A Novel Graph-Matching-Based Approach for Domain Adaptation in Classification of Remote Sensing Image Pair

Biplab Banerjee, Francesca Bovolo, *Senior Member, IEEE*, Avik Bhattacharya, *Member, IEEE*, Lorenzo Bruzzone, *Fellow, IEEE*, Subhasis Chaudhuri, *Fellow, IEEE*, and Krishna Mohan Buddhiraju, *Member, IEEE*

***Abstract*—This paper addresses the problem of land-cover classification of remotely sensed image pairs in the context of domain adaptation. The primary assumption of the proposed method is that the training data are available only for one of the images (source domain), whereas for the other image (target domain), no labeled data are available. No assumption is made here on the number and the statistical properties of the land-cover classes that, in turn, may vary from one domain to the other. The only constraint is that at least one land-cover class is shared by the two domains. Under these assumptions, a novel graph theoretic cross-domain cluster mapping algorithm is proposed to detect efficiently the set of land-cover classes which are common to both domains as well as the additional or missing classes in the target domain image. An interdomain graph is introduced, which contains all of the class information of both images, and subsequently, an efficient subgraph-matching algorithm is proposed to highlight the changes between them. The proposed cluster mapping algorithm initially clusters the target domain data into an optimal number of groups given the available source domain training samples. To this end, a method based on information theory and a kernel-based clustering algorithm is proposed. Considering the fact that the spectral signature of land-cover classes may overlap significantly, a postprocessing step is applied to refine the classification map produced by the clustering algorithm. Two multispectral data sets with medium and very high geometrical resolution and one hyperspectral data set are considered to evaluate the robustness of the proposed technique. Two of the data sets consist of multitemporal image pairs, while the remaining one contains images of spatially disjoint geographical areas. The experiments confirm the effectiveness of the proposed framework in different complex scenarios.**

***Index Terms*—Clustering, cross-domain graph, domain adaptation (DA), graph matching.**

I. Introduction

IN SUPERVISED classification, one basic assumption is that the labeled training data are generated from the same probability density function (pdf) of the test samples on which the trained model will be applied. However, in many fields including remote sensing [1], it may be possible that, due to the cost required to acquire the labeled training data, training and test samples are associated with related but not identical distributions. In the recent past, researchers have presented methods to tackle the mismatch between training and test domains, with a vision to build a mechanism that uses the labeled samples from the *source* domain to build the classifier that provides a fairly good performance on the test samples available on the *target* domain. This kind of approaches is usually termed as the domain adaptation (DA) or the transfer learning (TL) [2]–[4]. Learning under DA implies that the unlabeled target patterns are drawn from a domain different from the source domain. A detailed discussion of TL techniques for remote sensing image analysis is available in [5].

The problem of adapting models to previously unseen but relevant data sets is one of the important challenges in building a general prediction model for a pair of remote sensing images. It has gained enormous importance with the advent of some new age satellite systems which are capable of acquiring images of the Earth surface almost on a daily basis. This makes it impossible to collect enough reference samples for each available image because of the high cost and high time required. Thus, supervised land-cover mapping for each available image is an impossible task. One of the effective solutions considered in such circumstances relies on the available ground truth data for one of the images (the source domain image) and to propagate this training information to the rest of the images for which no labeled data are available (the target domain images). However, usually DA or TL models do not rely only on the available source domain information. This is because, even if the target and the source domains are similar, differences in the atmospheric conditions, ground reflectance, etc., may impose some local changes in the pdfs of the corresponding land-cover classes in different images. Moreover, some entirely new land-cover classes may also be present in the target domain, or some of the source domain classes may disappear in the target domain image. In this context, DA techniques are particularly useful in building an automated monitoring system aimed at classifying the land-cover classes in the target domain images considering the information from both domains.

A few DA methods available in the remote sensing literature are primarily based on the assumption that the source and the target domains contain the same set of land-cover classes. This implies that only the statistical parameters of the land-cover classes may vary between the acquisitions of multiple images but

Manuscript received August 4, 2014; revised October 29, 2014 and January 2, 2015; accepted January 2, 2015.

B. Banerjee, A. Bhattacharya, and K. M. Buddhiraju are with the Centre of Studies in Resources Engineering, Indian Institute of Technology, Bombay, Mumbai 400076, India.

F. Bovolo is with the Remote Sensing for Digital Earth, Fondazione Bruno Kessler, 38122 Trento, Italy.

L. Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38050 Trento, Italy.

S. Chaudhuri is with the Department of Electrical Engineering, Indian Institute of Technology, Bombay, Mumbai 400076, India.

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

Digital Object Identifier 10.1109/TGRS.2015.2389520

0196-2892 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images

Graph matching for adaptation in remote sensing

not their number and kind. Under such an assumption, the authors in [6] proposed an unsupervised retraining technique for a partially supervised maximum likelihood (ML) classifier for the land-cover mapping in the target domain image given the source domain training data. The method allows the classifier parameters, obtained by exploring the source domain training samples, to be updated in an unsupervised fashion using the expectation maximization (EM) technique on the basis of the class statistical distributions of the target domain image. In [7], an adaptation technique is proposed, which is aimed at finding the correspondence between the data manifolds of both domains. A simple and scalable solution has been presented there by focusing on the description of the changes in the manifold by defining a nonlinear transformation based on vector quantization and graph matching. In the recent past, several active-learning-based classification techniques have been proposed for DA in the remote sensing literature for both multispectral and hyperspectral data [8]–[10]. Some adaptation methods specific to the hyperspectral data have been introduced in [11] and [12]. Large-margin classifiers like support vector machines (SVM) have been modified for the cross-domain data classification purpose [5]. A DA support vector machine (DASVM) technique has been introduced in [5]. The DASVM model is initially built based on the source domain training data. The maximum margin hyperplane of the SVM is then modified iteratively considering only those target domain samples which are very close to the hyperplane, or another convergence criterion is satisfied. The process continues until no target domain samples reside within the SVM margin. In [13], the constraint on the fixed number of land-cover classes is relaxed in the context of DA for multitemporal images. A change vector analysis based method has been used to identify the classwise set of changed/unchanged pixels from the images of both domains. The changed pixels are further analyzed to investigate the existence of some new target domain classes based on statistical divergence measures.

一致性

基于向量量化和图匹配的非线性变换

However, to the authors knowledge, very little endeavor has been put up to develop DA approaches that deal with the problem of identifying the presence of new or disappeared classes when adaptation is required outside the context of multitemporal images. When adaptation is to be applied on images acquired over spatially disjoint locations, temporal correlation cannot be employed for the detection of new/disappeared classes in the target domain. In addition, the same land-cover class may show severe differences in statistical properties when observed in source and target domain images acquired over spatially disjoint areas.

努力

In view of the aforementioned discussion, an efficient and robust DA algorithm is needed, which is able to perform the following: 1) be effective in the presence of significant differences in the statistical and spatial properties of the underlying land-cover classes of both images and 2) detect new/disappeared classes even when temporal correlation is not considered.

为了解决上述问题，我们可以看到，尽管从一个域到另外一个，类统计发生了改变

公共地表类别的相对拓扑结构仍然在图片中保持

在这种情况下，召回图是一个在光谱域中有效且能够建立得很好的拓扑工具

已经广泛使用用图表示图像上的地表类别

To deal with the aforementioned issues, let us observe that, even if the class statistics change from one domain to the other, the relative topological structures of the common land-cover classes are preserved in the images. In this context, let us recall that graphs are an effective and well-established topological tool that has been used extensively to represent the images as a graph of the underlying land-cover classes in the spectral domain [14]. Accordingly, the identification of common cross-domain classes can now be seen as the problem of finding out the maximum common subgraphs (MCSs) [15] given the source and the target domain graphs, followed by identifying the similar matching nodes from both graphs. The remaining nodes (classes) of the individual graphs (not a part of the MCS) represent the added/deleted target domain land-covers.

According to the aforementioned assertion, in this paper, an unsupervised DA technique for land-cover classification of a pair of remotely sensed satellite images is proposed. The method considers that any kind of changes can occur between the source and the target domain images. The proposed method initially clusters the available target domain pixels optimally and simultaneously estimates the number of land-cover classes present in the target domain image. The well-known Kullback–Leibler (KL) divergence is used along with a kernel-based clustering technique for this purpose. This step is followed by an interdomain cluster correspondence step which is the main novel contribution of the proposed DA algorithm. Here, the common set of the source and the target domain classes are mapped into pairs. The new or disappeared target domain classes with respect to the source domain classes are also highlighted. The proposed algorithm is able to handle effectively the data misclassification problem due to the presence of overlapping samples from different land-cover classes. This is achieved by refining the classification results produced by the clustering algorithm according to a partially supervised classifier. Here, an ML classifier combined with an EM-based retraining scheme is employed since it demonstrated to be robust when the statistical distribution of the data points is known or assumed (however, any other classifier working under the assumption that Ω_S and Ω_T might include different sets of classes can be adopted) [6]. The initialization of the class statistical parameters required for the refinement step is performed by blending data from both domains. The adaptation of the statistical parameters of the classes common to both domains takes place in this step. The overall system provides a general unsupervised yet simple framework for cross-domain data classification that can be used in remote sensing data analysis as well as other kind of images.

先尽可能地对目标域像素进行聚类，同时估计在目标域图像中出现的地表类别数

出于此目的，使用的是基于KL散度的核聚类技术

紧接着是域间聚类一致性步骤，这是提出的DA算法的最主要的新贡献。

至此，源域和目标域的公共类别集合映射成对

对于源域类别来说新的或者没有的目标域域类别也凸显出来

提出的算法能够有效地解决数据的误分类问题，由于来自不同地表类别存在重叠样本

根据局部监督分类器的聚类算法产生的精化分类结果

至此，使用一种结合基于最大期望的再训练计划的最大似然分类器

因为它表明当数据的统计分布知道或者假设的时候具有鲁棒性

精化步骤需要的初始类统计参数由两个域的混合数据得到

两个域共同的类统计参数的调整也是在这步

总的系统提供了一种一般化的非监督且简单的框架，用于跨域数据分类，

能够用于遥感图像分析以及其他类型的图像

The rest of this paper is organized as follows. Section II describes the problem of the land-cover classification of a pair of remote sensing images in the context of DA. The proposed DA algorithm is detailed in Section III. The experimental results are presented in Section IV. Section V concludes this paper and discusses relevant possible future research directions.

II. Insight to the Classification Problem of a Pair of Images on the Context of DA

对域适应背景下图像对的分类问题的见解

Let I_S and I_T denote the source and the target domain multispectral remotely sensed images consisting of B spectral bands each. Without loss of validity, I_S and I_T may be acquired from two entirely disjoint locations or represent a multitemporal pair of images acquired over the same geographical area at different times. I_S and I_T may show an identical set of land-cover classes or may not. Moreover, common classes may show significant differences in statistical properties. Let $\Omega_S = \{\omega_1, \omega_2,$

$\ldots, \omega_N\}$ represent the set of land-cover classes that characterize the geographical area represented by I_S. N is the number of land-cover classes modeled in the training set $\text{TR}_S = \{(x_{Sl}, \omega_i)\}$, $(x_{Sl} \in \mathbb{R}^B, 1 \leq i \leq N)$, where x_{Sl} is the lth source domain image pixel.

For I_T where no training data are available, DA techniques mainly assume that the set of land-cover classes Ω_T that characterizes I_T is the same as Ω_S, i.e., they assume a high correlation between the target and source domain sets of classes. Despite this assumption, often the classifier trained on TR_S does not exhibit a good generalization performance on I_T. This is because the estimated classwise parameters of I_S do not provide an accurate estimate of the similar terms in I_T due to the reasons raised in the introduction. However, the classifier trained on I_S represents a reasonable rough estimate of the optimal classifier for I_T. The goal of the DA procedure is to adapt the classifier trained on I_S to the properties of I_T. 领域适应的目的是调整源域图像上的分类器用于目标域的内容

In real applications, there is another element that contributes to the possible poor correlation between Ω_S and Ω_T. This is the possibility that Ω_S and Ω_T may include different sets of classes. The following cases can be identified in this regard.

