

© © 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Title: A novel graph matching based approach for domain adaptation in classification of remote sensing image pair

This paper appears in: IEEE Transactions on Geoscience and Remote Sensing

Date of Publication: 09 February 2015

Author(s): B. Banerjee, F. Bovolo, A. Bhattacharya, L. Bruzzone, S. Chaudhuri, B.K. Mohan

Volume: 53, Issue: 7

Page(s): 4045 - 4062

DOI: 10.1109/TGRS.2015.2389520

A novel graph matching based approach for domain adaptation in classification of remote sensing image pair

Biplab Banerjee, Francesca Bovolo, Avik Bhattacharya, Lorenzo Bruzzone, Subhasis Chaudhuri, B Krishna Mohan

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 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 the domains as well as the additional or missing classes in the target domain image. An inter-domain graph is introduced which contains all the class information of both the 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 post-processing step is applied to refine the classification map produced by the clustering algorithm. Two multi-spectral datasets with medium and very high geometrical resolution and one hyper-spectral dataset are considered to evaluate the robustness of the proposed technique. Two of the datasets consist of multi-temporal 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

Domain Adaptation, Graph Matching, Cross-Domain Graph, Clustering.

I. INTRODUCTION

IN supervised classification, one basic assumption is that the labeled training data are generated from the same probability density function 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] [3] [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 datasets 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 probability density functions (PDF) 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 the 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 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 the domains. A simple and scalable solution has been presented there by focusing on the description of the changes in the manifold by defining a non-linear transformation based on vector quantization and graph matching. In the recent past, several active learning based classification techniques have been proposed for domain adaptation in the remote sensing literature for both multi-spectral and hyper-spectral data [8] [9] [10]. Some adaptation method specific to the hyper-spectral data have been introduced in [11] [12]. Large margin classifiers like Support Vector Machines (SVM) have been modified for the cross domain data classification purpose [5]. A Domain Adaptation 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 multi-temporal images. A Change Vector Analysis (CVA) based method has been used to identify the class-wise set of changed/unchanged pixels from the images of both the domains. The changed pixels are further analyzed to investigate the existence of some new target domain classes based on statistical divergence measures.

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 multi-temporal images. When adaptation is to be applied on images acquired over spatially disjoint locations, temporal correlation can not 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.

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

To deal with the above mentioned 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 is 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 (MCS) [15] given the source and the target domain graphs followed by identifying the similar matching nodes from both the 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 above assertion, in this paper, an unsupervised domain adaptation 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 inter domain 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 Maximum-Likelihood (ML) classifier combined with an Expectation-Maximization (EM) based retraining scheme is employed since it demonstrated to be robust when the statistical distribution of the data points are 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 class statistical parameters required for the refinement step is performed by blending data from both the domains. The adaptation of the statistical parameters of the classes common to both the 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.

The rest of the 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 the paper and discusses relevant possible future research directions.

II. AN INSIGHT TO THE CLASSIFICATION PROBLEM OF A PAIR OF IMAGES IN THE CONTEXT OF DOMAIN ADAPTATION

Let I_S and I_T denote the source and the target domain multi-spectral 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 multi-temporal 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 $\Omega_S = \{\omega_1, \omega_2, \dots, \omega_N\}$ 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 $TR_S = \{(x_{SI}, \omega_i)\}$, ($x_{SI} \in \mathbb{R}^B$, $1 \leq i \leq N$) where x_{SI} is the i^{th} 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 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 class-wise 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 , represent 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 .

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 set of classes. The following cases can be identified in this regard:

- **CASE A** The source and the target domain contain identical set of land-cover classes, $\Omega_S = \Omega_T$.
- **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 .
- **CASE C** There are less target domain classes than source domain classes, i.e. $\Omega_T = \Omega_S \setminus \{\omega_k\}$.
- **CASE D** k source domain classes disappear in the target domain and l new classes appear in that place, $\Omega_T = \Omega_S \cup \{\omega_l\} \setminus \{\omega_k\}$ with l either equal to or different from k .

The proposed adaptation method handles both the differences in terms of classes (new and disappeared ones) as well as those in terms of the statistical properties of the common classes. The adaptation procedure is based on EM algorithm and is able to handle the aforementioned cases. If the classes common to both the 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.

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

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.

III. THE PROPOSED TECHNIQUE

In order to cope with the problem exposed in Section II, a three-step unsupervised DA approach is proposed based on:

- **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 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.
- **Cross domain cluster mapping:** This is an important step and the main contribution of the proposed algorithm where all 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].
- **Definition of the domain adaptation classifier for post-processing:** Under the specific hypothesis formulated in the above two steps, this step performs the adaptation of the classifier built primarily for I_S to the properties of I_T . Here a Maximum-Likelihood (ML) classifier is used and further retrained by the iterative Expectation-Maximization (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 are also identified in this step. This step can also be considered as the post-processing step which refines the land-cover map produced by the clustering technique.

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

A. Definition of the 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).

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.

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 non-linearly 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 parameters 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 (RBF) kernel is used along with Kernel k-means. The kernel hyper-parameter (γ) is selected by grid search technique.

