal hyperspectral image classification task, while providing a domain adaptation framework [20] for hyperspectral image analysis from the geometric learning point of view. The proposed methods are also compared with several baseline approaches and the state-of-the-art domain adaptation methods in order to validate the effectiveness. The quality of data manifolds is related to spectral neighbors, which are critical for MA. For remotely sensed images, local spatial information also provides important input for determining spectral neighbors. Incorporation of spatial proximity into the proposed MA framework is also investigated.

This paper explores the data geometry problem where only one labeled image is assumed to be available, and the goal is to classify another image. The problems of 1) limited training samples and 2) domain adaptation exploiting data geometry through successive images are addressed. The paper is organized as follows. Literature pertaining to multitemporal image analysis is reviewed, and gaps in existing capability for applying manifold learning to multitemporal hyperspectral image classification are discussed in Section II. The proposed framework for addressing these issues is outlined in Section III. The experiments that were conducted in the proposed framework are described in Section IV, and results are presented and discussed in Section V. Section VI summarizes the contributions and outlines possible future directions for the research.

II. BACKGROUND AND MOTIVATION

The following sections review important studies related to multitemporal image classification and manifold learning. We also discuss the unresolved issues in the existing research relative to our problem of interest.

A. Multitemporal Image Classification

Multitemporal image analysis has been actively explored in the remote sensing community. Analysis of land cover signatures from images acquired over the same location at different times provides capability to study multitemporal processes. However, nonstationarity of spectral signatures in multitemporal data limits effective utilization of multitemporal images. Many methods attempt to overcome this spectral shift problem for classification. In the early work by Swain [21], a cascade method was implemented to estimate class labels of new images based on a Bayesian strategy using past classification results. A land cover transition matrix that models temporal dependence between two temporal images can be estimated using Bayes rule [22] or the expectation–maximization (E–M)-based algorithm [23]. A class of popular approaches updates the classifier's parameters by retraining a predefined classifier [24]. Unlabeled samples are included through semi-supervised learning to adapt the predefined classifier to the new scene. This updating scheme has been successfully demonstrated for both parametric and nonparametric classifiers [25]. Algorithms incorporating recent machine learning developments have also been proposed. In particular, advances in kernel machines

provide opportunities for effective analysis of multitemporal remote sensing data. The decision boundary in the kernel space can be redefined with an updating scheme, as shown in [3], [4]. Domain adaptation support vector machine (SVM) proposed in [3] gradually incorporates reliable semi-labeled samples [26] from the target image and deletes samples from the source image to refine the decision boundary in the kernel space. Kernel-based methods have also been used to fuse information of two temporal images with composite kernels in [27]. These approaches, however, were applied to multitemporal multispectral images. Feature extraction is commonly utilized as a preprocessing step in classification of hyperspectral data, but the resulting feature spaces differ if signatures change between images. Constrained by the feature extraction step, updating classifiers or estimates of probability distributions is nontrivial when the feature space changes, as compared with operations in the original spectral space. An attempt to utilize extracted features was made in [5], where samples from new images were used to update the Fisher discriminant features in the binary hierarchical classifier via the E–M algorithm for classification of multitemporal hyperspectral data. Kim and Crawford [4] demonstrated an adaptive kernel classifier with regularized manifolds for spatially disjoint hyperspectral data sets, as well as for temporal hyperspectral images. In [4], [5], and [24], the requirement for good initial parameters or conditions to ensure learning and achieving convergence was emphasized. That is, updating schemes may not be efficient or may converge to incorrect final values if significant changes occur in spectral signatures. Some approaches that minimize differences between two distributions [28], [29] might be applicable for multitemporal hyperspectral image analysis. However, those methods typically require access to labeled data from each image. In some practical cases, the targeted geographical region may only have training samples for one time period; however, exploration of other temporal images is desired.

B. Manifold Learning

For the classification or clustering problem, data samples that share the same labels are assumed to be close in a feature space. Machine learning approaches that seek to retain the essential clustering conditions based on distances and find a lower dimensional feature space are known as manifold learning [30], [31]. A data manifold is expected to be geometrically supported by data points. The greater the number of data points, the better the data structure is represented, as long as the points are not redundant and are well distributed across the data manifold. This philosophy potentially enables us to make use of the large quantities of available unlabeled samples, in addition to the labeled data in the scene.

Connections defined by distances between data points are usually used to uncover the data manifold. Popular manifold learning methods can be generally formulated in a graph embedding framework to exploit distance and point relationships [32]. Let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^l \in \mathbb{R}^p$ denote l pixels from a p-band hyperspectral image. In the graph embedding framework, we denote the undirected graph $G = (V, \mathbf{W})$, where a vertex $v_{i,j} \in V$ indicates that data points $\mathbf{x}_i$ and $\mathbf{x}_j$ are connected based on a

local neighborhood set $\mathrm{N}(\mathbf{x}_i) = \{\mathbf{x}_j \in \mathbf{X} \backslash \mathbf{x}_i\}$. A nonnegative weight $w_{i,j} \in \mathbf{W}$ associated with each vertex $v_{i,j}$ of the graph symbolizes the strength of the associated spectral similarities of $\mathbf{x}_i$ and $\mathbf{x}_j$. With a Gaussian function, spectral similarity of $\mathbf{X}$ can be represented in the l-by-l $\mathbf{W}$ matrix, whose elements are pixel-wise similarity, i.e.,

$$w_{i,j} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_{\mathrm{spe}}^2}\right). \tag{1}$$

The similarity matrix implicitly codes geometric relationships of samples and can be used to estimate an underlying data manifold. A diagonal degree matrix $\mathbf{D}$ is defined as $\mathbf{D}(i,i) = \sum_j w_{i,j}$, and the graph Laplacian is denoted by $\mathbf{L} = \mathbf{D} - \mathbf{W}$. A lower dimension data manifold embedding $\mathbf{F} = \{\mathbf{f}_i\}_{i=1}^l \in \mathbb{R}^d$ for $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^l \in \mathbb{R}^p$ can be learned by minimizing a smoothness function $S = (1/2)\sum_{i,j}\|\mathbf{f}_i - \mathbf{f}_j\|^2 w_{i,j}$. The optimization problem is then

$$\mathbf{F} = \arg\min_{\mathbf{f}_i, \mathbf{f}_j} \frac{1}{2}\sum_{i,j}\|\mathbf{f}_i - \mathbf{f}_j\|^2 w_{i,j}$$
$$= \arg\min_{\mathbf{F}} \mathbf{F}^T \mathbf{L} \mathbf{F}. \tag{2}$$

The solution to (2) is composed of the smallest eigenvectors of L. Alternatively, one can use the normalized Laplacian $\tilde{\mathbf{L}} = \mathbf{D}^{1/2}\mathbf{L}\mathbf{D}^{1/2}$ and solve the generalized eigenvalue problem $\tilde{\mathbf{L}}\mathbf{f} = \lambda\mathbf{D}\mathbf{f}$. The manifold developed based on the graph Laplacian, as in (2), can be also interpreted equivalently as kernel principal component analysis (PCA) from the kernel learning point of view. Readers are referred to [33] for more details.

By minimizing (2), the local spectral neighbor pair $(\mathbf{x}_i, \mathbf{x}_j)$ that is associated with larger values of $w_{i,j}$ (high similarity implied) results in small values of $\|\mathbf{f}_i - \mathbf{f}_j\|$ and thus are likely to be neighbors in the manifold approximated by $\mathbf{F}$. In this manner, clusters of $\mathbf{F}$ contribute to classification naturally. As previously discussed, a data manifold strongly depends on the smoothness criterion. For the single-image scenario, it is appropriate to impose a smoothness condition with spectral similarities between data points so that samples potentially sharing the same class labels meet the clustering assumption in the resulting manifolds. However, as previously noted, spectral drift may occur over temporal images, leading to unreliable spectral similarity of data points from different images. As shown in a two-class toy example in Fig. 1, the spectral drift results in differences in the two conditional distributions of Class 1 C_1, i.e., $p(\mathbf{X}^s|C_1)$ and $p(\mathbf{X}^t|C_1)$. Furthermore, the drifting phenomenon may cause the conditional distributions of Class 1 C_1 and Class 2 C_2 to overlap in some region, such as $p(\mathbf{X}^t|C_1)$ and $p(\mathbf{X}^s|C_2)$ in Fig. 1. Some samples of C_1 from $\mathbf{X}^t$ might be assigned as local neighbors of samples of C_2 from $\mathbf{X}^s$, if they are in the overlap region. Therefore, retaining nearest spectral neighbors that depend on this misleading cluster condition and enforcing smoothness across data sets is inadequate for multitemporal hyperspectral image analysis.

Another possibility is to explore data manifolds that are developed separately. However, as data-driven methodologies, data manifolds developed with individual images may be represented in different manifold coordinates whose
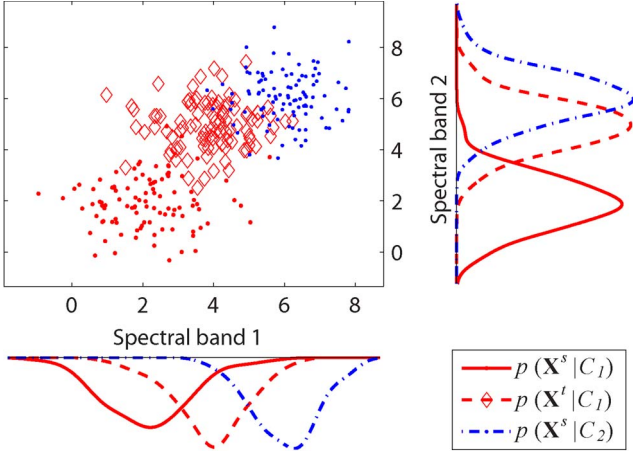
Fig. 1. Toy example of spectral shift. Blue and red lines indicate two classes, and dots and diamonds indicate two data sets.

correspondences are not easy to model [34]. However, it may be reasonable to assume that the intrinsic local structures of classes contained in these sequential images are similar. With this assumption, a cluster describing the same class can be obtained by aligning the geometric resemblances of temporal images in a common manifold space. The impact of spectral discrepancies across a series of images can then be reduced by grouping spectral neighbors within each image and mapping these clusters into the common latent space. Aligning image manifolds in a joint latent space is also proposed in [35] for handling multiple sensor problems.