1) **CASE A**. The source and the target domain contain identical sets of land-cover classes ($\Omega_S = \Omega_T$). 类别一致
2) **CASE B**. There are new classes in the target domain, i.e., $\Omega_T = \Omega_S \cup \{\omega_u\}$, where $\{\omega_u\}$ is the set of new unknown classes detected in I_T. 目标域新增了类别
3) **CASE C**. There are less target domain classes than source domain classes, i.e., $\Omega_T = \Omega_S \backslash \{\omega_k\}$. 目标域减少了类别
4) **CASE D**. k source domain classes disappear in the target domain, and l new classes appear in that place, $\Omega_T = \Omega_S \cup \{\omega_l\} \backslash \{\omega_k\}$, with l either equal to or different from k. 目标域类别有增有减，增减数量要么相等要么不等

The proposed adaptation method (see Fig. 1) handles both the differences in terms of classes (new and disappeared ones) and those in terms of the statistical properties of the common classes. The adaptation procedure is based on the EM algorithm and is able to handle the aforementioned cases. If the classes common to both images can be identified properly, the adaptation can be iteratively performed by EM as in [6]. A sophisticated method is employed for analyzing the remaining target domain classes, for which the corresponding land-cover labels are unknown. 复杂的，精致的

The set of variables used that play important roles in formulating the proposed algorithm is listed in Table I.

III. Proposed Technique

In order to cope with the problem exposed in Section II, a three-step unsupervised DA approach is proposed based on the following. 三步非监督领域适应方法

1) **Generation of an optimal clustering of I_T**: Given $|\Omega_S| = N$ source domain land-cover classes and assuming that the target domain image may contain some additional/disappeared classes as compared to N, the target domain image is clustered iteratively in a range $[N - n, N + n]$, where $n \in \mathbb{N}$ is the maximum difference in the number of classes between Ω_S and Ω_T. Any clustering technique can be used for this purpose. Considering the problem of possible class overlapping, a kernel-based clustering technique is considered here [16]. A KL divergence [17] based optimization mechanism is proposed to select the optimal number of land-cover classes $M \in [N - n, N + n]$ for I_T. Clustering can be performed by any technique. 目标域最佳聚类的生成 提出基于KL散度的优化机制用于地表最佳类别数选择

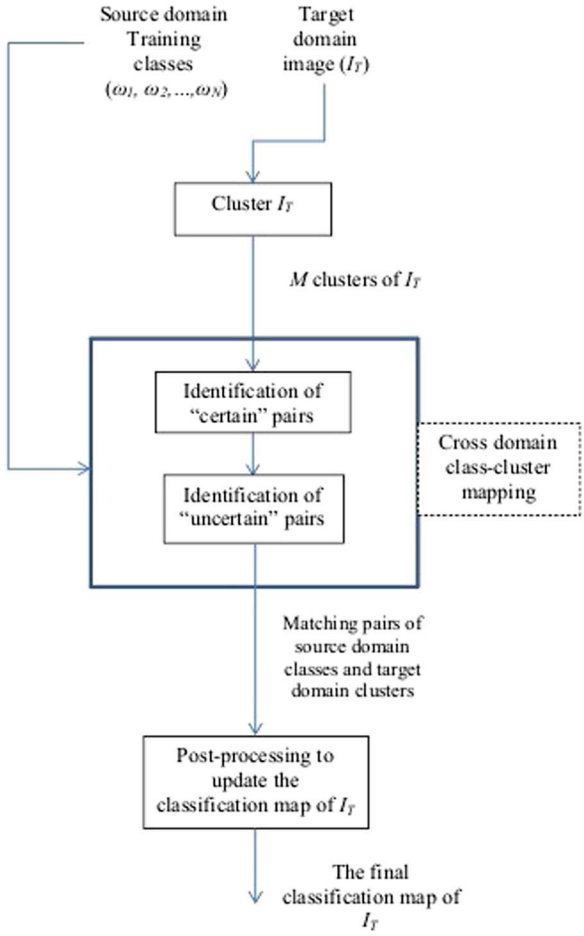


Fig. 1. Flowchart of the proposed method.

2) **Cross-domain cluster mapping**: This is an important step and the main contribution of the proposed algorithm where all of the matching pairs of the source domain land-cover classes and the target domain clusters are found out by using the concepts of subgraph matching [18]. 跨域集群映射 子图匹配
3) **Definition of the DA classifier for postprocessing**: Under the specific hypothesis formulated in the aforementioned two steps, this step performs the adaptation of the classifier built primarily for I_S to the properties of I_T. Here, an ML classifier is used and further retrained by the iterative EM algorithm. Any other classifier working under the assumption that Ω_S and Ω_T might include different sets of classes can be adopted. A set of reliable samples from each target domain cluster is also identified in this step. This step can also be considered as the postprocessing step which refines the land-cover map produced by the clustering technique. 阐述 采用、接受 这步也辨认出每个目标域簇中的可靠样本 这步也被当做后处理步骤，精化由聚类得到的地表图

The first two steps aim at deducing the hypothesis on the classes present in Ω_T and their equivalence with the ones in 推导

TABLE I
LIST OF THE IMPORTANT VARIABLES USED IN THE PROPOSED ALGORITHM

Variable name	Significance
I_S	The source image.
I_T	The target image.
Ω_S	The set of source domain land-cover classes.
Ω_T	The set of target domain land-cover classes.
$\widehat{\Omega_T}$	The set of target domain clusters (estimated land-cover classes).
Q	A measure of cluster compactness.
TR_S	The set of source domain training samples.
ω_i	i^{th} source domain land-cover class.
α_i	The i^{th} target domain cluster.
N	The number of land-cover classes present in the source domain image.
M	The optimal number of clusters in I_T.
Rank_S, Rank_T	Lists for storing the closest source (target) domain classes(clusters) given clusters (classes) from the opposite domain.
R_T, R_S	Temporary lists for storing the source domain classes and the target domain clusters respectively.
X_S, X_T	Random variables representing the observations of source and target domains respectively.
$P_S(\omega_i)$, $P_T(\alpha_j)$	Class prior probabilities in the source and target domains respectively.
$p_S(X_S\|\omega_i)$, $p_T(X_T\|\alpha_j)$	Class conditional probabilities in both the domains.
i, j, k, l	They are used to denote the temporary variables.

这两步共同的目标是检查源域类别和目标域类别是否相同

Ω_S. The cumulative goal of these two steps is to check whether $\Omega_S = \Omega_T$ or not. Accordingly, the class statistical parameters for I_T are initialized and updated in the last step.

相应的，最后一步是目标域图像类统计参数的初始化和更新

A. Definition of an Optimal Number of Clusters (M) for the Target Domain

The goal of this step is to estimate the number (M) of land-covers present in I_T by exploiting the information in the training set of the source domain (TR_S).

这一步的目的是估计目标域中地表类别数，通过利用源域中训练集的信息。

Given N source domain classes, the target domain data are clustered in the range $[N-n, N+n]$ iteratively. n is usually a small integer since it is unlikely that the selected source and target domains involved in the adaptation process are completely different from each other. The final value of M is estimated by means of a cluster validity measure (F) involving relative entropy measure between the domains using KL divergence and the compactness of the newly generated target domain clusters.

n通常是一个小的整数，因为在适应过程中涉及的源域和目标域不太可能完全不同

聚类正确性评价

域间KL散度的相对熵方法，和新生成的目标域簇的紧致度

Here, kernel k-means [19], [20] is used for the clustering purpose. Kernel k-means is preferred to other available methods as it has demonstrated to be effective in handling nonlinearly separable data. This is achieved by first projecting the data into some unknown higher dimensional kernel-induced feature space and performing linear class separation in the kernel space. This allows us to deal with possible class overlapping in the feature space, where standard linear clustering techniques fail. Some fuzzy clustering techniques [21] can also be used with proper parameter setting. However, if it is already known that the samples from different classes are not heavily overlapped, less complex clustering algorithms may be used as well. Radial basis function kernel is used along with kernel k-means. The kernel hyperparameter (γ) is selected by a grid search technique.

径向基函数

In the current scenario, let $p_S(X_S|\omega_i)$ and $p_T(X_T|\alpha_j)$ represent the discrete marginal distributions of the ith source domain class and the jth target domain cluster, respectively. Both distributions are usually approximated by histogram-based approaches [22]. Thus, F is defined for a given $k \in [N-n, N+n]$ as in (1), shown at the bottom of the page, where $KL(p_S(X_S|\omega_i)p_T(X_T|\alpha_j))$ is the nonsymmetric KL divergence of the discrete pdfs $p_S(X_S|\omega_i)$ from $p_T(X_T|\alpha_j)$ and $Q(\widehat{\Omega}_{\mathrm{Temp}}(k))$ is a measure of the quality of the clustering results.

The KL divergence is computed as

$$KL\left(p_S(X_S|\omega_i), p_T(X_T|\alpha_j)\right) = \sum_{x_S \in X_S, x_T \in X_T} \ln \frac{p_S(x_S|\omega_i)}{p_T(x_T|\alpha_j)} p_S(x_S|\omega_i). \quad (2)$$

$KL(p_T(X_T|\alpha_j), p_S(X_S|\omega_i))$ represents the divergence of the discrete pdf $p_T(X_T|\alpha_j)$ from $p_S(X_S|\omega_i)$ and can be obtained by reversing the roles of $p_T(X_T|\alpha_j)$ and $p_S(X_S|\omega_i)$ in (2). $KL(p_S(X_S|\omega_i), p_T(X_T|\alpha_j))$ is a measure of the amount of the information loss when $p_T(X_T|\alpha_j)$ is used to approximate

$$F(k) = \frac{1}{\frac{1}{kN}\left|\sum_{i=0}^{N}\sum_{j=0}^{k} KL\left(p_S(X_S|\omega_i), p_T(X_T|\alpha_j)\right) - \sum_{j=0}^{k}\sum_{i=0}^{N} KL\left(p_T(X_T|\alpha_j), p_S(X_S|\omega_i)\right)\right| + 1} + Q\left(\widehat{\Omega}_{\mathrm{Temp}}(k)\right) \quad (1)$$

簇间方差和簇内方差倒数之和

$p_S(X_S|\omega_i)$. Thus, it is a measure of the differential entropy or the information shared between $p_S(X_S|\omega_i)$ and $p_T(X_T|\alpha_j)$. $KL(p_S(X_S|\omega_i), p_T(X_T|\alpha_j)) \neq KL(p_T(X_T|\alpha_j), p_S(X_S|\omega_i))$ if $p_S(X_S|\omega_i)$ and $p_T(X_T|\alpha_j)$ are not identical and in the optimal case, i.e., $(p_S(X_S|\omega_i) = p_T(X_T|\alpha_j))$, $KL(p_S(X_S|\omega_i), p_T(X_T|\alpha_j)) = KL(p_T(X_T|\alpha_j), p_S(X_S|\omega_i)) = 0$. It signifies that a small KL divergence indicates better similarity between $p_S(X_S|\omega_i)$ and $p_T(X_T|\alpha_j)$. $\sum_i \sum_j KL(p_S(X_S|\omega_i), p_T(X_T|\alpha_j))$ represents the sum of the KL divergence from each of the source domain classes to each of the target domain clusters. $\sum_j \sum_i KL(p_T(X_T|\alpha_j), p_S(X_S|\omega_i))$ represents the same but in the reverse direction. This information theoretic aspect of KL divergence is the prime driving force behind selecting it over the symmetric divergence measures. The asymmetric property of KL divergence allows us to evaluate the amount of information that a source domain class shares with the target domain clusters and vice versa independently. Thus, it also allows a way to handle the cases of highly overlapping data where the clusters are detected only partially by the clustering algorithms. The cluster validity measure F takes into account the contribution of both asymmetric divergences. In the ideal situation when the target domain clusters are highly similar to the source domain training classes, $\sum_i \sum_j KL(p_S(X_S|\omega_i), p_T(X_T|\alpha_j)) - \sum_j \sum_i KL(p_T(X_T|\alpha_j), p_S(X_S|\omega_i)) \to 0$. F is maximized when the target domain clustering result matches the best with the structure of TR_S. In that case, the difference between the entropy of $p_T(X_T|\alpha_j)$ from $p_S(X_S|\omega_i)$ and vice versa will attain the minimum value indicating higher similarity between them. This is important as the aim here is to assess the similarity among the classes and clusters in terms of the information content shared between them.