This framework for exploring similar intrinsic geometries of images is referred to as MA by the machine learning community [17]–[19]. The goal of MA is to build a mapping between two or more manifolds for data that originate from similar sources. MA research has been conducted in other domains such as protein structure comparisons [36] and video sequence alignment [37]. However, to date, few studies have investigated MA in the remote sensing community [35], [38], [39].

III. PROPOSED METHOD

Let $\mathbf{X}^s = \{\mathbf{x}_i^s\}_{i=1}^m \in \mathbb{R}^p$ and $\mathbf{X}^t = \{\mathbf{x}_j^t\}_{j=1}^n \in \mathbb{R}^p$ denote m and n samples from two temporal hyperspectral images, the source and target images, respectively. Assuming that the local geometric properties of the two multitemporal images are similar, we propose two alignment approaches with prior manifold representations and bridging pairs [17], respectively. Both MA algorithms utilize labeled data and preserve underlying local geometries in a joint latent space.

A. MA With Prior Manifold

In practical cases, certain prior information is often available for multitemporal image analysis, e.g., labeled samples from the source image. Because the local geometry within the two images is assumed to be similar, the data manifold depicting the geometric structure of the source image is regarded as a set of informative features. The manifold representation $\mathbf{M}^s = \{\mathbf{m}_i^s\}_{i=1}^m \in \mathbb{R}^{\hat{d}}$ of $\mathbf{X}^s$ obtained by some local-based manifold learning methods such as locally linear embedding (LLE) [30],

local tangent space alignment (LTSA) [40], and Laplacian eigenmaps (LE) [41] are assumed to represent this prior information. We can approximate manifold representations using labeled samples from the source image and estimated $\hat{d}$ based on overall accuracy. However, the manifold representation of the source image is not easy to exploit for the target image, and mapping functions between two feature spaces derived by nonlinear dimensional reduction approaches are particularly difficult to obtain.

To find the optimal joint manifold embeddings $\mathbf{F} = [\mathbf{F}^s; \mathbf{F}^t] \in \mathbb{R}^d$ in a joint manifold space that is supported by local structures of $\mathbf{X}^s$ and $\mathbf{X}^t$, the prior manifold configurations are utilized to constrain an optimization problem. A class of mappings $\{\mathbf{f}_i^s, \mathbf{f}_j^t | \mathbf{f}_i^s \in \mathbf{F}^s, \mathbf{f}_j^t \in \mathbf{F}^t; 1 \le i \le m, 1 \le j \le n\} \in \mathbf{F}$ matching the similar local structures is given by minimizing the cost function

$$C(\mathbf{F}) = \sum_i^m \mu \|\mathbf{f}_i^s - \mathbf{m}_i^s\|^2 + \mathbf{F}^T \mathbf{L} \mathbf{F}$$

where

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}^{s,s} & \mathbf{L}^{s,t} \\ \mathbf{L}^{t,s} & \mathbf{L}^{t,t} \end{bmatrix}. \tag{3}$$

The first term in (3) controls the fitting error as in a regression problem and preserves information contained in the prior manifold features $\mathbf{M}^s$, whereas the second term ensures the smoothness of the aligned manifold on the combined Laplacian graph. These two penalty terms are balanced by a tradeoff weight coefficient μ. The optimal solution of $\mathbf{F}$ is then derived and used for classification, i.e.,

$$\mathbf{F} = \begin{bmatrix} \mu I + \mathbf{L}^{s,s} & \mathbf{L}^{s,t} \\ \mathbf{L}^{t,s} & \mathbf{L}^{t,t} \end{bmatrix}^{-1} \begin{pmatrix} \mu I \\ 0 \end{pmatrix} \mathbf{M}^s. \tag{4}$$

To reduce computation on tuning μ and fully exploit the prior manifold, we can assume that embedding $\mathbf{F}^s$ for the source image is the same as the prior manifold $\mathbf{M}^s$ by setting $\mu \to \infty$. $\mathbf{F}^t$ is then given by

$$\mathbf{F}^t = -(\mathbf{L}^{t,t})^{-1}(\mathbf{L}^{t,s})\mathbf{M}^s$$
$$= (\mathbf{L}^{t,t})^{-1}(\mathbf{W}^{t,s})\mathbf{M}^s. \tag{5}$$

The dimension of the inverse problem of (5) is smaller than (4). In (5), manifold configurations of $\mathbf{X}^t$ are functions of the extracted manifold coordinates $\mathbf{M}^s$, the local structures of $\mathbf{X}^t$ (i.e., $\mathbf{L}^{t,t}$), and the relationships between $\mathbf{X}^s$ and $\mathbf{X}^t$ (i.e., $\mathbf{W}^{t,s}$). That is, the alignment strategy establishes a joint manifold that incorporates the optimal manifold space and similar localities of the two images, as well as the important local geometries of the target data $\mathbf{X}^t$.

B. MA With Bridging Pairs

The second approach fuses two similar local geometries based on certain bridging pairs $(\mathbf{x}_B^s \leftrightarrow \mathbf{x}_B^t)$, where $\mathbf{x}_B^s \in \mathbf{X}^s$ and $\mathbf{x}_B^t \in \mathbf{X}^t$, between the two images. A bridging pair $(\mathbf{x}_B^s \leftrightarrow \mathbf{x}_B^t)$ here is defined as an instance $\mathbf{x}_B^s$ from $\mathbf{X}^s$ that has a high possibility of sharing the same class label as $\mathbf{x}_B^t$ from $\mathbf{X}^t$. We

expect to align two similar local geometries with bridging pairs, as two samples in a bridging pair are assumed to be as close as possible. With this concept, the embeddings $\mathbf{F} = [\mathbf{F}^s; \mathbf{F}^t] \in \mathbb{R}^d$ in the common latent manifold space are discovered by minimizing

$$C(\mathbf{F}^s, \mathbf{F}^t) = \sum_B \mu \left\| \mathbf{f}_B^s - \mathbf{f}_B^t \right\|^2 + \mathbf{F}^{s^T} \mathbf{L}^s \mathbf{F}^s + \mathbf{F}^{t^T} \mathbf{L}^t \mathbf{F}^t \quad (6)$$

$\mathbf{F}^s = \mathbf{f}_B^s \cup \mathbf{f}_b^s$ and $\mathbf{F}^t = \mathbf{f}_B^t \cup \mathbf{f}_b^t$ are the resultant features of $\mathbf{X}^s$ and $\mathbf{X}^t$, respectively, in the common manifold, where B and b are index members included or excluded from the set of bridging pairs, respectively. $\mathbf{L}^s$ and $\mathbf{L}^t$ are graph Laplacian matrices that represent local characteristics of the images. The criterion can be interpreted as the cost function that imposes the manifold smoothness $\mathbf{F}^{s^T} \mathbf{L}^s \mathbf{F}^s$ and $\mathbf{F}^{t^T} \mathbf{L}^t \mathbf{F}^t$ and preserves correspondences by constraining $\|\mathbf{f}_B^s - \mathbf{f}_B^t\|$ to be small. The new embedding space obtained by optimizing (6) can be seen as a joint manifold where the two disparate graph Laplacian matrices $\mathbf{L}^s$ and $\mathbf{L}^t$ align on the basis of bridging pairs. Since labels of $\mathbf{x}_i^s$ are known, we can select an adequate number of bridging pairs for each class to ensure that each class and corresponding clusters can be anchored in the joint manifold.

Similar to MA with a prior manifold, we can assume that two samples in the bridging pairs are perfectly aligned, in order to reduce model selection efforts and computation time. Solutions to (6) without μ are obtained by solving the following problem [17]:

$$C(\mathbf{h}) = \frac{\mathbf{h}^T \mathbf{L} \mathbf{h}}{\mathbf{h}^T \mathbf{h}}$$

where

$$\mathbf{h} = \begin{bmatrix} \mathbf{f}_B^s = \mathbf{f}_B^t \\ \mathbf{f}_b^s \\ \mathbf{f}_b^t \end{bmatrix}$$

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{B,B}^s + \mathbf{L}_{B,B}^t & \mathbf{L}_{B,b}^s & \mathbf{L}_{B,b}^t \\ \mathbf{L}_{b,B}^s & \mathbf{L}_{b,b}^s & 0 \\ \mathbf{L}_{b,B}^s & 0 & \mathbf{L}_{b,b}^t \end{bmatrix}. \quad (7)$$

The embeddings of $\mathbf{X}^s$ and $\mathbf{X}^t$ in the joint latent manifold are based on the d smallest nonzero eigenvectors of $\mathbf{L}$.

The two proposed frameworks differ from standard manifold learning in the use of prior information (derived from the labeled data) and the consideration of local geometries of the spectra of both images. In addition, instead of adaptively updating the decision boundary using manifold regularization as in [4], the MA exploits similar underlying geometries of two data sets and aligns them in a common manifold space in one step, potentially avoiding the need to redefine classifier rules.