KL散度的不对称性使我们能够评价源域类别和目标域簇的信息共享量

反之亦然

处理利用聚类算法只能得到高重叠数据的部分簇的情况

在这种情况下，T对S的熵差将得到最小值，反之亦然，表明更高的相似度

The $Q(\widehat{\Omega}_{\mathrm{Temp}}(k))$ term measures the quality of the clustering results in terms of the sum of intercluster variance and the inverse of intracluster variance. Thus, F is a trade-off between the degree of similarity between the domains and the goodness of the target domain clustering. The optimal number of clusters M is obtained when the information shared between the source domain classes and target domain clusters and the cluster compactness for the target domain clusters are maximized over k.

簇间方差和簇内方差倒数之和

因此，F是域间相似度和目标域聚类好坏的一种权衡

最佳的聚类数量M是在源域类和目标域簇之间的信息共享量和目标域聚类紧致度最大时的k

The output of this step is the M clusters $\widehat{\Omega}_T = \{\alpha_1, \alpha_2, \ldots, \alpha_M\}$ of I_T. The cluster map obtained in this step gives a rough estimate of the land-cover map of I_T (see Algorithm 1).

Algorithm 1

Input: Ω_S, TR_S, I_T, $k \in [N-n, N+n]$

Output: The optimal cluster assignments $\widehat{\Omega}_T$ for the pixels of I_T

1: **for** $k = N - n$ to $N + n$ **do**
2: Cluster the pixels of I_T into k number of clusters using kernel k-means. Let $\Omega_{\mathrm{Temp}}(k) = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ be the clusters obtained.
3: $\mathrm{Mean}_{S\to T}$ = the mean KL divergence from all of the classes in TR_S to the clusters in Ω_{Temp}.
4: $\mathrm{Mean}_{T\to S}$ = the mean KL divergence from all of the clusters in Ω_{Temp} to the classes in TR_S.
5: $F(k) = (1/(|\mathrm{Mean}_{S\to T} - \mathrm{Mean}_{T\to S}| + 1)) + Q(\widehat{\Omega}_{\mathrm{Temp}}(k))$ (Alternatively Equation (1))
6: **end for**
7: $M = \underset{k\in[N-n,N+n]}{\operatorname{argmax}} F(k)$
8: $\widehat{\Omega}_T = \Omega_{\mathrm{Temp}}(M) = \{\alpha_1, \alpha_2, \ldots, \alpha_M\}$

B. Cross-Domain Cluster Mapping

跨域群集映射

This is the most important step and the main contribution of the proposed DA technique where the unique one-to-one mapping between all of the matching source domain classes and the target domain clusters is performed and the new land-covers in the target domain are identified. Four possible cases to model the changes between I_S and I_T are mentioned in Section II. In order to handle a special situation of CASE D, where a set of k land-cover classes of I_S is replaced by k new classes in I_T, a postprocessing check is included in the cluster mapping algorithm to identify all of the nonequivalent pairs of the source domain classes and the target domain clusters.

A graph theoretic approach is proposed here for the cross-domain cluster mapping purpose. A graph is a set composed of nodes connected by edges. The edges are weighted, and the weight of a given edge connecting two nodes is a measure of similarity between the adjoining nodes. Graphs have been used extensively in the domain of TL [23]. Cross-domain graphs are a special kind of undirected bipartite graph topology where the nodes are from different domains and edges only exist between the nodes from one domain to the other. The edge weights in this case measure the degree of similarity between the inter-domain nodes. The cross-domain graph $(G(V, E))$ topology is used here to establish the correspondence between classes and clusters of Ω_S and $\widehat{\Omega}_T$, respectively. $\{\omega_i\}_{i=1}^N$ and $\{\alpha_j\}_{j=1}^M$ represent the source and the target domain nodes of G, i.e., $V = \{\{\omega_i\}_{i=1}^N \cup \{\alpha_j\}_{j=1}^M\}$. There are $(N \times M)$ edges present in G which connect all of the ω_i's to the α_j's and vice versa. Fig. 2 depicts a cross-domain graph with 3 source domain classes, 4 target domain clusters, and 12 edges.

出于跨域聚类映射的目的，提出了一种图理论方法

跨域图是一种特殊的无向二分图拓扑结构

节点是来自不同的域，而边只存在于一个域的节点到另一个域的节点之间

这种情况下的边权重衡量域间节点相似度

跨域拓扑图是用来建立类和簇之间的对应

In this respect, two kinds of (ω_i, α_j) pairs are defined.

1) A (ω_i, α_j) pair is called a *certain* pair if α_j is the unique best matched target domain cluster for ω_i, i.e., there is an injective mapping between ω_i and α_j. The first part of the cluster mapping algorithm identifies the presence of these pairs in G.
2) A (ω_i, α_j) pair is called an *uncertain* pair if there is no injective correspondence between ω_i and α_j. A new round of analysis is needed to handle the uncertain cases. It defines two intradomain graphs first and then tries to find their MCS. A subtractive iterative algorithm is followed to find the matching pairs from the MCS.

确定的配对，如果j类是目标域簇i独一无二的最佳匹配

也就是说在i和j之间存在映射图

映射图算法的第一部分确定了在G中这些对的存在

新一轮的分析用来解决不确定的情况

它先定义了两种域间图，然后试图找出他们之间的最大共同子图

Maximum Common Subgraph

接着就是一种删减迭代算法用于从MCS中找出匹配的对

The edge weights (E_W) of G are defined using a symmetric divergence measure depending on the KL divergence, i.e., given in the equation shown at the bottom of the page, where $X_{\omega_i \cup \alpha_j}$ is the observations related to the samples of $\omega_i \cup \alpha_j$. $|E_W|^2$ is called the Jensen–Shannon (JS) divergence [24]. As the square root of the JS divergence is a metric, it is selected for defining the statistical distance between a given ω_i and a given α_j. A

因为JS散度的平方根是一种距离，选择他用了定义i和j之间的统计距离

JS散度，对称且有界(对数底为2时：[0,1])

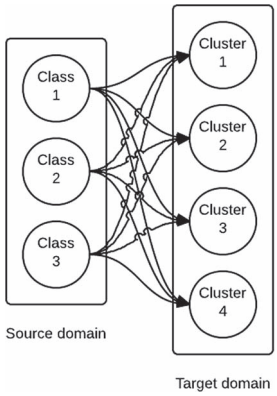


Fig. 2. Example of a cross-domain graph.

small $E_W(\alpha_j, \omega_i)$ indicates high similarity between ω_i and α_j. G is represented efficiently by a symmetric weighted adjacency matrix A_G. Each row of A_G is considered as the array of edge weights from a given source domain class to each of the target domain clusters. The columns of A_G are the arrays of edge weights from a given target cluster to each source domain class.

Two lists Rank_S and Rank_T are maintained in all of the iterations of the cluster mapping algorithm. Rank_S is an $N \times 1$ matrix which defines the closest target domain nodes for each of the N source domain nodes based on the E_W measures, whereas Rank_T is an $M \times 1$ matrix which denotes the same as that of Rank_S but from the target to the source domain.

All of the corresponding pairs of source classes and target clusters are identified in two steps. The first step focuses on the mapping of the *certain* pairs of classes. The second step analyzes the more uncertain correspondences and identifies new target domain classes. A flowchart of the cluster mapping algorithm is depicted in Fig. 3. These steps are detailed as follows.

The **identification of the certain pairs of source domain classes and the target domain clusters from** G requires us to calculate the unique one-to-one mapping between the source domain classes and the target domain clusters from G. Two functions are defined to this end based on Rank_S and Rank_T

$$\nabla_{S\mapsto T}(\omega_i) = \underset{\alpha_{j'}\in\{\alpha_1,\alpha_2,\ldots,\alpha_M\}}{\text{argmin}} E_W(\omega_i, \alpha_{j'}) = \alpha_j \quad (3)$$

$$\nabla_{T\mapsto S}(\alpha_k) = \underset{\omega_{l'}\in\{\omega_1,\omega_2,\ldots,\omega_N\}}{\text{argmin}} E_W(\omega_{l'}, \alpha_k) = \omega_l. \quad (4)$$

$\nabla_{S\mapsto T}$ is a mapping function from a source class to a target cluster, and according to (3), α'_j is the best matched target domain cluster of the source domain class ω_i. $\nabla_{T\mapsto S}$ is a mapping function from a target cluster to a source class, and according to (4), the target domain cluster α_k maps to ω'_l.

If $\nabla_{S\mapsto T}(\omega_i) = \alpha_j$ and $\nabla_{T\mapsto S}(\alpha_j) = \omega_i$ and $|\nabla_{S\mapsto T}(\omega_i)| = |\nabla_{T\mapsto S}(\alpha_j)| = 1$ (where $|\cdot|$ denotes the cardinality of a set), then it is expected that ω_i and α_j represent the same class in both

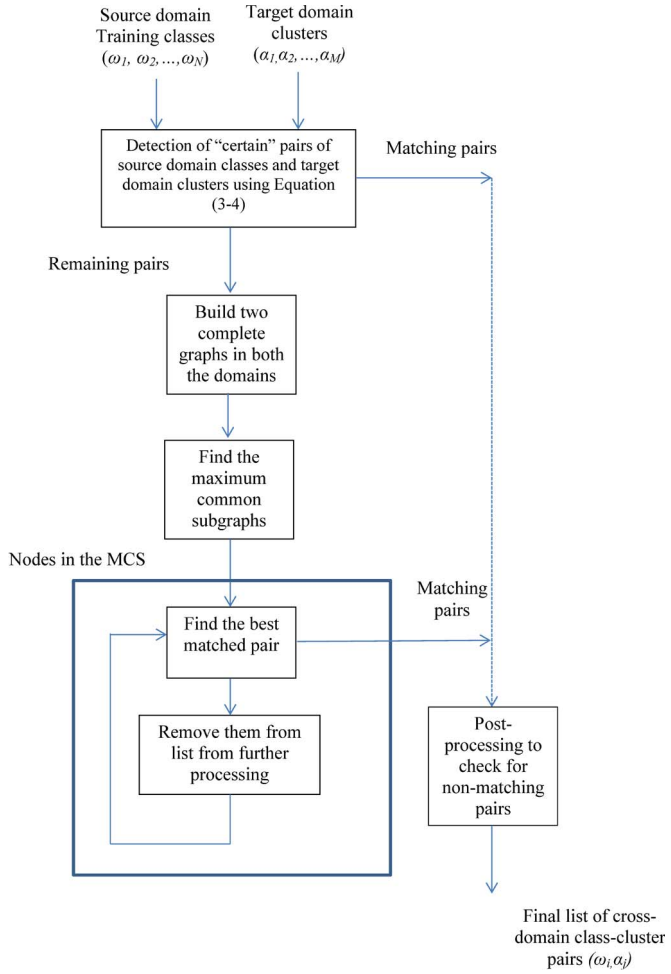


Fig. 3. Flowchart of the cluster mapping algorithm.

domains. $|\nabla_{S\mapsto T}(\omega_i)| = 1$ means that only one target domain node has ω_i as its best matched source domain counterpart. $|\nabla_{T\mapsto S}(\alpha_j)| = 1$ signifies that only one source class is the best unique match for the target cluster α_j. If there exists a one-to-one mapping between a given pair of source and target domain nodes and none of the remaining nodes has any of these two particular nodes as its best match, that source–target pair of nodes is considered to represent the identical land-cover class on the ground. It is also possible to find no pair in this step in some cases if the classes undergo severe changes between the considered images. The set of all of the certain pairs of (ω_i, α_j) and the JS divergence between the corresponding ω_i and α_j of such pairs are stored to check further for the nonequivalence of a given pair.

The source and the target domain nodes (classes/clusters) which are left unused after this step are considered to be the uncertain cases. Let R_S and R_T represent the set of all such classes and clusters of both domains separately. The next part of the proposed cluster mapping algorithm investigates the possible correspondences of these nodes. The

$$E_W(\alpha_j, \omega_i) = E_W(\omega_i, \alpha_j) = \sqrt{(\text{Min}(\text{KL}(p_T(X_T|\alpha_j), p_{ST}(X_{\alpha_j\cup\omega_i})), \text{KL}(p_S(X_S|\omega_i), p_{ST}(X_{\alpha_j\cup\omega_i}))))}$$

new target-domain-specific nodes (CASE B and CASE D; Section II) are also identified in that step.

The **identification of the uncertain pairs of source domain classes and the target domain clusters from** G is based on an iterative subgraph-matching algorithm. It aims at uniquely mapping the uncertain nodes reported by step 1 of the cluster mapping algorithm. It first builds two complete graphs, one for each domain excluding the already correctly detected classes and the clusters. The MCS of these two graphs defines the possible set of the common source domain classes and the target domain clusters. If some source domain nodes (classes) remain unused in the MCS, they are considered to disappear in I_T (CASE C; Section II). If it is the case with the target domain subgraph, they represent the new classes that appeared in I_T (CASE B).