As compared with LapSVM [42], which is a benchmark of the semi-supervised graph learning framework, MA based on information from the manifold of the source image is supplied with additional information $\mathbf{M}^s$, which serves as a transfer learning mechanism via features. MA with bridging pairs extends the graph representation of LapSVM to the two image scenario and minimizes spectral discrepancies resulting from temporal spectral drift. Illustrations of the proposed MA algorithms are shown in Figs. 2 and 3, where diamonds and
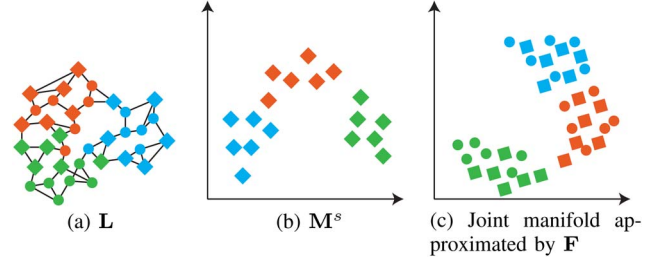

Fig. 2. MA with a prior manifold (derived from a previous data set). (a) Graph Laplacian $\mathbf{L}$ and (b) prior manifold $\mathbf{M}^s$ defined first. (c) Individual local clusters in the joint manifold indicating different classes.
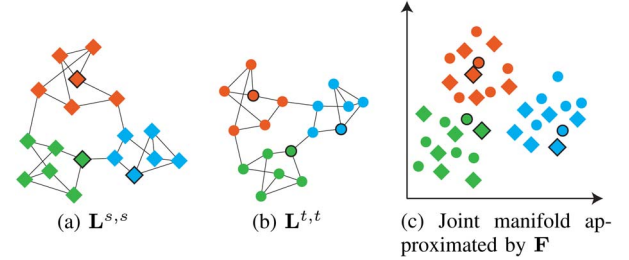

Fig. 3. MA with bridging pairs. Graph Laplacians (a) $\mathbf{L}^{s,s}$ and (b) $\mathbf{L}^{t,t}$ of two images are computed separately. (c) With some given bridging pairs (samples in black edges), one pair per class in this example, local manifolds implicitly coded in $\mathbf{L}^{s,s}$ and $\mathbf{L}^{t,t}$ are aligned in the joint manifold space. Samples in the same bridging pairs are forced to be as close as possible in the joint latent space.

circles indicate samples from the source and target images, respectively. Each color corresponds to one class. Compared with current studies of MA for remote sensing data, the work in [35] requires labeled samples from both the source and target images to perform MA, and the graph-matching method proposed in [39] needs k-means or fuzzy c-means as a vector quantification a preprocessing step, which may be inefficient for high-dimensional data. In addition, these two papers handle multispectral images, whereas our methods focus on hyperspectral data where nonlinear manifold learning is exploited.

C. Integrating Spatial Proximity into MA

As indicated in [14], manifold learning is sensitive to the choice of spectral neighbors. While exploiting spectral neighbors to build a graph Laplacian for MA, difficult classes, whose spectral signatures are similar, are often easily confused. That is, some spectral neighbors in $N(\mathbf{x}_i)$ may actually not share the same class label as $\mathbf{x}_i$. To address this potential issue, we propose to incorporate spatial proximity to stabilize spectral neighbor relations. While traditional methods that focus on spectral classification ignore spatial relations inherent in remote sensing images, recent research has demonstrated the value of spatial proximity for hyperspectral image classification [43], [44]. Similar to clustering with spectral neighbors, local neighborhood and geographical coordinates reveal a natural clustering condition whereby spatially contiguous pixels have higher probabilities of being from the same class.

Considering the spatial domain, each pixel $\mathbf{x}_i$ can be represented in a 2-D feature space that is associated with the geographical locations $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^l \in \mathbb{R}^2$. Spatial proximity then can be given by $\|\mathbf{z}_i - \mathbf{z}_j\|$ for a Euclidean distance metric.

TABLE I
HYPERION DATA PAIRS FOR MA EXPERIMENTS AND COVERAGE DESCRIPTION

	Source image	Target image	Remarks
Data Pair 1	May	June	Source and target images have a small overlap in spatial domain
Data Pair 2	June	July	Source and target images cover the same geographical area
Data Pair 3	May	July	Source and target images have a small overlap in spatial domain

Applying a Gaussian function, a candidate weighting function that considers spectral and spatial characteristics is given by

$$w_{i,j} = \exp\left(-\frac{\|\mathbf{z}_i - \mathbf{z}_j\|^2}{2\sigma_{\mathrm{spa}}^2}\right)\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_{\mathrm{spe}}^2}\right) + \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_{\mathrm{spe}}^2}\right) \tag{8}$$

where σ_{spa}^2 and σ_{spe}^2 are user-specified parameters. This weighting reflects our expectation of increased similarity when two samples are close in the spatial domain. The second term of (8) prevents $w_{i,j} = 0$ or $w_{i,j} \cong 0$ when pairs $(\mathbf{x}_i, \mathbf{x}_j)$ are spectrally similar but spatially distant. The weighting of (1) in the MA algorithms is then replaced by (8) for incorporating spatial and spectral proximity.

Inclusion of spatial proximity for improving the data manifold is also motivated by spatial regularization, where smoothness concept is not only in the spectral domain but also included in the penalty term that considers spatial consistency [43], [45].

IV. EXPERIMENTAL DATA AND SETTINGS

A. Data Description

The proposed MA frameworks were implemented on three Hyperion images collected over the Okavango Delta, Botswana, in May, June, and July of 2001. The Hyperion on the NASA EO-1 satellite acquires 224 bands at 30-m spatial resolution in 10-nm bands from 400 to 2500 nm. Removal of uncalibrated and noisy bands resulted in 145 bands ([10–55, 82–97, 102–119, 134–164, and 187–220]) that were used in the experiments. The images have nine common identified classes, as listed in Table I and displayed in Fig. 4. Among these three images, June and July data cover the same geographic area, but the May imagery is not exactly geographically colocated with the other two images because satellite pointing was changed. Three data pairs, each composed of a source image and a target image, were tested to validate the proposed MA approaches. The description of the data pairs is given in Table II. For each data pair, all labels of the target image were discarded, and all available labeled data of the source image were considered as the training data.

B. Experimental Settings

In our previous work [38], some popular local-based manifold learning methods (LLE [30], LTSA [40], and LE [41]) were evaluated for feature extraction in the MA with a prior manifold framework. MA with a local manifold based on LE provided the best results. Therefore, the prior manifold used in this paper was provided by LE.

The bridging pairs link two clusters from the source and target images if they are likely from the same class. In this paper, we select bridging pairs by considering both spectral proxim-
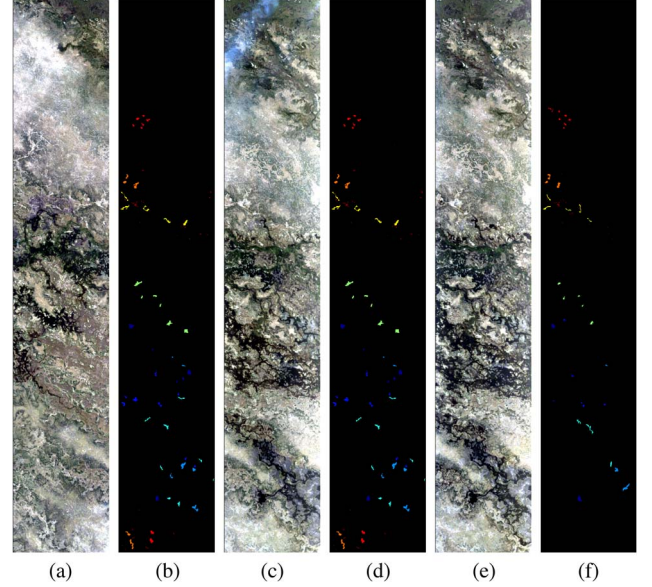

Fig. 4. Multitemporal Hyperion data. Red–green–blue composite images of (a) May, (c) June, and (e) July. Labeled data for (b) May, (d) June, and (f) July.

TABLE II
CLASS NAMES AND NUMBER OF LABELED
SAMPLES OF HYPERION DATA

ID	Class	May	June	July
1	Water	297	361	185
2	Primary Floodplain	437	308	96
3	Riparian	448	303	164
4	Firescar	354	335	186
5	Island Interior	337	370	131
6	Woodlands	357	324	169
7	Savanna	330	342	171
8	Short Mopane	239	299	152
9	Exposed Soils	215	229	96

ities and spatial locations to obtain a robust pairing condition for the multitemporal data sets. Let $\mathbf{Z}^s = \{\mathbf{z}_i{}^s\}_{i=1}^m \in \mathbb{R}^2$ and $\mathbf{Z}^t = \{\mathbf{z}_j{}^t\}_{j=1}^n \in \mathbb{R}^2$ represent geographical coordinates of m and n pixels of the source and target images, respectively. The B corresponding pairs $(\mathbf{x}_B^s \leftrightarrow \mathbf{x}_B^t)$ are determined by samples that have the B smallest values of $(\|\mathbf{x}_i^s - \mathbf{x}_j^t\| + a\|\mathbf{z}_i^s - \mathbf{z}_j^t\|)$, where a is a balancing coefficient. The two samples in a bridging pair should be spectrally similar and geographically close and would be informative for linking to local clusters from these two images. In this paper, a was set to 0.2 based on a grid search, and 30 bridging pairs per class were used.

The relative weighting coefficient μ in (3) and (6) was tested over a wide range of values, and the optimal μ tended to be large, which implies that the assumption of no uncertainty in (5) and (7) is justified for these data. Therefore, the MA results presented here were obtained using (5) and (7).

TABLE III
OVERALL PRODUCER/USER ACCURACIES AND KAPPA STATISTICS OF MAY/JUNE PAIR

May/June		Class ID		1	2	3	4	5	6	7	8	9
Spe.	Naive k-NN		Prod. Acc.	100.0	98.4	62.7	99.7	95.9	40.1	71.1	44.5	53.3
	OA/Kappa	75.6/0.72	User Acc.	99.7	63.8	64.6	98.5	61.5	53.9	74.1	100.0	100.0
	LE-pooled		Prod. Acc.	71.5	98.4	60.4	99.4	97.0	42.9	70.5	53.8	65.5
	OA/Kappa	74.1/0.71	User Acc.	99.6	53.2	66.1	89.5	68.4	53.7	80.9	100.0	100.0
	LapSVM		Prod. Acc.	100.0	98.7	57.8	99.1	98.6	60.2	72.8	52.8	37.6
	OA/Kappa	77.5/0.75	User Acc.	99.2	68.5	69.4	97.4	69.8	57.0	76.1	82.7	98.9
	k-NN-Stacked		Prod. Acc.	100.0	98.4	63.4	99.7	91.9	43.2	81.3	55.5	56.3
	OA/Kappa	78.1/0.75	User Acc.	99.7	68.4	65.5	96.3	66.1	57.1	78.5	100.0	87.8
	TCA		Prod. Acc	100.0	96.8	52.8	99.4	83.8	40.1	78.1	56.5	81.2
	OA/Kappa	77.1/0.74	User Acc.	100.0	71.6	60.4	79.9	70.5	54.9	79.5	100.0	80.9
	Feature MA		Prod. Acc.	99.7	73.1	72.1	98.8	70.0	57.3	71.7	99.1	74.8
	OA/Kappa	78.8/0.76	User Acc.	99.7	71.8	77.3	99.1	72.3	58.6	72.1	99.1	97.5
	Prior MA		Prod. Acc.	100.0	98.4	58.1	99.7	95.7	44.1	87.4	79.3	54.1
	OA/Kappa	81.2/0.79	User Acc.	99.7	71.0	64.7	98.8	81.6	53.0	73.5	100.0	100.0
	Bridging MA		Prod. Acc.	100.0	82.1	82.5	97.9	72.4	27.2	86.0	80.9	83.8
	OA/Kappa	79.3/0.77	User Acc.	97.3	80.6	53.4	89.4	81.7	63.3	82.6	97.6	68.6
	k-NN-Target		Prod. Acc.	100.0	96.5	62.7	100.0	96.0	59.6	96.2	99.2	94.8
	OA/Kappa	89.6/0.88	User Acc.	100.0	94.4	61.1	100.0	96.4	61.9	96.8	100.0	93.2
Spe.+Spa.	Prior MA		Prod. Acc.	100.0	99.0	96.4	99.7	98.4	86.1	97.7	50.8	22.7
	OA/Kappa	86.1/0.84	User Acc.	82.6	96.8	98.6	99.4	96.3	95.5	54.7	100.0	96.3
	Bridging MA		Prod. Acc.	100.0	97.1	91.1	89.6	89.5	35.2	98.8	85.6	69.4
	OA/Kappa	79.3/0.77	User Acc.	91.6	91.7	71.5	98.7	86.4	84.4	69.1	96.2	84.6
	k-NN-Target		Prod. Acc.	100.0	98.2	97.3	100.0	99.4	98.0	99.4	99.4	88.4
	OA/Kappa	98.9/0.98	User Acc.	100.0	99.4	98.0	100.0	93.4	97.4	97.8	100.0	98.0

The proposed methods were first compared with four baselines, whose classification results were obtained without using MA or considering domain adaptation. The first baseline result was provided by applying naive k-nearest neighbor (k-NN) classifier, denoted by *naive k-NN*, in the original space. The second baseline performed k-NN classification in a joint manifold feature space developed by LE with pooled data (source and target data sets), which is denoted by *LE-pooled*. The third baseline method was LapSVM, denoted by *LapSVM*, which is often considered to be the benchmark of semi-supervised manifold learning as in [46]. A Gaussian radial basis function kernel was used in LapSVM. The fourth baseline was also created in a semi-supervised way with bridging pairs, denoted by k-NN-Stacked: we stacked the training samples from the source image and the bridged samples from the target image, and performed k-NN classification in the original space. In addition to these four baselines, we also compared the proposed methods to transfer component analysis [47], denoted by *TCA*, a state-of-the-art domain adaptation method, and the feature-level MA method proposed in [19], which is denoted by *Feature MA*. The k-NN classifier was also applied in the latent space learned by these two methods. Note that TCA can be seen as a domain adaptation solution of kernel PCA whose kernels are constructed across domains. Our LE-pooled baseline is similar to [33] that constructs Gram matrices based on pixel-wise similarity (i.e., kernel) across two images.

Samples of the source image were provided for training the classifiers (SVM for LapSVM and k-NN for the rest of the methods). The classifiers were then used to classify samples of the target image, and the classification results were evaluated in terms of overall accuracy, producer/user accuracy, and Kappa statistic using the previously blinded class labels of the target image. Differences in geographical locations and temporal dynamics of May/June and May/July pairs make these two pairs more challenging, particularly since no labeled data of the target image are used in classifications.

As a reference of the upper bound of our proposed methods, we also performed k-NN classification with training data from 50% of the labeled samples of the target data. The classification results of this reference, denoted by *k-NN-Target*, were evaluated with the rest of the labeled data.

Free parameters such as k for k-NN, σ_{spe}, and σ_{spa} were tuned by grid search in the ranges $k = [1, 20]$, $\sigma_{\mathrm{spe}} = [50, 1000]$, and $\sigma_{\mathrm{spa}} = [50, 200]$. The neighborhood size for determining $\mathbf{N}(\mathbf{x}_i)$ was also searched in increments of 5 over the range $[5, 10, \ldots, 60]$. The optimal parameters were determined by fivefold cross validation on the training data set. The free parameters of LapSVM and TCA were tuned as in [46] and [47], respectively.

To quantitatively evaluate the difference between two data sets, we used symmetric Kullback–Leibler (K–L) divergence [48], [49] to measure similarity between $p(\mathbf{X}^s|C_i)$ and $p(\mathbf{X}^t|C_i)$. With this indicator, we can estimate the impact of spectral drift, as illustrated in Fig. 1, between two images for a specific class.

V. RESULTS AND DISCUSSIONS

A. Classification Accuracies

The effectiveness of the proposed methods with spectral information is first evaluated and compared with the baseline methods introduced previously. Then, we compare the MA results obtained with our methods by including the spatial information. The overall accuracies and Kappa statistics obtained by the proposed methods and the baselines with spectral information are listed in Tables III–V. The accuracies achieved by *naive k-NN* are about 5% lower than the proposed MA methods in the

TABLE IV
OVERALL PRODUCER/USER ACCURACIES AND KAPPA STATISTICS OF JUNE/JULY PAIR

June/July		Class ID		1	2	3	4	5	6	7	8	9
Spe.	Naive k-NN		Prod. Acc.	100.0	70.8	56.1	95.7	99.2	52.1	95.9	100.0	97.9
		85.3/0.83	User Acc.	100.0	80.0	60.1	98.9	91.5	56.8	81.2	99.3	98.9
	LE-pooled		Prod. Acc.	100.0	70.8	56.1	65.1	93.1	50.3	95.3	100.0	95.8
		80.0/0.77	User Acc.	74.6	77.3	59.4	100.0	89.7	54.8	82.7	98.7	95.8
	LapSVM		Prod. Acc.	100.0	63.5	78.7	94.1	98.5	42.6	82.5	99.3	93.8
		83.9/0.82	User Acc.	94.9	85.9	67.2	96.7	83.2	77.4	82.0	94.4	68.7
	k-NN-Stacked		Prod. Acc.	100.0	72.9	59.8	98.4	97.7	56.2	95.3	100.0	97.9
		86.5/0.85	User Acc.	100.0	73.7	68.1	98.4	90.8	60.9	83.6	99.3	98.9
	TCA		Prod. Acc.	99.5	65.6	70.1	98.4	93.1	47.3	90.1	100.0	99.0
		85.1/0.83	User Acc.	100.0	81.8	61.2	100.0	87.8	61.5	87.5	98.7	79.8
	Feature MA		Prod. Acc.	100.0	78.1	43.9	98.4	93.1	68.0	96.5	96.7	96.9
		85.7/0.84	User Acc.	100.0	65.2	82.8	100.0	90.4	56.1	83.8	100.0	96.9
	Prior MA		Prod. Acc.	100.0	74.0	53.7	100.0	97.7	54.4	98.8	100.0	95.8
		86.2/0.84	User Acc.	100.0	85.5	60.3	97.9	93.4	56.4	84.1	99.3	100.0
	Bridging MA		Prod. Acc.	100.0	77.1	58.5	100.0	97.7	52.1	95.3	99.3	97.9
		86.3/0.85	User Acc.	99.5	55.6	71.1	97.4	93.4	65.2	87.6	99.3	98.9
	k-NN-Target		Prod. Acc.	100.0	94.9	61.0	100.0	94.4	64.8	90.8	100.0	93.6
		88.2/0.87	User Acc.	100.0	77.1	60.3	98.9	95.8	65.6	96.7	100.0	100.0
Spe.+Spa.	Prior MA		Prod. Acc.	100.0	77.1	96.3	100.0	100.0	95.3	99.4	100.0	97.9
		97.1/0.97	User Acc.	89.4	98.7	99.4	97.9	98.5	99.4	96.6	99.3	98.9
	Bridging MA		Prod. Acc.	100.0	86.5	95.1	100.0	99.2	94.1	100.0	100.0	95.8
		97.3/0.97	User Acc.	93.0	94.3	98.1	98.9	97.7	100.0	95.5	99.3	100.0
	k-NN-Target		Prod. Acc.	100.0	100.0	98.8	100.0	98.5	96.5	98.8	100.0	95.8
		98.8/0.99	User Acc.	99.0	97.9	97.7	100.0	98.5	98.8	98.8	100.0	97.9