Two individual intradomain complete graphs $G_S(V_S, E_S)$ and $G_T(V_T, E_T)$ are built using the remaining classes and the clusters in R_S and R_T, independently. The square root of the JS distance between a pair of classes (clusters) is used to define the edge weights in both graphs. The MCSs are found from G_S and G_T according to the following analysis. If $|V_S| <= |V_T|$, then the MCSs contain $|V_S|$ nodes; else, they contain $|V_T|$ nodes as both G_S and G_T are complete graphs. Several subgraphs of G_S and G_T can be isomorphic to each other. The particular pair of the subgraphs of G_S and G_T for which the sum of the following:

1) the absolute difference between the cumulative edge weights of the subgraphs;
2) the sum of the absolute difference between the corresponding maximum eigenvalues of the graph adjacency matrix of both subgraphs along with the mean edge weight (or the representative maximum eigenvalue) of the cross-domain graph considering the nodes of the MCS [25]

is minimized, is processed further. Since these subgraphs of G_S and G_T exhibit highest similarity in terms of the graph edge properties, they are considered to contain the identical set of underlying nodes (classes/clusters). A proper matching strategy is needed to map the equivalent nodes from both subgraphs. The remaining source domain or the target domain nodes, which are not part of the MCS, represent CASE B and CASE C. R_S and R_T are updated further only with the nodes in the MCS.

A cross-domain graph (G_{ST}) is built again considering the nodes in the updated R_S and R_T and using the same topology as of G, i.e., only cross-domain nodes are connected by edges. The edge weights are defined using the square root of the JS divergence measure as also used in G, G_T, and G_S and are stored in the weighted adjacency matrix A_{ST}.

An iterative algorithm is proposed to obtain all of the unique one-to-one mappings of the cross-domain nodes from (G_{ST}). It is considered that, at any iteration of the method, the specific source and the target domain nodes associated with the minimum JS distance from A_{ST} are identified to be matching. Once such a pair is found, the related row and column are deleted from the A_{ST} along with their entries in R_S and R_T. The successive iterations of the algorithm follow the same set of steps until R_S and R_T become empty (see Algorithm 2).

Algorithm 2

Input: R_S, R_T, and A_{ST}, the weighted adjacency matrix of the cross-domain graph built with the nodes from R_S and R_T

Output: The unique one-to-one mapping between the nodes in R_S, R_T

1: **while** $\mathrm{R}_S \neq \emptyset$ or $\mathrm{R}_T \neq \emptyset$ **do**
2: The minimum value of A_{ST} is identified. Let i and j represent the related row and column indices.
3: $R_S(i)$ and $R_T(j)$ are declared to be a matching pair and the corresponding entries are removed from both sets along with their entries in A_{ST}.
4: A new round of the processing takes place with the updated R_S, R_T, and A_{ST}.
5: **end while**
6: All of the matching pairs and the corresponding JS divergence between the members of each pair obtained in this step are stored for future references.

Once all of the cross-domain node pairs (classes and clusters) are obtained, CASE D (Section II) is analyzed. An analysis is carried out to verify if a given source domain class and a target domain cluster form a pair, yet they are not highlighting the same land-cover class. If they are not equivalent, their JS distance is much higher than that of the actual equivalent set of pairs. Those nodes are also not referenced together frequently in Rank_S and Rank_T as their closest neighbors from opposite domains tend to change at every iteration of the cluster mapping algorithm. These two properties are combined to check whether a pair is nonequivalent, i.e., although the corresponding nodes have been mapped by the proposed cluster mapping algorithm, they represent two different land-cover classes (see Algorithm 3).

According to Algorithm 3, it is checked, for each class and cluster, which is a part of the set of matching pairs produced by the cross-domain cluster mapping algorithm, whether the aforementioned conditions are satisfied. If the conditions are satisfied for a class (cluster), the corresponding pair of the source domain class and the target domain cluster is declared to be nonmatching.

Algorithm 3

Input: Set of all of the matching pairs produced by the cluster mapping stage, the JS divergence of the candidates of each pair and the lists of all of the Rank_S and Rank_T

Output: The inconsistent pairs which do not point to the same underlying land-cover classes

---

1: For each pair of (ω_i, α_j) produced by the cluster mapping algorithm, the following conditions are checked.
2: **Condition 1**: $JS(\omega_i, \alpha_j) >= \upsilon \times m$ {m is the minimum JS divergence of the pairs produced by the cluster mapping algorithm. υ is an user-defined constant value used to establish how much the distance between classes can be larger than m prior to concluding that the considered pair is a nonmatching one. It is dependent on m and is directly

proportional to the absolute difference between the average divergence of all of the pairs found and m.}
3: If k denotes the number of times (ω_i, α_j) pair has appeared in Rank_S and Rank_T during the cluster mapping process, another condition is checked.
4: **Condition 2**: $k \leq C$ where C is the total number of matching pairs produced by the cluster mapping algorithm.
5: If Condition (1) and (2) are satisfied simultaneously, (ω_i, α_j) is declared to be an inconsistent pair and hence, removed.

每个迭代的Rank都得记录下来吗？

The set of common land-cover classes which have undergone minimum or no changes between the acquisitions tends to have small JS divergence measures between them. It is comparatively easy to obtain a one-to-one mapping for these class–cluster pairs in the first stage of the proposed cluster mapping strategy. However, if some different classes get shifted largely and overlap with them, then correct mapping for the stable classes would be difficult to obtain in the first stage. This is due to the possible presence of many-to-many correspondences between the classes and clusters in terms of the smallest divergence measure. In order to resolve this issue, the second stage of the cluster mapping method breaks the many-to-many correspondences by mapping the class–cluster pair which is closest and subsequently removing both of them from the lists. In this way, it is ensured that, at each iteration of cluster mapping, the source domain class and the target domain cluster which have maximum similarity in the spectral domain until that moment are mapped together. A false matching can occur for a given class (cluster) if some other classes (clusters) entirely overlap with the considered class (cluster), making it impossible to distinguish them in the spectral domain. This issue can be resolved by selecting multiple robust pixel level features along with the spectral properties of the same to ensure that different classes do not overlap entirely in the feature space. However, inconsistent matching can only take place during the latter iterations of the cluster mapping algorithm which is detected by the postprocessing stage, whereas during the initial iterations, only very similar classes and clusters are mapped.

这是因为在最小散度方法下，类和簇之间可能存在多对多的对应

为了解决这个问题，集群映射方法的第二阶段，打破了多对多的对应关系

通过构建类簇最近的映射对，而后将他们都从列表中移除

同时选择多个鲁棒性强的像素级特征和光谱特征一起，保证在特征空间中不同的类不会重叠在一起

After this step, all of the target domain clusters for which there are equivalent source domain classes are declared.

C. Cluster Refinement to Obtain the Final Target Domain Land-Cover Classification

集群精化获取最终的目标域地表分类

This step consists of updating the target domain classification result by exploring data from both domains. The main goals of this steps are the following.

1) Define a set of training samples per cluster of I_T. For all of the matching pairs, an algorithm is proposed to highlight a set of target-specific training samples from the information of both domains. For the new target domain classes, the corresponding cluster information is used.
2) The adaptation of source domain properties to I_T ones is achieved in this step through the iterative EM algorithm.
3) The final land-cover map of I_T is generated by an ML classifier with the updated parameter set.

Algorithm 4 describes the steps for defining the set of training samples for all of the target domain clusters for which matching source domain classes are present. Considering the fact that the data of some of the classes may be overlapped, the proposed training sample selection algorithm overclusters both ω_i and α_j of a given matching pair of (ω_i, α_j). The particular subcluster of α_j which is closest to ω_i in terms of the JS divergence is merged with ω_i, and this updated set of samples α_j^{Tr} is used as the training data for the class represented by α_j. For the remaining new target domain clusters, for which the actual class labels are unknown, the entire cluster information is used to define the training set considering random labels for the underlying unknown land-covers.

Here, classification is conducted in the context of Bayes decision theory by maximizing the posterior probability of a sample x_{Tl} for a class α_k [26] according to

$$x_{Tl} \in \alpha_k \Leftrightarrow \underset{\alpha_j \in \text{I}_T}{\text{argmax}} \{P_T(\alpha_j) p_T(x_{Tl}|\alpha_j)\} \tag{5}$$

where $P_T(\alpha_j)$ is an estimate of the prior probability of the class α_j in the target domain image. $p_T(x_{Tl}|\alpha_j)$ is the conditional probability estimated for $p_T(X_T|\alpha_j)$ for the feature vector x_{Tl} given α_j in I_T. X_T is the random variable associated with I_T. It is worth noting that the subscript T has been used here to stress the dependences of both statistical terms on the considered target image I_T. According to (5), the training phase of the ML classifier requires the estimation of the prior probability $P_T(\alpha_j)$ and the conditional probability $p_T(X_T|\alpha_j)$ for each class $\alpha_j \in I_T$. Such estimates can be obtained by exploiting the information in α^{Tr}. It is usually assumed in the remote sensing community that natural classes present in the images acquired by passive sensors are Gaussian distributed [6], i.e., the conditional probability term in (5) can be modeled as a Gaussian function. Under such an assumption, the marginal probability $p_T(X_T)$ of the pixel values in I_T can be approximated by a mixture of Gaussians. Hence, the solution of the ML classifier relies on the estimations of the set of parameters $\theta_{Tj} = \{\mu_{Tj}, \Sigma_{Tj}, P_T(\alpha_j)\}$, where μ_{Tj} and Σ_{Tj} define the mean vector and the covariance matrix for class α_j and $P_T(\alpha_j)$ is the prior probability of class α_i calculated as the relative frequency. Even if, in this work, classes are assumed to be Gaussian distributed, the same logic can be applied for any other model.

Algorithm 4
Input: The set of matching pairs, I_S, I_T
Output: A set of reliable training samples for each classes in I_T

1: **for** Each pair of matching (ω_i, α_j) **do**
2: Vector quantize (Cluster) the points in ω_i and α_j into same number of clusters l using any clustering algorithm.
3: $\omega_i = \{\omega_{i1}, \omega_{i2}, \ldots, \omega_{il}\}$.
4: $\alpha_j = \{\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{jl}\}$.
5: Mat is a $l \times l$ matrix which stores the pairwise JS divergence of all of the clusters of ω_i and α_j. {ω's are in rows and α's are in columns}
6: Obtain the indices of the minimum value of Mat, (Min(Mat)), i.e., (k, m), which corresponds to the cluster

pair of ω_i and α_j which are closest to each other in terms of the JS distance measure.

7: The training set for the jth target domain class is defined as $\alpha_j^{\mathrm{Tr}} = (\omega_i \cup \alpha_{jm})$. 源域对应的类别样本+目标域对应簇的小簇样本

8: **end for**

Considering the initial class statistical parameter estimates obtained from the newly defined training set (including information from both source and target domain) and under the assumption of the Gaussian distributed classes, the iterative EM equations are to be applied for adaptation of the statistical parameters. The true statistical terms $\theta_{Tj} = \{\mu_{Tj}, \Sigma_{Tj}, P_T(\alpha_j)\}$ 协方差 associated with each cluster α_j of I_T are computed as follows

$$P_T^{k+1}(\alpha_j) = \frac{1}{|I_T|} \sum_{x_{Tl} \in X_T} \frac{P_T^k(\alpha_j) p_T(x_{Tl}|\alpha_j)}{P_T^k(x_{Tl})} \tag{6}$$

$$\mu_{Tj}^{k+1} = \frac{\sum_{x_{Tl} \in X_T} \frac{P_T^k(\alpha_j) p_T(x_{Tl}|\alpha_j)}{P_T^k(x_{Tl})} x_{Tl}}{\sum_{x_{Tl} \in X_T} \frac{P_T^k(\alpha_j) p_T(x_{Tl}|\alpha_j)}{P_T^k(x_{Tl})}} \tag{7}$$

$$\Sigma_{Tj}^{k+1} = \frac{\sum_{x_{Tl} \in X_T} \frac{P_T^k(\alpha_j) p_T(x_{Tl}|\alpha_j)}{P_T^k(x_{Tl})} (x_{Tl} - \mu_{Tj}^{k+1})^2}{\sum_{x_{Tl} \in X_T} \frac{P_T^k(\alpha_j) p_T(x_{Tl}|\alpha_j)}{P_T^k(x_{Tl})}}. \tag{8}$$

The superscripts k and $k+1$ refer to the values of the parameters at two consecutive iterations, respectively. The iterative process terminates when a local maximum of the negative log-likelihood function $L(\theta_T)$ is achieved.