TABLE V
OVERALL PRODUCER/USER ACCURACIES AND KAPPA STATISTICS OF MAY/JULY PAIR

May/July		Class ID		1	2	3	4	5	6	7	8	9
Spe.	Naive k-NN		Prod. Acc.	100.0	77.1	54.9	100.0	87.8	35.5	88.9	59.9	81.3
		76.4/0.73	User Acc.	100.0	52.1	60.4	94.9	59.0	52.6	76.0	100.0	100.0
	LE-pooled		Prod. Acc.	100.0	53.1	56.1	61.8	95.4	34.3	93.6	73.7	79.2
		72.2/0.69	User Acc.	100.0	40.5	59.4	80.4	70.2	50.0	80.4	65.1	100.0
	LapSVM		Prod. Acc.	97.3	64.6	64.0	100.0	98.5	39.1	94.7	79.6	52.1
		78.6/0.76	User Acc.	100.0	59.0	57.7	94.4	76.3	53.7	81.8	84.0	96.2
	k-NN-Stacked		Prod. Acc.	100.0	69.8	47.6	100.0	88.5	39.1	95.3	82.9	93.8
		79.8/0.77	User Acc.	100.0	51.9	63.9	80.5	85.3	56.9	85.8	99.2	78.9
	TCA		Prod. Acc.	100.0	75.0	48.2	100.0	70.2	39.6	89.5	66.4	91.7
		75.8/0.73	User Acc.	100.0	49.0	61.7	94.4	60.9	51.1	69.5	99.0	98.9
	Feature MA		Prod. Acc.	100.0	77.1	51.8	99.5	87.0	57.4	82.5	80.3	89.6
		80.7/0.78	UserAcc.	100.0	57.4	72.0	97.9	64.8	59.5	80.1	95.3	100.0
	Prior MA		Prod. Acc.	100.0	75.0	53.7	100.0	96.9	40.8	95.3	90.1	78.1
		81.6/0.79	User Acc.	100.0	48.0	65.2	97.4	79.9	55.6	84.9	98.6	100.0
	Bridging MA		Prod. Acc.	100.0	72.9	72.6	99.5	75.6	43.8	90.1	85.6	86.5
		81.4/0.79	User Acc.	93.0	75.3	61.0	93.4	78.0	65.5	81.1	99.2	79.8
	k-NN-Target		Prod. Acc.	100.0	94.9	61.0	100.0	94.4	64.8	90.8	100.0	93.6
		88.2/0.87	User Acc.	100.0	77.1	60.3	98.9	95.8	65.6	96.7	100.0	100.0
Spe.+Spa.	Prior MA		Prod. Acc.	100.0	76.0	65.9	92.5	100.0	70.4	98.2	99.3	79.2
		87.6/0.86	User Acc.	62.1	93.6	99.1	87.8	92.3	93.0	98.2	99.3	100.0
	Bridging MA		Prod. Acc.	100.0	82.3	97.0	96.8	96.9	69.2	89.5	83.7	92.7
		90.1/0.89	User Acc.	100.0	78.2	89.8	91.8	79.9	98.3	85.5	94.8	89.0
	k-NN-Target		Prod. Acc.	100.0	100.0	98.8	100.0	98.5	96.5	98.8	100.0	95.8
		98.8/0.99	User Acc.	99.0	97.9	97.7	100.0	98.5	98.8	98.8	100.0	97.9

May/June and May/July pairs and had approximately the same classification accuracy for the June/July pair. Compared with the *LE-pooled* baseline, overall accuracies achieved using MA with a prior manifold were $\sim 7\%$, $\sim 6\%$, and $\sim 9\%$ higher for the May/June, June/July, and May/July data pairs, respectively. Compared with *LapSVM*, overall accuracies were $\sim 3\%$, $\sim 2\%$, and 3% higher for these three data pairs. Roughly, the same level of improvement was achieved by MA with bridging pairs.

As compared with k-*NN-Stacked*, that including bridged samples from the target image in the training set, the two proposed methods deliver higher classification results for these three data pairs. Specifically, we also observe that the classification accuracies are higher than those from *naive k-NN* and *LE-pooled*. We can further see that the k-*NN-Stacked* method delivers higher accuracies than *LapSVM* and *TCA*. These observations suggest some potential of exploiting bridging pairs for domain adaption problems.

When compared with *TCA* and *Feature MA*, the two proposed methods also result in higher overall accuracy and Kappa statistic. Note that *TCA* seeks a feature space where distances between kernel means are globally minimized, which differs from our local MA methods where local structures are

TABLE VI
OPTIMAL NEIGHBORHOOD SIZE

	LE-pooled	Prior w/ Spe.	Prior w/ Spe.+Spa.	Bridging w/ Spe.	Bridging w/ Spe. +Spa.
May/June	35	5	5	5	5
June/July	45	5	5	5	5
May/July	55	5	5	10	10

TABLE VII
SYMMETRIC KL-DIVERGENCE OF $p(\mathbf{X}^s|C_i)$ AND $p(\mathbf{X}^t|C_i)$: MAY/JUNE PAIR

i	1	2	3	4	5	6	7	8	9
1	244.2	3772.6	3064.3	2592.8	6438.3	3301.8	7054.5	6512.5	12227.3
2	3772.6	148.3[+]	539.5	325.3	247.1	214.7	232.5[+]	286.5	635.5
3	3064.3	539.5	378.2[*]	1098.9	1898.1	333.5[*]	1837.2	1422.2	6084.6
4	2592.8	325.3	1098.9	158.7	659.8	421.6	502.4	657.7	654.1
5	6438.3	247.1	1898.1	659.8	203.6	472.3	237.7	277.4	378.6
6	3301.8	214.7	333.5[*]	421.6	472.3	137.4[*]	356.6	359.5	989.2
7	7054.5	232.5[+]	1837.2	502.4	237.7	356.6	180.6[+]	250.4	308.4
8	6512.5	286.5	1422.2	657.7	277.4	359.5	250.4	179.7	365.3
9	12227.3	635.5	6084.6	654.1	378.6	989.2	308.4	365.3	236.9

[*] denotes the difficult class pair Class 3/Class 6
[+] denotes the difficult class pair Class 2/Class 7 that is also discussed in Figure 6.

TABLE VIII
SYMMETRIC KL-DIVERGENCE OF $p(\mathbf{X}^s|C_i)$ AND $p(\mathbf{X}^t|C_i)$: JUNE/JULY PAIR

i	1	2	3	4	5	6	7	8	9
1	11410.4	14978.8	11388.1	7694.4	19771.0	14910.3	17977.3	19946.0	31680.3
2	14978.8	78.8	425.9	332.7	122.6	188.8	111.9	190.9	413.1
3	11388.1	425.9	88.3[*]	654.1	846.9	89.6[*]	465.7	428.2	3373.5
4	7694.4	332.7	654.1	63.6	479.1	293.5	283.0	534.0	531.2
5	19771.0	122.6	846.9	479.1	59.8	316.4	119.3	222.2	261.4
6	14910.3	188.8	89.6[*]	293.5	316.4	73.6[*]	145.9	183.7	798.7
7	17977.3	111.9	465.7	283.0	119.3	145.9	46.6	158.6	175.6
8	19946.0	190.9	428.2	534.0	222.2	183.7	158.6	46.7	321.1
9	31680.3	413.1	3373.5	531.2	261.4	798.7	175.6	321.1	63.6

[*] denotes the difficult class pair Class 3/Class 6.

encouraged to aligned. Since spectral drift may occur at local scale in the spectral domain and vary class by class, our local MA methods may be beneficial in such case. *Feature MA* can be seen as a linear version of our bridging pair MA. The linear version could be computationally advantageous but would deliver lower classification accuracies since certain nonlinearity within hyperspectral data might be ignored with linear operations.

The optimal numbers of local neighbors for *LE-pooled* and the proposed methods are listed in Table VI. The *LE-pooled* baseline method required significantly more spectral neighbors than the MA approaches to obtain optimal classification results. However, since spectral drift may be class dependent, the optimal number of neighbors may introduce undesirable connected spectral neighbors, resulting in an unfaithful joint manifold where a large number of omission and commission classification errors are observed for some classes.

Class-dependent accuracies of three data pairs are shown in Tables III–V, respectively. Classes 2, 3, 4, and 6 tend to be spectrally similar. Without MA, the overlaps in spectral signatures result in misclassifications with *LapSVM*. Classes 5, 7, 8, and 9 suffer from the same problem, and the resulting accuracies are also low. These tendencies are more noticeable for the May/June and May/July pairs because of the significant spectral drift and their reduced overlap in the spatial domain.

Increases of $\sim 5\%$, $\sim 11\%$, and 6% in overall accuracies for these three data pairs, respectively, were achieved when spatial proximity was included in the MA with a prior manifold strategy. Similarly, including spatial information for the MA with bridging pairs method results in respective $\sim 6\%$, $\sim 10\%$, and $\sim 9\%$ overall accuracy gains for these data pairs. The increases in accuracy are primarily due to the higher class accuracies of some classes that are difficult to discriminate. By incorporating spatial proximity into the similarity function, the resulting graph leads to a stable and robust manifold representation. In Tables III–V, critical classes such as Classes 3 and 6, as indicated by the small divergence distances in Tables VII–IX, have significantly higher class accuracies, whereas most classes had at least marginally higher accuracies. A labeled subset that contains Classes 6, 7, and 9 of the June data is shown in Fig. 5. With addition spatial information, we observe that Class 6 (yellow) has a more homogeneous classified area in Fig. 5(c) and (e) than in Fig. 5(b) and (d), where only spectral information is exploited. In addition, the number of samples that are misclassified as Class 3 (light blue colored) is reduced.

Scatter plots of the embeddings in the joint manifold feature space are included in Fig. 6, where labeled data in the first two intrinsic dimensions are shown. For the MA scenario using

TABLE IX
SYMMETRIC KL-DIVERGENCE OF $p(\mathbf{X}^s|C_i)$ AND $p(\mathbf{X}^t|C_i)$: MAY/JULY PAIR

i	1	2	3	4	5	6	7	8	9
1	21517.4	12878.8	16061.1	11866.6	21513.0	20136.4	17405.2	16835.3	25301.7
2	12878.8	193.9	306.1	343.5	278.0	252.7	234.1	305.7	672.4
3	16061.1	306.1	133.2*	477.1	545.7	142.6*	363.7	322.7	1128.7
4	11866.6	343.5	477.1	148.5	607.4	389.1	453.5	657.7	629.1
5	21513.0	278.0	545.7	607.4	308.4	443.2	293.1	399.7	587.7
6	20136.4	252.7	142.6*	389.1	443.2	140.3*	266.4	273.2	822.8
7	17405.2	234.1	363.7	453.5	293.1	266.4	186.0+	265.6	383.8+
8	16835.3	305.7	322.7	657.7	399.7	273.2	265.6	173.7	628.6
9	25301.7	672.4	1128.7	629.1	587.7	822.8	383.8+	628.6	368.1+

* denotes the difficult class pair Class 3/Class 6.
+ denotes the difficult class pair Class 7/Class 9 that is also discussed in Figure 6.
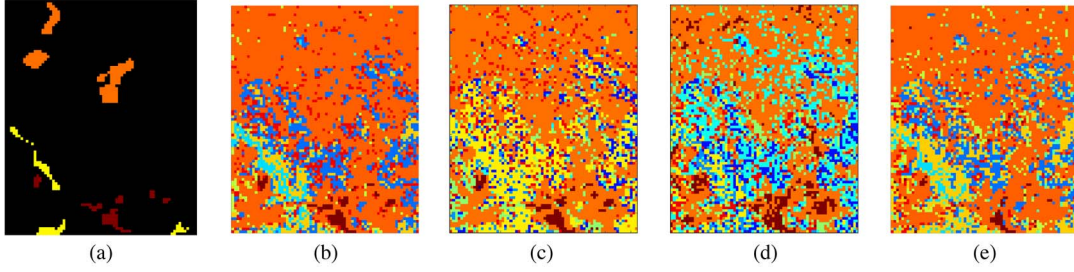
Fig. 5. May/June pair: Classified map of a subset data. (a) Labeled Class (subset data), (b) classified via a prior manifold (spectral), (c) classified via prior manifold (spectral + spatial), (d) classified via bridging pairs (spectral), and (e) classified via bridging pairs (spectral + spatial). The colors correspond to the class list in Table II.
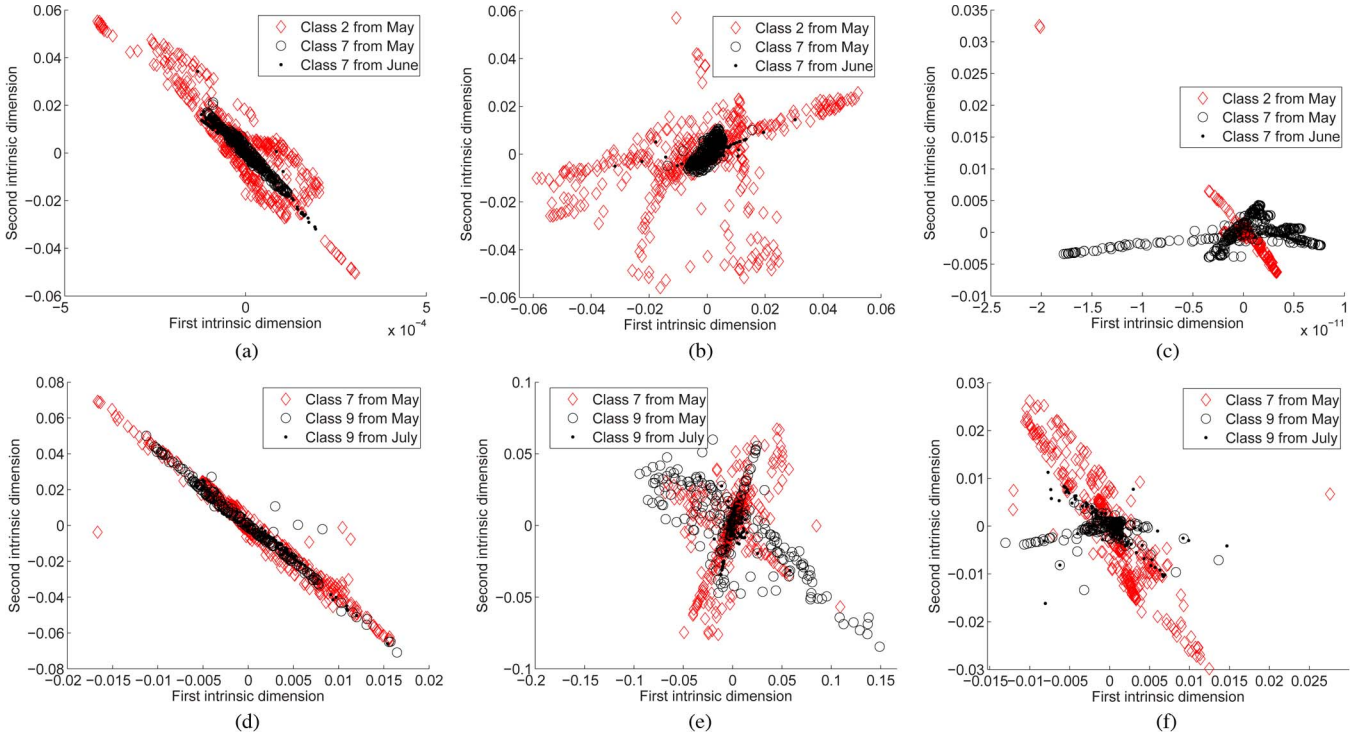

Fig. 6. Scatter plots of resulting joint manifold spaces that were estimated using LE-pooled, MA with prior manifold, and MA with bridging pairs methods. (a) LE-pooled. Accuracy of Class 7: 70.4%. (b) MA with prior manifold (spectral). Accuracy of Class 7: 87.4%. (c) MA with prior manifold (spectral + spatial). Accuracy of Class 7: 97.7%. (d) LE-pooled. Accuracy of Class 9: 79.2%. (e) MA with bridging pairs (spectral). Accuracy of Class 9: 86.5%. (f) MA with bridging pairs (spectral + spatial). Accuracy of Class 9: 92.7%.

features derived from previous images, the Class 2/Class 7 pair of the May/June pair demonstrates why the baseline *LE-pooled* was unable to successfully distinguish some classes.

The enlarged neighborhood search required by *LE-pooled* often includes samples that are spectrally similar but actually do not have the same class label. In this case, small symmetric K–L

TABLE X
OPTIMAL INTRINSIC DIMENSION OF THE DATA PAIRS

	Prior w/ Spe.	Prior w/ Spe.+Spa.	Bridging w/ Spe.	Bridging w/ Spe. +Spa.
May/June	49	41	21	28
June/July	49	48	32	36
May/July	49	41	27	16

divergence between these two classes, as marked in Table VII, indicates that samples of Class 7 from the June data are very similar to Class 2 from the May data. Clustering conditions are imposed by misleading neighborhoods and result in inseparable clusters of these two classes in the learned joint manifold. As shown in Fig. 6(a), there is no clear separation between the two classes. With training data from the May data (the source image), samples of Class 7 from the June data (the target image) are prone to be misclassified as Class 2. By exploiting manifold features derived from previous images, a higher producer accuracy was achieved for Class 7, and Class 2 had a higher user accuracy (see Table III). Fig. 6(b) shows that the cluster associated with Class 7 is more dense. The scatter plot [see Fig. 6(c)] further shows that two classes can be classified correctly using the k-NN classifier by exploiting spatial proximity.

For the bridging pairs MA strategy, Classes 7 and 9 that also have small symmetric K–L distances for the May/July pair, as listed in Table IX, illustrating why MA can result in higher classification accuracies when spectral drift occurs. Analogous to the Class 2/Class 7 example, similar spectral signatures lead to non-informative spectral neighbors with the *LE-pooled* method and subsequently confuse cluster conditions for developing joint manifolds, as shown in Fig. 6(d). The signatures of these two classes are overlapped in the first two intrinsic dimensions. In Fig. 6(e) and (f), the two classes are associated with different clusters; clearer separation between two clusters also occurs with auxiliary spatial information. The corresponding higher class accuracies of Class 9 are also indicated in Fig. 6.

Results of k-*NN-Target* are also listed in Tables III–V. The classification results provide an upper bound of the proposed MA methods, which do not use labeled samples from the target image. Using only spectral features, the differences between our proposed MA methods and the supervised classifications are as small as $\sim 3\%$ (June/July pair) and $\leq 10\%$ for the other two data sets, where there is a greater change between individual images in the pairs. Similarly, the most successful classification results of the two MA frameworks with spectral and spatial information are less than 3% for the June/July pair, as compared with the traditional supervised classification results. For the other two challenging data sets, the classification accuracies for the spectral–spatial MA frameworks are within 13% of the accuracies obtained from the supervised experiments. In these two challenging pairs, MA with bridging pairs for the May/July pair achieved the result closest to the upper bound, with an 8.6% overall accuracy difference.

It should also be noted that the effectiveness of the proposed methods is not restricted to any order of constituent data pairs. For example, we applied the two MA methods to another June/May pair, where June is the source image and May is the
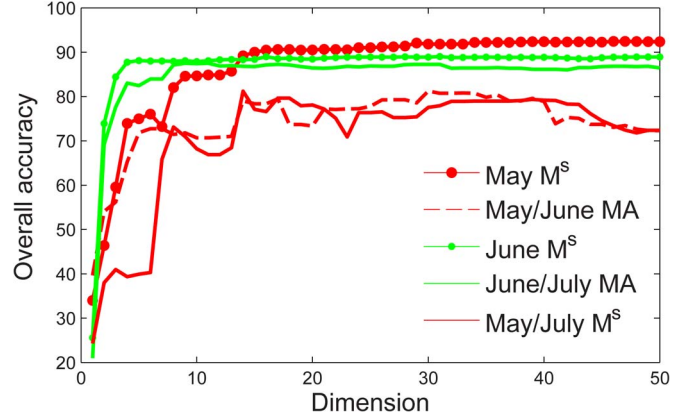

Fig. 7. Sensitivity analysis of $\mathbf{M}^s$ to dimension d.

target image. The overall accuracies provided by the two MA methods are about 63.5% and are better than *naive k-NN* and *TCA*, i.e., 62.8% and 62.4% respectively, and equivalent to the *Feature MA* method, which is 63.0%.