For each of the four cases mentioned in Section II, an adequate initialization of the retraining process should be defined.

1) **CASE A**: $\Omega_S = \widehat{\Omega_T}$: In this case, no change other than only a spatial shift of the classes has been detected in both domains. In this situation, the initialization of the statistical terms is based on the training set obtained by Algorithm 4. Let θ_T^{New} denote the initial estimates of the class statistical parameters obtained from the newly defined training set for the common classes of Ω_S and $\widehat{\Omega_T}$. The iterative EM retraining is hence followed after.
2) **CASE B**: $\widehat{\Omega_T} = \Omega_S \cup \{\alpha_u\}$: This is the case where new classes have been detected in I_T. For all of the classes common to I_S and I_T, the same initialization approach as in CASE A can be adopted, whereas this is not possible for classes in $\{\alpha_u\}$. The statistical terms for those classes are calculated from the corresponding clusters obtained in I_T. The initial set of class statistical parameters to be used in the iterative adaptation process is defined as

$$\theta_T^0 = \theta_T^{\mathrm{New}} \cup \left\{\mu_u^0, \Sigma_u^0, P_T^0(\alpha_u)\right\}. \tag{9}$$

The prior probabilities are scaled based on the newly defined training set for I_T.
3) **CASE C**: $\widehat{\Omega_T} = \Omega_S \backslash \{\omega_u\}$: This is the case where some classes disappear between I_S and I_T. For all of the classes common to I_S and I_T, the same initialization approach as in CASE A is followed. As the classes in $\{\omega_u\}$ are no longer present in I_T, the statistical terms related to those classes are not needed for the updating process. Therefore, θ_T^0 in this case is defined as

$$\theta_T^0 = \theta_T^{\mathrm{New}} \backslash \{\mu_u^0, \Sigma_u^0, P_T^0(\omega_u)\}. \tag{10}$$

4) **CASE D**: $\widehat{\Omega_T} = \Omega_S \cup \{\alpha_k\} \backslash \{\omega_u\}$: This case is a combination of CASE B and CASE C. Accordingly, θ_T^0 in this case is defined as

$$\theta_T^0 = \theta_T^{\mathrm{New}} \cup \left\{\mu_k^0, \Sigma_k^0, P_T^0(\alpha_k)\right\} \setminus \left\{\mu_u^0, \Sigma_u^0, P_T^0(\omega_u)\right\}. \tag{11}$$

From these initial estimates, the iterative EM algorithm provides an estimate of the true class statistical parameters in I_T. The Bayesian classifier in (5) is used to generate the final land-cover map of I_T. Any other classifier working under the assumption that Ω_S and Ω_T might include different sets of classes can be adopted.

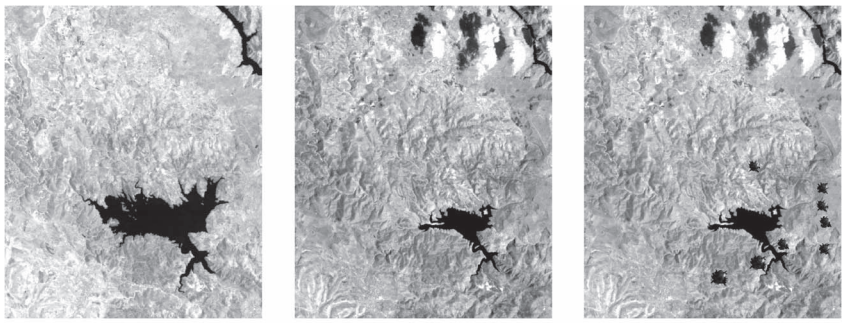
IV. Experimental Results

A. Design of the Experiments

In order to assess the performance of the proposed DA algorithm, different experiments on two multispectral and one hyperspectral data sets were conducted. Experiments have been designed at three different levels of the proposed framework as follows.

1) The target domain data are required to be clustered optimally in order to create provision for perfect cross-domain cluster mapping. In order to validate the outcome of the initial clustering stage (Section III-A), the well-known Silhouette cluster validity [27] measure has been used. It ranges in $[-1, +1]$, and a large value of this index indicates a stable clustering result. We opted for this particular index measure as it has been reported to provide a better measure on the quality of clustering than other validity indices [28].

   The kernel parameter γ has been selected for $k = N - n$ by grid search strategy. For the given k, different values of γ have been considered in the range $(0, 1]$ with a step of 5×10^{-4}. The particular clustering result which maximizes the Silhouette index has been declared to be the optimal clustering result with k clusters. Furthermore, M is selected to be the k which produces clusters with the maximum Silhouette measure within $[N - n, N + n]$. The γ value obtained for $k = N - n$ has been used for the remaining cases. $n = \lceil N/2 \rceil$ has been considered for all of the data sets. It has been observed that kernel k-means along with the γ found in this way has produced good clustering results for all of the cases. However, as previously mentioned, different ranges of the same are equally plausible. The kernel k-means algorithm has been used using the median centroids.
2) The cross-domain cluster mapping step outputs pairs of the similar classes shared by both domains. The proposed cluster mapping technique has been compared with a recent cross-domain cluster mapping strategy used for DA [29]. The cluster mapping is achieved in [29] by means of a transformation matrix based on the geometric means of the covariances, estimated from the covariance matrices of the data from both domains. It further assumes that



(a) (b) (c)

Fig. 4. Band 4 of the (a) July 96 image (Z), (b) September 95 image (Y), and (c) simulated September 95 image with burned area (Y').

TABLE II
CARDINALITY ($|\cdot|$) OF TRAINING (Tr) AND TEST (Ts) SETS PER CLASS AVAILABLE FOR THE SARDINIA DATA SET

	Z (July '96)		Y' (September '95)		Y (September '95)	
Land Cover	$\|Tr\|$	$\|Ts\|$	$\|Tr\|$	$\|Ts\|$	$\|Tr\|$	$\|Ts\|$
Pasture	554	589	554	170	554	170
Forest	304	274	128	159	304	274
Urban	408	418	408	418	408	418
Water	1120	551	804	551	804	551
Vineyard	179	117	179	117	179	117
Burned Area	-	-	176	115	-	-

the domains share the same set of classes. Hence, the comparison performed in this case has been based on checking how well the proposed method captures the correspondences between the common set of classes shared by the domains without any assumption.

3) Two different strategies have been followed to evaluate the performance of the proposed EM+ML-based classifier system in producing a reliable land-cover map of the target domain data.
   a) For the common set of classes, the class-by-class classification accuracies of the proposed algorithm in producing the land-cover map for the target domain image have been compared with three well-studied techniques from the literature, e.g., the clustering result of kernel k-means, an ML classifier trained solely on the source domain data, and an EM+ML technique with class statistical parameters initialized from the available source domain training samples [6]. In the latter two cases, the trained classifier is used in the target domain for testing. It is worth noting that the proposed framework automatically selects a set of reliable target domain samples per cluster to be used in the EM-based retraining step along with the available labeled source domain samples from the similar class. However, this analysis is not possible for the additional target domain classes.
   b) The overall classification accuracy of the proposed technique has been compared with the one of a supervised ML classifier trained using the reliable target-domain-specific training samples and kernel k-means.

To carry out the experiments, it has been assumed that, for each of the image pairs, only the training set associated with the image considered as the source was available, whereas the reference ground data associated with the image considered as the target were used only to evaluate the performance of the proposed technique. It is to be noted that the proposed method considers a single image in both domains. If multiple images are to be used for the domains, it must be assured that classes are well distributed among the images to prevent redundancy in data. Furthermore, for multitemporal images, the spatial correlation between the images is of no use for the proposed technique considering that the algorithm works only with spectral features. Five independent realizations of all of the experiments have been conducted, and the average results are reported here.

Parameter v (Algorithm 3) is user specified. For all of the data sets, the best value for v has been found to lie between 2 and 4.

B. Medium-Resolution Sardinia Data Set

The first study area consists of two coregistered multispectral images acquired by the Thematic Mapper sensor of the Landsat-5 satellite. The images consist of six bands (1–5 and 7), and band 6 is neglected due to its lower geometrical resolution. The selected test site is a section of 412 × 493 pixels of a scene, including the area surrounding the Lake Mulargia on the Island of Sardinia (Italy). The images were acquired in September 1995 (Y) and July 1996 (Z), respectively. Fig. 4(a) and (b) depicts band 4 of both images. The images share five land-cover classes, i.e., pasture, forest, water, vineyard, and urban. The September 1995 image contains an additional simulated land-cover class named burned area. Fig. 4(c) shows band 4 of the September 1995 image with the burned-area class (Y'). Table II lists the classwise distributions of the training and test samples of all of the images used for experimental purposes. All of the images have been used alternatively as source and

核kmeans聚类结果
只用源域数据训练的最大似然分类器
类统计参数由源域训练样本初始化的EM+ML方法
实验过程中，对应每个图像对，只有与源域图像相关联的训练样本可以使用，
然而目标域参考数据只是用来评价

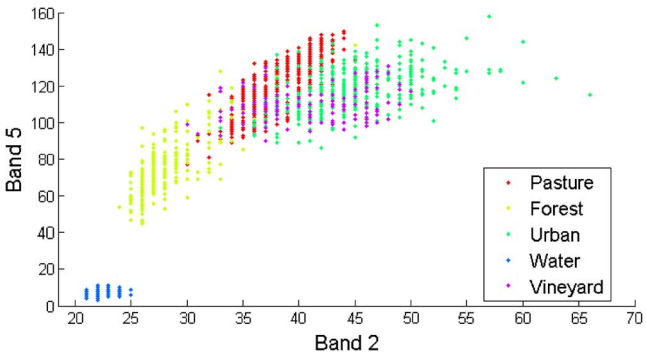


Fig. 5. Scatter plots of bands 2 and 5 of the July 96 image (Z).

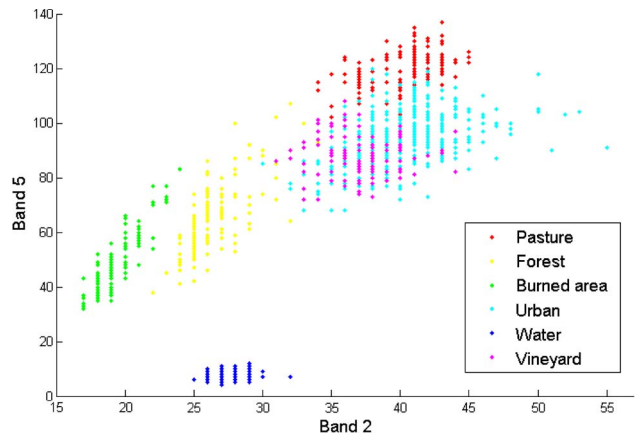


Fig. 6. Scatter plots of bands 2 and 5 of the simulated September 95 image with burned area (Y').

target domain. Figs. 5 and 6 depict the scatter plot of the test sets of (Z) and (Y') of bands 2 and 5, respectively. It can be observed from the scatter plots that pasture and the vineyard classes have highly similar spectral signatures, i.e., these two classes are highly overlapped. The JS divergence measures between the training samples of Z and Y indicate that pasture (1.32) and water (1.70) have undergone a significant shift, while the remaining classes are stable in terms of the statistical properties.

Nine experimental cases have been considered for the Sardinia data set with Z and Y or Y' representing the source or target domain alternatively. In the current setup, the experimental case Sardinia $Z5Y5-1$ indicates that the Z and Y are the source and target domain images, and they contain five classes each, with one common class shared between them.

1) **Sardinia 1** $Z3Y4-1$: Z and Y have been considered to be the source and the target domain images, respectively. Pasture, vineyard, and water classes have been considered for Z, while water, forest, burned area, and urban have been selected for Y.
2) **Sardinia 2** $Z5Y5-5$: Z and Y have been considered to be the source and the target domain images, respectively. The same set of five classes has been considered for the images. $\Omega_S = \Omega_T$.
3) **Sardinia 3** $Z5Y'6-5$: We have considered Z and Y' to be the source and the target domain images in this case. Y' contains an additional class of burned area other than the five common classes. $\Omega_T = \Omega_S \cup \{\text{Burned-area}\}$.
4) **Sardinia 4** $Z5Y'5-4$: Z and Y' have been considered to be the source and the target domain images, respectively. In Y', the burned-area class has been added, while the vineyard class has not been considered. Hence, although the same number of land-cover classes is present in both domains, one of the pairs is nonequivalent. $\Omega_T = \Omega_S \cup \{\text{Burned-area}\}\backslash\{\text{Vineyard}\}$.
5) **Sardinia 5** $Z4Y'6-4$: Z and Y' have been used as the source and the target domain images, respectively. Four classes have been considered in Z, excluding the urban class. In Y', six classes have been considered. In addition to the four classes of Z, burned area and urban have been included in Y', i.e., $\Omega_T = \Omega_S \cup \{\text{Burned-area,Urban}\}$.
6) **Sardinia 6** $Y3Z3-1$: Y and Z have been selected to be the source and the target domain images, respectively. Pasture, vineyard, and water classes have been considered for Y while selecting water, forest, and urban for Z.
7) **Sardinia 7** $Y5Z5-5$: We have used Y and Z as the source and the target domain images, respectively. The same set of five classes has been considered for both images. $\Omega_S = \Omega_T$.
8) **Sardinia 8** $Y'6Z5-5$: Y' and Z have been considered to be the source and the target domain images, respectively. Y' contains an additional class of burned area other than the five common classes, i.e., $\Omega_T = \Omega_S\backslash\{\text{Burned-area}\}$. This is the same as experiment 2 but inverting the roles of the source and the target domains.
9) **Sardinia 9** $Y'6Z4-4$: Y' and Z are used in the source and the target domains. Four classes have been considered in Z, excluding the urban class. In Y', six classes have been considered; in addition to the four classes of Z, burned area and urban have been included in Y', i.e., $\Omega_T = \Omega_S\backslash\{\text{Burned-area,Urban}\}$.

Sardinia 1–5 consider Z and $Y(Y')$ as the source and the target domain images, respectively, whereas their roles are reverted in experiments 6–9. Experiments 1 and 6 consider the case where only one common class exists in both domains. Similarly, experiments 2 and 7 consider the cases where both domains contain the same set of classes. The remaining experiments consider different combinations of classes in both domains.

The initial clustering step of the proposed method correctly found out the optimal target domain clustering in all of the cases. The result has also been validated by the maximization of the Silhouette index for the number of clusters M found corresponding to the minimum values of F (1) in all of the cases. A typical γ value obtained for Sardinia 1 was 0.00075, which has further been used for all of the remaining experimental cases. υ has been set to an average value of 3.75 to compensate for the high variability of the divergence measures for all of the matching pairs of the source domain classes and target domain clusters found in the cross-domain cluster mapping stage for all of the experiments conducted.

In the cross-domain cluster mapping stage, the water and forest classes are easy to map as one-to-one mapping exists in both cases with respect to the JS divergence. Unique injective mappings are difficult to obtain for the pasture, vineyard, and

TABLE III
AVERAGE OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) COMPUTED ON THE TARGET DOMAIN FOR THE COMMON SET OF CLASSES OVER FIVE ITERATIONS, SARDINIA EXPERIMENTS 1–9

Experiment	Kernel k-means	ML trained on source domain	EM+ML classifier of [6]	Proposed method
Sardinia 1 $Z3Y4-1$	100.00	36.56	100.00	100.00
Sardinia 2 $Z5Y5-5$	83.39	55.47	86.71	94.34
Sardinia 3 $Z5Y'6-5$	83.49	55.47	86.71	94.34
Sardinia 4 $Z5Y'5-4$	90.11	60.09	92.06	98.22
Sardinia 5 $Z4Y'6-4$	89.95	54.76	97.89	98.49
Sardinia 6 $Y3Z3-1$	100.00	41.39	100.00	100.00
Sardinia 7 $Y5Z5-5$	73.98	67.06	79.52	84.14
Sardinia 8 $Y'6Z5-5$	73.98	67.06	79.52	84.14
Sardinia 9 $Y'6Z4-4$	75.63	64.01	90.65	91.70

TABLE IV
AVERAGE OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) COMPUTED ON ALL OF THE TARGET DOMAIN CLASSES OVER FIVE ITERATIONS, SARDINIA EXPERIMENTS 1–9

Experiment	Kernel k-means	Proposed method	Supervised ML trained on target domain
Sardinia 1 $Z3Y4-1$	98.06	99.19	99.43
Sardinia 2 $Z5Y5-5$	83.39	94.34	94.41
Sardinia 3 $Z5Y'6-5$	84.24	92.94	94.77
Sardinia 4 $Z5Y'5-4$	91.86	98.51	98.58
Sardinia 5 $Z4Y'6-4$	84.24	93.65	94.77
Sardinia 6 $Y3Z3-1$	98.47	98.71	99.03
Sardinia 7 $Y5Z5-5$	73.98	84.14	92.66
Sardinia 8 $Y'6Z5-5$	73.98	84.14	92.66
Sardinia 9 $Y'6Z4-4$	75.63	91.70	94.05

urban classes due to high overlapping. The interdomain cluster mapping strategy of [29] fails to obtain the correct mappings for the common set of classes, particularly for experiments concerning the aforementioned three classes even if a sample point per class is shared between the domains for those classes. In [29], data from both domains are aligned in the same dimensions using a rotation matrix involving their covariance matrices. This causes data from the overlapping classes difficult to distinguish from each other.

Tables III and IV show the comparisons of classification accuracies of the proposed method for the experimental cases mentioned previously for the common set of classes and the entire target domain, respectively. The water class is well separated from the remaining land-cover classes, but it has undergone a substantial spectral shift from one domain to the other. Kernel k-means is able to detect the water class properly (100%). For the ML classifier trained on the source domain samples, the water class is heavily misclassified due to the shift in the spectral domain between acquisitions of the respective images. It is reflected in the classification accuracies of the water class in Sardinia 1 and 6 (36.56% and 41.39%). The application of the EM algorithm enhances the classification accuracy of the water class successfully to 100% for the EM+ML classifier. It is observed from Table III that kernel k-means clustering is affected by the problem due to data overlapping to some extent (73.98%–90.11% for Sardinia 2–5 and 7–9). The ML classifier trained on the source domain performs poorly in all of the experimental cases, with the best performance of 67.06% for Sardinia 7 and 8 in classifying the common set of land-cover classes shared between the domains. A sharp enhancement in terms of the classification accuracy is observed when EM retraining is used along with ML classifier (79.52%–97.89% excluding Sardinia 1 and 6, where only water class is highlighted). The ML+EM classifier is able to adopt to the target domain better with the proposed mixed set of training samples and shows a superior performance than the ML+EM classifier initialized on the source domain training data with an enhancement of 1.5%–6% in generalization accuracy for the common set of classes. The overall generalization performance of the proposed classifier (84.14%–99.19%) is far better than the kernel k-means-based classification results (73.98%–98.47%) and is very close to the results of a supervised ML classifier trained on the reliable target domain training samples (92.66%–99.43%).

In order to exhibit the working of the proposed cross-domain cluster mapping algorithm, a special case of the Sardinia data has been considered. In the source domain, pasture, forest, urban, and water (source classes 1–4) classes have been selected, whereas pasture, forest, burned area, and water (target classes 1–4) have been used in the target domain. Table V shows the adjacency matrix of the cross-domain graph built using these classes of both domains.

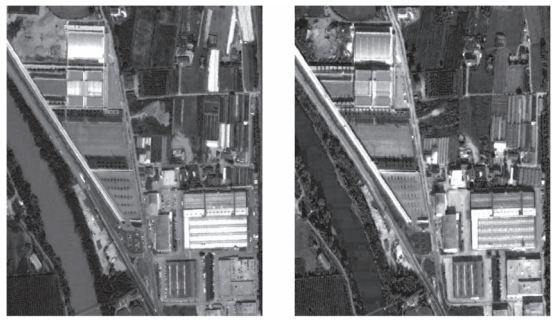
TABLE V
ADJACENCY MATRIX OF THE CROSS-DOMAIN GRAPH

	Target Class 1	Target Class 2	Target Class 3	Target Class 4
Source Class 1	1.48	1.51	2.57	4.24
Source Class 2	2.01	0.72	2.49	4.19
Source Class 3	1.86	1.79	3.09	4.49
Source Class 4	4.24	3.86	4.35	1.50

The cluster mapping for the certain cases has indicated unique one-to-one mappings for (source class 1, target class 1) and (source class 4, target class 4) pairs. The remaining source and target domain classes (source class 2, source class 3, target class 2, and target class 3) have been processed further in the second stage of the cluster mapping algorithm as unique one-to-one mappings have not been found for them. Two complete graphs have been constructed in both domains using the remaining classes independently, and the MCSs of both domains have two nodes each. The analysis of the uncertain cases has indicated two more pairs (source class 2, target class 2) and (source class 3, target class 3). The JS divergence between class 2 of both domains is 0.72, which is also the minimum JS divergence between all pairs. On the other hand, the distance is 3.09 between class 3 of both domains, which is greater than 3.75×0.72 (Algorithm 3). It indicates that class 3 of both domains is not pointing toward the same land-cover class, and these classes reside far apart in the spectral domain. Hence, (source class 3, target class 3) is not a matching pair according to the postprocessing step.

C. Very High Resolution QuickBird Data Set

The second data set is made up of two coregistered and pan-sharpened multispectral very high geometrical resolution (VHR) images acquired by the QuickBird satellite. All of the four spectral bands of QuickBird have been considered in the experiments. The selected test site is a section of 992×992 pixels of a scene including an area on the southern part of the city of Trento (Italy). The two images were acquired in October 2005 (Y) and July 2006 (Z), respectively. The available prior information about the considered area has been used to build a training set and a test set for each image. Four main land-cover classes common to both dates were identified, i.e., water, red-roof, asphalt, and field. For the image acquired in July 2006, one additional class has been detected, i.e., plastic-mulched fields. Band 1 of both images is shown in Fig. 7. Table VI lists the classwise number of training and test samples for both images used for the experiments, along with the JS divergences between the pairs of common training classes. It is evident from the classwise JS divergences of the training samples of both images that all of the classes have undergone a significant spectral shift, thus making the cross-domain cluster mapping process more complex. Considering the seasonal change that has been reflected in the classwise divergence measures for both images, the minimum divergence (m) is comparatively higher, while the absolute difference between the average and minimum divergences of the matching pairs is lesser in comparison to the Sardinia data set. Accordingly, υ has been set to an average value of 2.



(a) (b)

Fig. 7. Band 4 of the (a) July 2006 image (Z) and (b) October 2005 image (Y).

The classes have undergone a considerable shift between acquisitions. However, the classes are well separated in the feature space, thus making it possible to properly detect them. Two experimental cases have been considered here, with the same naming convention as that of the experimental cases of the Sardinia data set being followed here.

1) **VHR 1** $Y4Z3-3$: Y and Z have been considered to be the source and the target domain images, respectively. Y contains an additional class of field in addition to the three land-cover classes common to them, i.e., water-body, red-roof, and asphalt. $\Omega_S = |\Omega_T| \cup \{\text{Field}\}$.
2) **VHR 2** $Z4Y3-3$: Z and Y are selected as the source and the target domain images in this case. Z contains an additional class of field. The remaining three land-cover classes, i.e., water-body, red-roof, and asphalt, are present in both images, i.e., $\Omega_S = |\Omega_T| \cup \{\text{Field}\}$.

The clustering step has easily been performed in this case as the clusters are well defined as well as well separated from one another. This notion has also helped in performing injective mapping between the common set of classes and clusters without much confusion.