B. Sensitivity Analysis

The optimal dimension $\hat{d}$ for each image pair is listed in Table X. Note that the optimal dimension for MA with a prior manifold depends on the dimension of $\mathbf{M}^s$, as implied in (5). For spectral experiments, the first 49 features of the manifold derived for the May data and the first 50 features of the June data were selected according to the highest accuracy criterion. In the spectral and spatial proximity cases, 41 and 50 features were selected for the source data May and June, respectively. As shown in Table X, the optimal dimensions for MA based on the manifold derived for the prior month are very close to the dimension of $\mathbf{M}^s$. Fig. 7 shows the classification accuracies of $\mathbf{M}^s$ obtained with various d and corresponding $\mathbf{F}^t$ to demonstrate the impact of the dimension of $\mathbf{M}^s$ on the resulting accuracies provided by MA. This indicates that the optimal dimension of a prior manifold could be an important factor for obtaining satisfactory MA results. Since we have labeled samples from $\mathbf{X}^s$, it is easy to exploit the labeled samples to estimate optimal $\hat{d}$ for MA.

The optimal dimensions for MA with bridging pairs, and the accuracies are shown in Table X and Fig. 8, respectively. In most of the cases, the optimal classification accuracies were achieved with fewer than 30 intrinsic features. Although more than 30 features are needed for the June/July pair, the accuracy changes are small after 20 features.

We also investigated the usefulness of including spatial locations for determining the bridging pairs. We varied a in the range $[0, 0.1, \ldots, 1]$, where $a = 0$ has no spatial information is included, whereas $a = 1$ implies that the bridging pairs
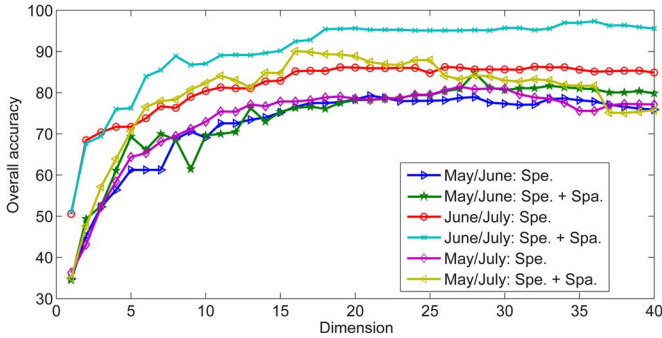
Fig. 8. Overall accuracy versus intrinsic dimension d of $\mathbf{F}$ using bridging pairs.
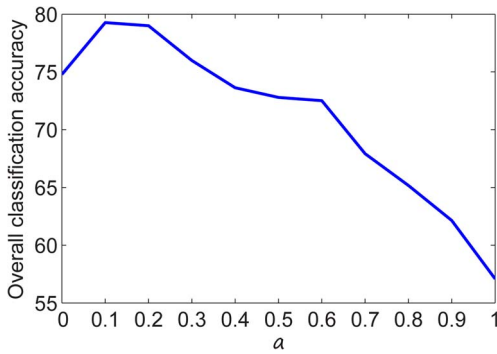

Fig. 9. Sensitivity analysis of bridging pairs to values of a (May/June pair).

are determined heavily depending on the spatial locations. In the latter case, bridging pairs usually do not contain useful information, as spectral proximity has a too low weight. As an example using the May/June pair, the result of MA is not satisfactory, as shown in Fig. 9. We used $a = 0.2$ in all of our experiments based on grid search and empirical results since we have similar observations for these three data pairs as in Fig. 9.

C. Classification With Limited Training Samples

To investigate the effect of small training samples, random samples were selected from the training set with a sampling rate r in increments of 0.1 within $[0, 1]$. Classifications were performed with ten trials of each r. The average overall accuracies and standard deviations of the *LE-pooled* baseline method and the proposed methods are shown in Fig. 10. Considering the upper and lower bounds of overall accuracies, indicated by standard deviations, the accuracies obtained by the proposed MA methods were consistently higher than those from the *LE-pooled* baseline for the multitemporal test data, regardless of how many training samples were provided to the k-NN classifier.

With $r > 0.2$, the two proposed MA methods with spectral and spatial proximity provide consistently higher accuracies, as compared with MA with spectral characteristics. These observations suggest that our MA frameworks can be a useful representation learning [50] approach, where manifold is a powerful assumption for learning representations. Clustering conditions built by samples in an unsupervised mode are well represented in the latent space $\mathbf{F}$ with our MA frameworks.

We also investigated the computation complexity of the proposed methods. The most demanding part of the methods is the inverse step and the eigendecomposition step in the prior manifold and the bridging pairs methods, respectively. Both steps require $O((|\mathbf{X}^s| + |\mathbf{X}^t|)^3)$. Recent developments on large-scale matrix operations [51], [52] could be incorporated when the number of samples is large.

VI. CONCLUSION

This paper has addressed the issue of classification task in the presence of spectral shifts within multitemporal hyperspectral images. The spectral variations were primarily induced by dynamic environments associated with flooding and seasonal vegetation phenology. The issue may be handled from a geometric learning point of view. We propose to exploit limited labeled data and align similar local manifolds of two temporal images in a common latent space for classification. The scheme is referred to as MA.

Based on different scenarios, we investigated two MA methods. The MA method based on a prior manifold developed from a previous image attempts to solve a domain adaptation problem by jointly exploiting current optimal features space, local-based data manifolds of the target image, and the relationships between target and source images. Bridging-pair-based MA finds a joint latent space where data geometries of two data sets are aligned with some bridging pairs. The two MA methods deliver comparable classification results and outperform the four baseline methods in terms of classification accuracies, as well as two state-of-the-art domain adaptation methods.

The promising results demonstrate the potential of utilizing similar local geometries of two temporal images without involving any labeled data from the target image. The proposed MA strategies exploit similar underlying manifolds of two data sets and align them in a common latent space in one step, potentially avoiding the need to redefine classifier rules. The benefits of integrating spatial information for discriminating difficult classes in the joint manifold were also demonstrated. Since few studies have investigated transfer learning in the manifold space in the remote sensing community, our proposed methods establish a framework for aligning data manifolds of hyperspectral images from a domain adaptation perspective.

A possible fruitful direction for extending these results would be to consider the use of submanifolds for classification tasks. Classification performed in manifold domains is based on the high-density assumption that pixels are classified according to their neighbors. At a local scale, the regions with different densities can be viewed as submanifolds, which are combined at global level. This topic was recently discussed in [53]–[55]. Such submanifolds might be learned at the local scale and prove beneficial to our local MA concepts. In the future, we will also further investigate a hybrid global and local manifold learning approach for MA. Local and global-based approaches concentrate on different characteristics of data manifolds. Local approaches are more favored in discriminating difficult classes, whereas global methods are advantageous for data representation over sequences of images. As discussed in [38], changes in global manifolds developed from two sequential temporal
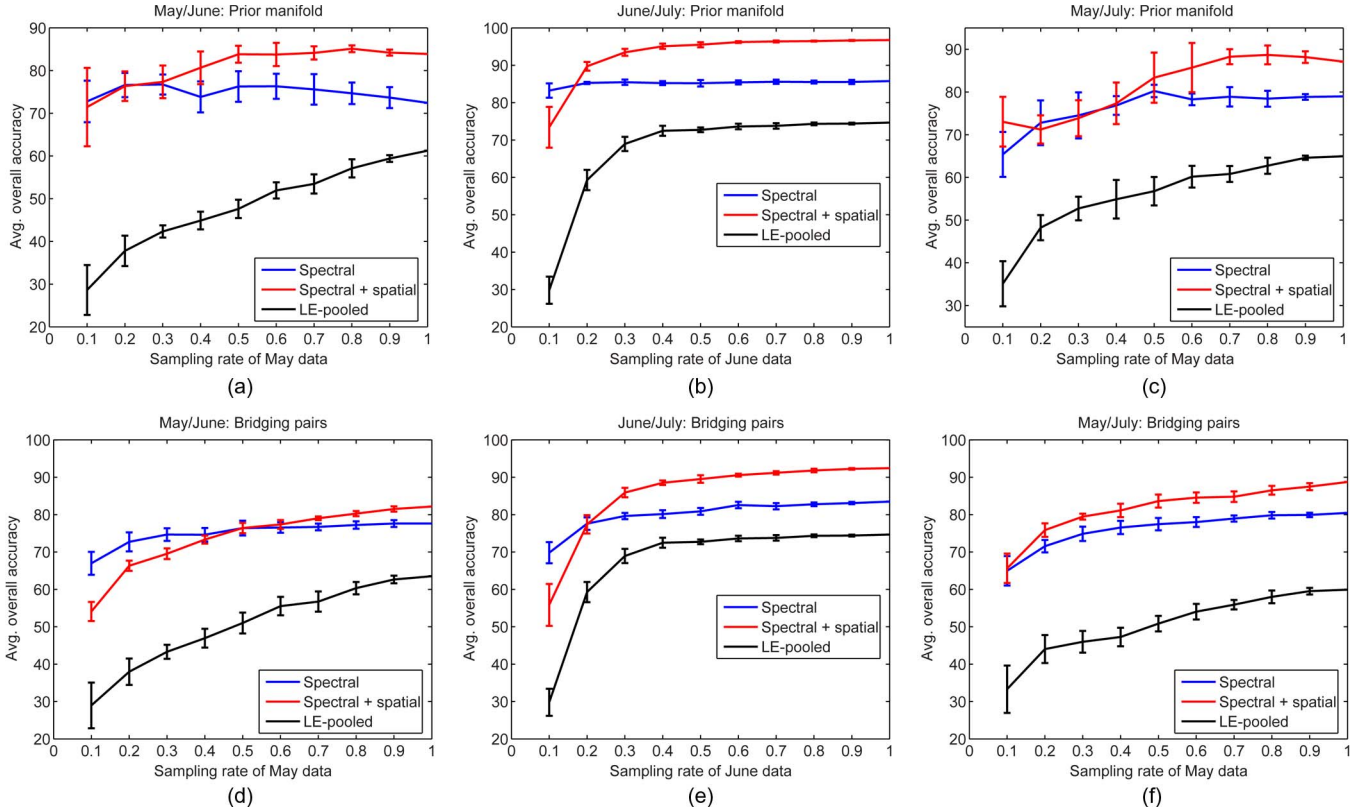
Fig. 10. Classification with varied training sample size.

images are often small for many classes. Combining global structure and similar locality can potentially better exploit the advantages of global and local manifold learning approaches. In addition, based on the classification results, both the k-NN-Stacked method and bridging pairs MA demonstrate the value of exploiting bridging pairs for addressing domain adaptations problems. We also intend to study other methods such as [56] to model correspondences between the source and target hyperspectral images and construct bridging pairs without requiring labeled data from the target image.