Tables VII and VIII report the classification accuracies of the proposed method for the common set of classes and the overall generalization accuracies for the different experimental cases, respectively. All of the classes are almost properly detected by all of the techniques (classification accuracy $\geq 99\%$), except the ML classifier trained on the source domain samples (49%–50%). The water-body class has undergone an extensive shift in the spectral domain, and it is entirely undetected in the target domain image. The application of the EM algorithm is also unable to alleviate the situation. The abrupt spectral shift is responsible for the misclassification caused by the ML classifier on the target domain data with the training set defined on the source domain. The proposed algorithm tackles such situation

TABLE VI
CARDINALITY ($|\cdot|$) OF TRAINING (Tr) AND TEST (Ts) SETS PER CLASS AVAILABLE FOR THE QUICKBIRD DATA SET

	Oct. 2005' image (Y)		July 2006' image (Z)		
Land Cover	$\|Tr\|$	$\|Ts\|$	$\|Tr\|$	$\|Ts\|$	JS divergence between training samples
Water	1099	1104	1099	1104	4.07
Red-roof	449	469	449	469	2.62
Asphalt	673	474	673	474	1.87
Field	647	534	647	534	2.14

TABLE VII
AVERAGE OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) COMPUTED ON THE COMMON SET OF TARGET DOMAIN CLASSES OVER FIVE ITERATIONS, QUICKBIRD EXPERIMENTS 1 AND 2

Experiment	Kernel k-means	ML trained on source domain	EM+ML classifier of [6]	Proposed method
VHR 1 $Y4Z3-3$	99.90	49.41	49.41	100.00
VHR 2 $Z4Y3-3$	100.00	50.02	50.02	100.00

TABLE VIII
AVERAGE OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) COMPUTED ON ALL OF THE TARGET DOMAIN CLASSES OVER FIVE ITERATIONS, QUICKBIRD EXPERIMENTS 1 AND 2

Experiment	Kernel k-means	Proposed method	Supervised ML trained on target domain
VHR 1 $Y4Z3-3$	99.90	100.00	100.00
VHR 2 $Z4Y3-3$	100.00	100.00	100.00

by incorporating the target domain samples in the classification process.

D. Hyperspectral Data Set

The third study area considered is represented by a pair of hyperspectral images that is used as a benchmark in the remote sensing literature and consists of data acquired by the Hyperion sensor of the EO-1 satellite over a 1476×256 pixel study area located in the Okavango Delta, Botswana, on May 31, 2001. The considered image has a spatial resolution of 30 m; thus, it covers a large strip of 7.7×44.3 km^2 with 145 out of 242 original spectral bands. The bands affected by noise and the water absorption bands are neglected. A detailed description of the data set can be obtained in [30]. Here, 14 land-cover classes are identified for two different spatially disjoint areas, denoted by Y and Z, respectively. Many of the classes from the identified set of 14 have highly similar spectral signatures which even a kernel-based clustering technique fails to distinguish. The goal of the proposed framework is not to optimize the clustering step but to demonstrate the effectiveness of the cross-domain cluster matching technique and the further postprocessing for land-cover map refinement. Thus, a subset of nine classes has been considered such that the clustering works at least to a certain extent. These nine classes have initially been divided between the domains, with each domain containing five classes, with one common class shared between the domains. However, most of the classes considered are shifted substantially in the feature space and are largely overlapped with each other. The distributions of the training and test samples extracted from both images are mentioned in Table IX. From the set of 145 spectral bands, 10 bands that maximizes the discrimination capability among the classes have been selected according to the method proposed in [31]. This step ensures the removal of the redundant and nondiscriminant bands and maintains the recommended ratio between the number of feature dimensions and the available number of samples. The JS divergence for the common pair of classes of the training set of both areas is also reported in Table IX. It is clear from the divergence measures that the classes floodplain grass-1, island interior, acacia grasslands, and exposed soil are associated with a significant shift (JS divergences of 1.59, 1.60, 1.52, and 1.60, respectively), while the remaining classes are more stable.

Two distinct sets of experiments totaling 12 have been performed for the hyperspectral data set considering Y and Z alternatively as the source and target domains. Initially, five land-cover classes have been selected for each domain, with one common class shared between them, i.e., water, floodplain grass-1, reeds1, firescar2, and island interior classes have been selected to represent Y, while water, acacia woodland, acacia grassland, mixed mopane, and exposed soil have been picked to characterize Z. In the first set of experiments, Y and Z have been considered to be the source and the target domain, respectively (Hyper 1–6). Their roles have been reverted for the remaining set of experiments (Hyper 7–12). Hyper 1 and Hyper 7 denote the cases where only one land-cover class is shared between Y and Z. Subsequently, similar source domain classes have been added to the target domain iteratively with the addition of a random class to the target domain in each iteration (Hyper 2–5 and Hyper 8–11). In addition, a special experimental case has been considered, where both domains share a common set of the nine aforementioned land-cover classes (Hyper 7 and 12).

TABLE IX
CARDINALITY ($|\cdot|$) OF TRAINING (Tr) AND TEST (Ts) SETS PER CLASS AVAILABLE FOR THE HYPERSPECTRAL DATA SET

	Area 1 (Y)		Area 2 (Z)		
Land Cover	$\|Tr\|$	$\|Ts\|$	$\|Tr\|$	$\|Ts\|$	JS div. for training data
Water	69	57	213	57	0.68
Floodplain grass-1	83	75	199	52	1.59
Reeds1	80	88	219	50	0.78
Firescar2	93	83	215	44	1.23
Island interior	77	77	166	37	1.60
Acacia woodland	84	67	253	61	1.07
Acacia grasslands	184	174	243	62	1.52
Mixed mopane	68	85	154	27	1.30
Exposed soil	41	48	81	14	1.60

TABLE X
AVERAGE OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) COMPUTED ON THE COMMON SET OF TARGET DOMAIN CLASSES OVER FIVE ITERATIONS, HYPERSPECTRAL EXPERIMENTS 1–6

Experiment	Kernel k-means	ML trained on the source domain	EM+ML classifier of [Bruzzone et. al.]	Proposed method
Hyper 1 $Y5Z5-1$	100.00	100.00	100.00	100.00
Hyper 2 $Y5Z6-2$	95.80	72.11	70.19	100.00
Hyper 3 $Y5Z7-3$	88.52	75.81	75.30	96.56
Hyper 4 $Y5Z8-4$	76.37	80.80	81.80	96.67
Hyper 5 $Y5Z9-5$	72.45	73.83	76.25	97.34
Hyper 6 $Y9Z9-9$	72.62	60.85	75.33	98.41

For Hyper 1, the target domain data have been found to be clustered into five classes while maximizing F (1) for $\gamma = 0.0000085$. The clustering result has further been validated by calculating the Silhouette index for each clustering result in the given range. The index measure is maximum (0.87) for the optimal case, i.e., when the target domain data have been clustered into five groups. The same value of γ has been used for the remaining experiments (Hyper 2–12).

The proposed cross-domain cluster mapping step has been performed henceforth. υ has been set to a mean value of 3.75, similar to the Sardinia data set. Only a valid one-to-one mapping has been obtained for the water class, which has also been represented by the pair with the smallest JS divergence (0.76) between all of the corresponding source domain classes and the target domain clusters. For the remaining set of pairs, the high JS divergence between the corresponding members (> 2.80) has confirmed the absence of any further matching pairs. For Hyper 1, [29] has performed a correct cluster mapping for the water class as indicated by the small JS divergence (0.73) between the members of the corresponding class-cluster pair. It is to be noted that the parallel clustering technique of [29] produces a minor degraded result compared to the kernel k-means used in the proposed setup as indicated by the JS divergence measures in both cases.

Hyper 2–5 have been carried out in a similar fashion. The target domain clustering result has been validated in the same way of Hyper 1 using the Silhouette index measure. The cluster mapping step in each of the cases was able to produce the correct number of matching pairs. It is interesting to see that [29] was unable to produce well-defined clusters when new classes were added to the target domain with high degree overlapping with the existing classes. In particular, [29] failed to detect the floodplain grass-1, island interior, and firescar2 properly given that these classes are very much overlapping in the spectral domain. However, the rest of the classes are not affected by such amount of severe cluster overlapping and were detected with high confidence. Likewise, for Hyper 7–12, the proposed technique has successfully performed optimal target domain clustering, followed by the cross-domain cluster mapping without any false matching.

Once the common set of classes and clusters is identified, the final target domain land-cover classification is performed using the proposed EM+ML-based classification method. The proposed classification method uses a mixture of training samples selected from both domains, thus allowing the classifier to be adopted to the target domain data significantly with more direct target-domain-oriented training samples. Tables X and XI mention the overall classification accuracies of the proposed classification system for the common set of land-cover classes shared by both domains and the Overall target domain classification performance as given by the proposed EM+ML classifier scheme for Hyper 1–6.

Kernel k-means has been able to detect all of the clusters in each of the experiments (Hyper 1–12). However, the generalization performance of kernel k-means is not so impressive in distinguishing the overlapping data points (overall accuracy ranges between 67.43% and 93.82% according to Table XI). ML classifier trained on the training samples of Y consistently

TABLE XI
AVERAGE OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) COMPUTED ON ALL OF THE TARGET DOMAIN CLASSES OVER FIVE ITERATIONS, HYPERSPECTRAL EXPERIMENTS 1–6

Experiment	Kernel k-means	Proposed method	Supervised ML of target domain
Hyper 1 $Y5Z5-1$	93.82	99.22	100.00
Hyper 2 $Y5Z6-2$	91.63	99.03	100.00
Hyper 3 $Y5Z7-3$	87.53	97.50	99.44
Hyper 4 $Y5Z8-4$	81.72	97.53	99.25
Hyper 5 $Y5Z9-5$	72.62	97.96	99.32
Hyper 6 $Y9Z9-9$	72.62	98.41	99.32

TABLE XII
AVERAGE OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) COMPUTED ON THE COMMON SET OF TARGET DOMAIN CLASSES OVER FIVE ITERATIONS, HYPERSPECTRAL EXPERIMENTS 7–12

Experiment	Kernel k-means	ML with source domain based training	EM+ML classifier of [6]	Proposed method
Hyper 7 $Z5Y5-1$	98.24	100.00	100.00	100.00
Hyper 8 $Z5Y6-2$	93.14	100.00	100.00	100.00
Hyper 9 $Z5Y7-3$	91.75	86.01	84.17	94.90
Hyper 10 $Z5Y8-4$	85.86	78.85	79.30	88.54
Hyper 11 $Z5Y9-5$	87.24	82.85	83.44	92.98
Hyper 12 $Z9Y9-9$	74.40	75.28	76.66	91.53

TABLE XIII
AVERAGE OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) COMPUTED ON ALL OF THE TARGET DOMAIN CLASSES OVER FIVE ITERATIONS, HYPERSPECTRAL EXPERIMENTS 7–12

Experiment	Kernel k-means	Proposed method	Supervised ML of target domain
Hyper 7 $Z5Y5-1$	86.05	93.94	100.00
Hyper 8 $Z5Y6-2$	81.43	93.73	98.21
Hyper 9 $Z5Y7-3$	77.29	90.01	98.06
Hyper 10 $Z5Y8-4$	73.83	87.71	96.39
Hyper 11 $Z5Y9-5$	74.40	90.58	96.48
Hyper 12 $Z9Y9-9$	74.40	91.53	96.48

exhibits poor generalization accuracy (72.11%–60.85% for experiment 12–16) for classifying the common set of classes, which is quite expected given that the statistical properties of the classes of Y and Z differ substantially. Application of the EM algorithm enhances the performance of the ML classifier by 1%–15% for Hyper 4–6. However, for Hyper 2 and 3, EM algorithm degrades the performance of the ML classifier by 0.08%–0.51%. One possible reason for such reduction in performance may be that EM gets stuck in some local optima in the feature space and thus is unable to enhance the classifier performance substantially.

For Hyper 5 and 6, the target domain contains nine land-cover classes. In Hyper 5, the source domain contains five classes, whereas it contains nine classes in Hyper 6. The performance of kernel k-means is almost identical in both cases (72.45% and 72.62%) in classifying the common set of classes, which signifies that kernel k-means detects the four target domain classes of Hyper 6 with high accuracy. However, the proposed classifier shows an extremely impressive performance in both cases (97.34% and 98.41%) in classifying the common set of classes present in the target domain, which is superior than the other classifiers considered for comparative study.

It can be observed from Table XI that the performance of the proposed classifier (97.53%–99.22%) is very close to the performance of an ML classifier trained solely based on the reliable training samples from Z (99.25% to 100%) for classifying the entire target domain image.

The performances of the classifier for Hyper 7–12 are depicted in Tables XII and XIII, respectively. Similar trends of Hyper 1–6 are also followed here. The proposed classifier system (91.53%–94.90% for Hyper 9–12) outperforms the ML classifier trained on the training samples of Z (75.28%–86.01%) and the ML classifier with EM-based retraining (76.66%–84.17%) for the common set of land-cover classes. The overall target domain generalization performance of the proposed classifier system (87.71%–93.94%) is consistent and is comparable to the performance of a supervised ML classifier training in Y (96.39%– 100%).

V. Conclusion

A novel unsupervised DA technique for the land-cover map updating of remote sensing image pairs has been proposed here. The proposed technique first clusters the target domain image optimally by exploiting the available source domain training data. A graph theoretic technique has been proposed next, which can efficiently identify all of the target domain clusters having correspondences to the identical source domain classes. The proposed cluster mapping method is nonparametric and demonstrated to be effective in highlighting new classes that appeared in the target domain image as well as the disappeared classes from the source domain. The proposed method is robust as it does not have assumptions regarding the number and properties of the target domain land-cover classes. A postprocessing method to refine further the classification map produced by the clustering algorithm based on the ML classifier and the EM retraining has been followed henceforth. The classifier has been selected since a DA paradigm has been developed for it to handle the situation in which the target and source domains do not share all of the classes. However, any other classifier designed to deal with such a situation can be used. The experimental results demonstrate the robustness of the proposed framework in addition/deletion of maximum of two classes. The algorithm has been demonstrated to be scalable, i.e., able to handle more complex situations involving simultaneous addition/deletion of classes. Even if testing has been carried out on multitemporal data, the method can be applied without any restriction to any DA problem. The current mode of research is to explore the possibility of overclustering as a measure which can further simplify the proposed framework. The properties of some gradient-based nonparametric kernel clustering techniques, like mean-shift, are being studied now for the same. Furthermore, it is expected that inclusion of some other pixel level features like texture, contextual information, etc., in addition to the spectral values may enhance the performance of each step of the proposed framework considerably.

Acknowledgment

The authors would like to thank Prof. M. Crawford of Purdue University, West Lafayette, IN, USA, for sharing the hyperspectral data set. This work was carried out under the framework of India Trento Program for Advance Research (ITPAR) during a stay of B. Banerjee at the University of Trento, Trento, Italy.

References

[1] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*. New York, NY, USA: Wiley, 2009, vol. 26.

[2] H. Daumè III and D. Marcu, "Domain adaptation for statistical classifiers" *J. Artif. Intell. Res.*, vol. 26, no. 1, pp. 101–126, May 2006.

[3] J. Jiang, "A literature survey on domain adaptation of statistical classifiers," 2008. [Online]. Available: http://sifaka.cs.uiuc.edu/jiang4/domainadaptation/survey

[4] 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.

[5] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A DASVM classification technique and a circular validation strategy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 770–787, May 2010.

[6] 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.

[7] D. Tuia, J. Munoz-Mari, L. Gomez-Chova, and J. Malo, "Graph matching for adaptation in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 329–341, Jan. 2013.

[8] D. Tuia, E. Pasolli, and W. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, Sep. 2011.

[9] C. Persello and L. Bruzzone, "A novel active learning strategy for domain adaptation in the classification of remote sensing images," in *Proc. IEEE IGARSS*, 2011, pp. 3720–3723.

[10] B. Demir, F. Bovolo, and L. Bruzzone, "Classification of time series of multispectral images with limited training data," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3219–3233, Aug. 2013.

[11] G. Jun and J. Ghosh, "Spatially adaptive classification of land cover with remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2662–2673, Jul. 2011.

[12] 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.

[13] K. Bahirat, F. Bovolo, L. Bruzzone, and S. Chaudhuri, "A novel domain adaptation Bayesian classifier for updating land-cover maps with class differences in source and target domains," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2810–2826, Jul. 2012.

[14] L. J. Grady, "Space-variant computer vision: A graph-theoretic approach," Ph.D. dissertation, Boston Univ., Boston, MA, USA, 2004.

[15] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento, "A (sub) graph isomorphism algorithm for matching large graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 10, pp. 1367–1372, Oct. 2004.

[16] M. Volpi, D. Tuia, G. Camps-Valls, and M. Kanevski, "Unsupervised change detection by kernel clustering," in *Proc. Int. Soc. Opt. Photon. Remote Sens.*, 2010, vol. 7830, pp. 1–8.

[17] J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, pp. 487–493.

[18] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty years of graph matching in pattern recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 18, no. 3, pp. 265–298, May 2004.

[19] R. Zhang and A. I. Rudnicky, "A large scale clustering scheme for kernel k-means," in *Proc. 16th Int. Conf. Pattern Recognit.*, 2002, vol. 4, pp. 289–292.

[20] I. Dhillon, Y. Guan, and B. Kulis, "A Unified View of Kernel k-Means, Spectral Clustering and Graph Cuts," Dept. Comput. Sci., Univ. Texas, Austin, TX, USA, 2004.

[21] M.-S. Yang, "A survey of fuzzy clustering," *Math. Comput. Model.*, vol. 18, no. 11, pp. 1–16, Dec. 1993.

[22] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. Hoboken, NJ, USA: Wiley, 2009, vol. 383.

[23] J. He, Y. Liu, and R. Lawrence, "Graph-based transfer learning," in *Proc. 18th ACM Conf. Inf. Knowl. Manag.*, 2009, pp. 937–946.

[24] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1858–1860, Jul. 2003.

[25] L. A. Zager and G. C. Verghese, "Graph similarity scoring and matching," *Appl. Math. Lett.*, vol. 21, no. 1, pp. 86–94, Jan. 2008.

[26] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Hoboken, NJ, USA: Wiley, 1974.

[27] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

[28] S. Petrovic, "A comparison between the Silhouette index and the Davies–Bouldin index in labelling IDS clusters," in *Proc. 11th Nordic Workshop Secure IT Syst.*, 2006, pp. 53–64.

[29] S. Samanta and S. Das, "Inter-domain cluster mapping and GMCV based transformation for domain adaptation," in *Proc. Pattern Recognit. Mach. Intell.*, 2013, pp. 74–81.

[30] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, Apr. 2008.

[31] L. Bruzzone and C. Persello, "A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3180–3191, Sep. 2009.

Biplab Banerjee received the M.E. degree in computer science from Jadavpur University, Kolkata, India.

He is currently a Research Scholar with the Satellite Image Analysis Laboratory, CSRE, IIT Bombay, Mumbai, India. He completed a three-month research stay at the Signal and Image Processing Department, Telecom ParisTech, France, from April 2012 to June 2012. He was a Visiting Researcher with the Remote Sensing Laboratory, University of Trento, Trento, Italy, from September 2013 to December 2013. His research interests include image processing, computer vision, machine learning, and algorithm analysis and design.

Francesca Bovolo (S'05–M'07–SM'13) received the Laurea (B.S.) degree, the Laurea Specialistica (M.S.) degree in telecommunication engineering (*summa cum laude*), and the Ph.D. degree in communication and information technologies from the University of Trento, Trento, Italy, in 2001, 2003, and 2006, respectively.

She was a Research Fellow with the University of Trento until June 2013. She is the Founder and the Head of the Remote Sensing for Digital Earth Unit, Fondazione Bruno Kessler (FBK), Trento, and a member of the Remote Sensing Laboratory (RSLab), Trento. She is a referee for several international journals. She is one of the Coinvestigators of the Radar for Icy Moon Exploration (RIME) instrument of the ESA JUpiter ICy moons Explorer (JUICE). Her main research activity is in the area of remote sensing image processing. Her interests are related to multitemporal remote sensing image analysis; change detection in multispectral, hyperspectral, and synthetic aperture radar images and very high resolution images; content-based time series retrieval; and domain adaptation. She conducts research on these topics within the context of several national and international projects.

Dr. Bovolo received the First Place in the Student Prize Paper Competition of the 2006 IEEE International Geoscience and Remote Sensing Symposium (Denver, August 2006). Since January 2011, she has been an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She has been a Guest Editor for the Special Issue on Analysis of Multitemporal Remote Sensing Data of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. She is the Technical Chair of the Sixth International Workshop on the Analysis of Multi-temporal Remote-Sensing Images (MultiTemp 2011). From 2006 to 2013, she was on the Scientific Committee of the SPIE International Conference on Signal and Image Processing for Remote Sensing. Since 2014, she has been the Cochair of the same conference. She is the Publication Chair of the International Geoscience and Remote Sensing Symposium 2015. Since 2012, she has been a member of the International Program Committee of the International Conference on Pattern Recognition Applications and Methods. She was on the Scientific Committee of the IEEE Fourth and Fifth International Workshop on the Analysis of Multi-Temporal Remote Sensing Images (MultiTemp 2007 and 2009) and of the IEEE GOLD Remote Sensing Conference in 2010, 2012, and 2014.

Avik Bhattacharya (M'08) received the M.Sc. degree in mathematics from the Indian Institute of Technology, Kharagpur, India, and the Ph.D. degree from the Telecom ParisTech, Paris, France, and the Ariana Research Group, Institut National de Recherche en Informatique et en Automatique (INRIA), Sophia Antipolis, Nice, France.

He is currently an Assistant Professor with the Centre of Studies In Resources Engineering, Indian Institute of Technology Bombay (IITB), Mumbai, India. Prior to joining IITB, he was a Canadian Government Research Fellow with the Canadian Centre for Remote Sensing (CCRS), Ottawa, ON, Canada. He received the Natural Sciences and Engineering Research Council of Canada Visiting Scientist Fellowship at the Canadian National Laboratories from 2008 to 2011. His current research includes synthetic aperture radar (SAR) polarimetric application to cryosphere and planetary exploration, statistical analysis of polarimetric SAR images, machine learning, and pattern recognition.

Lorenzo Bruzzone (S'95–M'98–SM'03–F'10) received the Laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Genova, Italy, in 1993 and 1998, respectively.

He is currently a Full Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, radar, pattern recognition, and electrical communications. He is the Founder and the Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. He is the author (or coauthor) of 161 scientific publications in referred international journals (111 in IEEE journals), more than 220 papers in conference proceedings, and 17 book chapters. He is the editor/coeditor of 15 books/conference proceedings and 1 scientific book. His papers are highly cited, as proven by the total number of citations (more than 11 200) and the value of the h-index (55; source: Google Scholar). His current research interests are in the areas of remote sensing, radar and synthetic aperture radar, signal processing, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. Among the others, he is the Principal Investigator of the Radar for Icy Moon Exploration (RIME) instrument in the framework of the JUpiter ICy moons Explorer (JUICE) mission of the European Space Agency.

Dr. Bruzzone has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society since 2009. He was invited as a keynote speaker in 24 international conferences and workshops. He was a guest coeditor of different special issues of international journals. He is the Cofounder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. Since 2013, he has been the founder Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE. He is currently an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the *Canadian Journal of Remote Sensing*. Since 2012, he has been appointed as a Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society. He received the First Place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). Since that time, he has been the recipient of many international and national honors and awards.

Subhasis Chaudhuri (M'99–SM'02–F'11) received the B.Tech degree from the Indian Institute of Technology, Kharagpur, India, in 1985, the M.Sc. degree from the University of Calgary, Calgary, AB, Canada, in 1987, and the Ph.D. degree from the University of California San Diego, La Jolla, CA, USA, in 1990.

He is currently the KN Bajaj Chair Professor with the Department of Electrical Engineering and the Deputy Director of the Indian Institute of Technology Bombay, Mumbai, India, where he was the Head of the department in 2005–2008 and the Dean (International Relations) in 2009–2013. His primary research interests include pattern recognition, image processing, computer vision, and haptics.

Dr. Chaudhuri is a Fellow of the science and engineering academies in India. He was the recipient of the Bhatnagar Prize in Engineering Sciences in 2004 and the GD Birla Award in 2011.

Krishna Mohan Buddhiraju (M'12) received the Ph.D. degree in electrical engineering from the Indian Institute of Technology Bombay, Mumbai, India, in 1991, specializing in digital image processing and analysis.

He is currently a Professor and the present Head of the Centre of Studies in Resources Engineering, IIT Bombay. He is currently supervising six Ph.D. students, and he has over 100 publications in journals, book chapters, and conference proceedings. His current interests are high-resolution remote sensing, machine learning, and multimedia educational content development for satellite image processing.

Dr. Buddhiraju is a Life Member of the Indian Society of Remote Sensing and Indian Society of Geomatics. He was the recipient of the Indian Society of Remote Sensing's National Geospatial Award for Excellence 2012 and IETE's M.N. Saha Memorial Award for Best Application Oriented Paper published in *IETE Journal of Research* in 2000.