ACKNOWLEDGMENT

REFERENCES

[1] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[2] R. E. Bellman, *Adaptive Control Processes*. Princeton, NJ, USA: Princeton Univ. Press, 1961.

[3] L. Bruzzone and M. Marconcini, "Toward the automatic updating of land-cover maps by a domain-adaptation SVM classifier and a circular validation strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1108–1122, Apr. 2009.

[4] W. Kim and M. Crawford, "A novel adaptive classification method for hyperspectral data using manifold regularization kernel machines," in *Proc. 1st Workshop Hyperspectr. Image Signal Process.—Evol. Remote Sens.*, 2009, pp. 2–5.

[5] S. Rajan, J. Ghosh, and M. M. Crawford, "Exploiting class hierarchies for knowledge transfer in hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3408–3417, Nov. 2006.

[6] W. Kim, "Manifold learning for robust classification of hyperspectral data," Ph.D. dissertation, Dept. Civil Eng., Purdue Univ., West Lafayette, IN, USA, 2011.

[7] F. Cozman, I. Cohen, and M. Cirelo, "Semi-supervised learning of mixture models," in *Proc. 20th Int. Conf. Mach. Learn.*, Washington, DC, USA, 2003, pp. 99–106.

[8] Y. Grandvalet and Y. Bengio, "Semi-supervised learning by entropy minimization," in *Proc. NIPS*, Vancouver, BC, Canada, 2005, pp. 529–536.

[9] C. Persello and L. Bruzzone, "Active learning for domain adaptation in the supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4468–4483, Nov. 2012.

[10] S. Sandmeier and E. Middleton, "The potential of hyperspectral bidirectional reflectance distribution function data for grass canopy characterization," *J. Geophys. Res.*, vol. 104, no. D8, pp. 9547–9560, Apr. 1999.

[11] C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters*. San Diego, CA, USA: Academic, 1994.

[12] D. A. Roberts, M. Smith, and J. Adams, "Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data," *Remote Sens. Environ.*, vol. 44, no. 2/3, pp. 255–269, May/Jun. 1993.

[13] L. Ma, M. Crawford, and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, Nov. 2010.

[14] M. M. Crawford, L. Ma, and W. Kim, "Exploring nonlinear manifold learning for classification of hyperspectral data," in *Optical Remote Sensing*, vol. 3, ser. Augmented Vision and Reality, S. Prasad, L. M. Bruce, J. Chanussot, R. I. Hammoud, and L. B. Wolff, Eds. Berlin, Germany: Springer-Verlag, 2011, pp. 207–234.

[15] W. Li and S. Prasad, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

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

[17] J. Ham, D. D. Lee, and L. K. Saul, "Semisupervised alignment of manifolds," in *Proc. Int. Workshop Artif. Intell. Stat.*, 2005, pp. 120–127.

[18] S. Lafon, Y. Keller, and R. R. Coifman, "Data fusion and multicue data matching by diffusion maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1784–1797, Nov. 2006.

[19] C. Wang and S. Mahadevan, "A general framework for manifold alignment," in *Proc. AAAI Fall Symp. Manifold Learn. Appl.*, 2009, pp. 53–58.

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

[21] P. H. Swain, "Bayesian classification in a time-varying environment," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 12, pp. 879–883, Dec. 1978.

[22] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 858–867, Jul. 1997.

[23] L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural–statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1350–1359, May 1999.

[24] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 456–460, Feb. 2001.

[25] L. Bruzzone, R. Cossu, and D. F. Prieto, "Combining parametric and nonparametric classifiers for an unsupervised updating of land-cover maps," in *Proc. 1st Int. Workshop Multiple Classifier Syst.*, 2000, pp. 290–299.

[26] Q. Jackson and D. A. Landgrebe, "An adaptive method for combined covariance estimation and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 5, pp. 1082–1087, May 2002.

[27] J. Muñnoz Marí *et al.*, "Multi-temporal image classification with kernels," in *Kernel Methods for Remote Sensing Data Analysis*, G. Camps-Valls and L. Bruzzone, Eds. Hoboken, NJ, USA: Wiley, 2009.

[28] W. Dai, Q. Yang, G. Xue, and Y. Yu, "Boosting for transfer learning," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 193–200.

[29] L. Duan, I. W. Tsang, and D. Xu, "Domain transfer multiple kernel learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 465–79, Mar. 2012.

[30] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[31] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[32] S. Yan, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[33] J. Ham, D. D. Lee, S. Mika, and B. Schölkopf, "A kernel view of the dimensionality reduction of manifolds," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 47.

[34] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, Mar. 2005.

[35] D. Tuia, M. Trolliet, and M. Volpi, "Multisensor alignment of image manifolds," in *Proc. IEEE IGARSS*, Melbourne, VIC, Australia, 2013, pp. 1246–1249.

[36] C. Wang and S. Mahadevan, "Manifold alignment using Procrustes analysis," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1120–1127.

[37] R. Li and R. Chellappa, "Aligning spatio-temporal signals on a special manifold," in *Proc. 11th ECCV*, 2010, vol. 5, pp. 547–560.

[38] H. L. Yang and M. Crawford, "Manifold alignment for multitemporal hyperspectral image classification," in *Proc. IEEE IGARSS*, 2011, pp. 4332–4335.

[39] D. Tuia, J. Muñoz Marí, L. Gómez-Chova, and J. Malo, "Graph matching for adaptation in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 329–341, Jan. 2012.

[40] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, 2004.

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

[42] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Dec. 2006.

[43] W. Di and M. M. Crawford, "Active learning via multi-view and local proximity co-regularization for hyperspectral image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 618–628, 2011.

[44] W. Kim, C. M. Melba, and S. Lee, "Integrating spatial proximity with manifold learning for hyperspectral data," *Korean J. Remote Sens.*, vol. 26, no. 6, pp. 693–703, 2010.

[45] R. Cuingnet, J. A. Glaunés, M. Chupin, H. Benali, and O. Colliot, "Spatial and anatomical regularization of SVM: A general framework for neuroimaging data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 682–696, Mar. 2013.

[46] W. Kim and M. M. Crawford, "Adaptive classification for hyperspectral image data using manifold regularization kernel machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4110–4121, Nov. 2010.

[47] 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.

[48] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proc. R. Soc. Lond. A, Math. Phys. Sci.*, vol. 186, no. 1007, pp. 453–61, Jan. 1946.

[49] S. Kullback and R. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.

[50] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[51] S. Li, W. Wu, and E. Darve, "A fast algorithm for sparse matrix computations related to inversion," *J. Comput. Phys.*, vol. 242, pp. 915–945, Jun. 2013.

[52] C. G. Baker, U. L. Hetmaniuk, R. B. Lehoucq, and H. K. Thornquist, "Anasazi software for the numerical solution of large-scale eigenvalue problems," *ACM Trans. Math. Softw.*, vol. 36, no. 3, pp. 1–23, Jul. 2009.

[53] G. Jun and J. Ghosh, "Nearest-manifold classification with Gaussian processes," in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 914–917.

[54] A. B. Goldberg, X. Zhu, A. Singh, Z. Xu, and R. Nowak, "Multi-manifold semi-supervised learning," in *Proc. Int. Conf. Artif. Intell. Stat.*, 2009, pp. 169–176.

[55] R. Xiao, Q. Zhao, D. Zhang, and P. Shi, "Facial expression recognition on multiple manifolds," *Pattern Recognit.*, vol. 44, no. 1, pp. 107–116, Jan. 2011.

[56] K. Grauman and T. Darrell, "Approximate correspondences in high dimensions," in *Proc. Adv. NIPS 19*, 2006, pp. 505–512.

Hsiuhan Lexie Yang (S'10–M'15) was born in Taiwan. She received the M.S. degree from National Taiwan University, Taipei, Taiwan, in 2007 and the Ph.D. degree in geomatics in civil engineering from Purdue University, West Lafayette, IN, USA, in 2014.

She is currently a Postdoctoral Researcher at the School of Civil Engineering, Purdue University. Her research interests include unmanned aerial vehicles with imaging platform developments, statistical pattern recognition, geospatial data analysis, machine learning, and its related applications to remote sensing data.

Melba M. Crawford (M'90–SM'05–F'08) received the B.S. and M.S. degrees in civil engineering from the University of Illinois at Urbana-Champaign, IL, USA, and the Ph.D. degree in systems engineering from The Ohio State University, Columbus, OH, USA.

She is the Purdue Professor of Excellence in Earth Observation, the Director of the Laboratory for Applications of Remote Sensing, and the Associate Dean of Engineering for Research, Purdue University, West Lafayette, IN, USA. Previously, she was a faculty member at the University of Texas at Austin, TX, USA, where she founded an interdisciplinary research and applications development program in space-based and airborne remote sensing. She is currently leading a joint initiative between the Purdue Colleges of Agriculture and Engineering in development of advanced sensing technologies and analysis methodology for wheeled and unmanned aerial vehicle platforms. Her research interests focus on development of advanced methods for image analysis, including manifold learning, active learning, classification and unmixing, and applications of these methods to hyperspectral and LIDAR data for land cover characterization, mapping, and monitoring.

Dr. Crawford is a Past President of the IEEE Geoscience and Remote Sensing Society. She was also a member of the NASA Earth System Science and Applications Advisory Committee, the NASA EO-1 Science Validation Team, and the Advisory Committee to the NASA Socioeconomic Applications and Data Center (SEDAC). She was a Jefferson Senior Science Fellow at the U.S. Department of State, where she coordinated Science Sector activities within the U.S. National Commission to UNESCO and served as an Advisor to the U.S. Ambassador to UNESCO.