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 i^{th} source domain class and the j^{th} target domain cluster, respectively. Both the distributions are usually approximated by histogram based approaches [22]. Thus, F is defined for a given $k \in [N - n, N + n]$ as:

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

where $KL(p_S(X_S|\omega_i), p_T(X_T|\alpha_j))$ is the non-symmetric KL divergence of the discrete PDF's $p_S(X_S|\omega_i)$ from $p_T(X_T|\alpha_j)$ and $Q(\hat{\Omega}_{Temp}(k))$ is a measure of the quality of the clustering results.

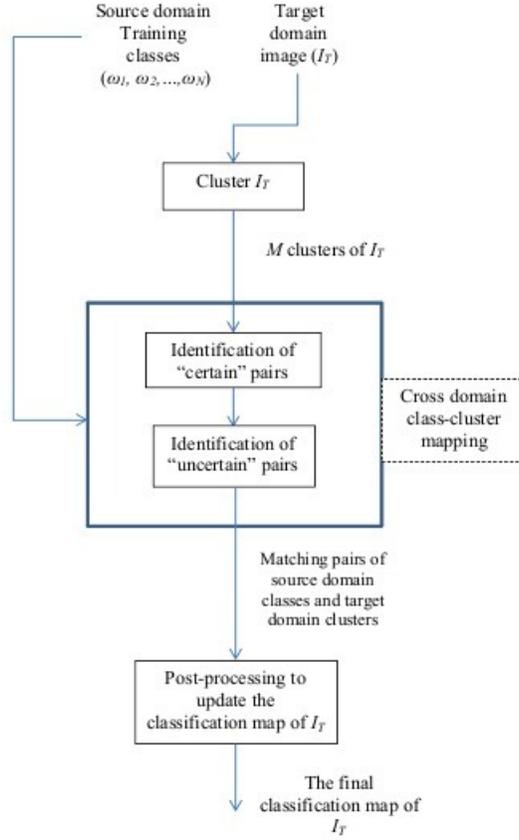


Fig. 1: The flowchart of the proposed method

The KL divergence is computed as

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

$KL(p_T(X_T|\alpha_j), p_S(X_S|\omega_i))$ 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 $p_S(X_S|\omega_i)$. 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, $\sum_i \sum_j KL(p_S(X_S|\omega_i), p_T(X_T|\alpha_j)) - \sum_j \sum_i KL(p_T(X_T|\alpha_j), p_S(X_S|\omega_i)) \rightarrow 0$. 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 term of the information content shared between them.

The $Q(\widehat{\Omega}_{Temp}(k))$ term measures the quality of the clustering results in terms of the sum of inter-cluster variance and the inverse of intra-cluster 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 .

The output of this step are the M clusters $\widehat{\Omega}_T = \{\alpha_1, \alpha_2, \dots, \alpha_M\}$ 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: $\Omega_S, TR_S, I_T, k \in [N - n, N + n]$

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 $\Omega_{Temp}(k) = \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ be the clusters obtained.
 - 3: $Mean_{S \rightarrow T} =$ the mean KL divergence from all the classes in TR_S to the clusters in Ω_{Temp} .
 - 4: $Mean_{T \rightarrow S} =$ the mean KL divergence from all the clusters in Ω_{Temp} to the classes in TR_S .
 - 5: $F(k) = \frac{1}{|Mean_{S \rightarrow T} - Mean_{T \rightarrow S}| + 1} + Q(\widehat{\Omega}_{Temp}(k))$ (Alternatively Equation 1)
 - 6: **end for**
 - 7: $M = \underset{k \in [N - n, N + n]}{\operatorname{argmax}} F(k)$
 - 8: $\widehat{\Omega}_T = \Omega_{Temp}(M) = \{\alpha_1, \alpha_2, \dots, \alpha_M\}$
-

B. Cross domain cluster mapping

This is the most important step and the main contribution of the proposed DA technique where the unique one-to-one mapping between all 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 are replaced by k new classes in I_T , a post-processing check is included in the cluster mapping algorithm to identify all the non-equivalent 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 transfer learning [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 the ω_i 's to the α_j 's and vice versa. Figure 2 depicts a cross-domain graph with 3 source domain classes, 4 target domain clusters and 12 edges.

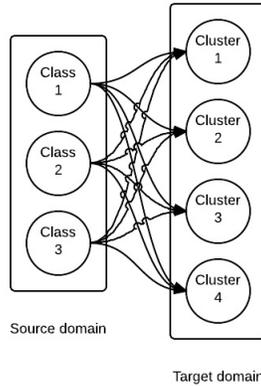


Fig. 2: An example of a cross-domain graph

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

- 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 .
- 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 intra-domain graphs first and then tries to find the maximum common subgraph (MCS) of them. A subtractive iterative algorithm is followed to find the matching pairs from the MCS.

The edge weights (E_W) of G are defined using a symmetric divergence measure depending on the KL-divergence, i.e. $E_W(\alpha_j, \omega_i) = E_W(\omega_i, \alpha_j) = \sqrt{(\text{Min}(\text{KL}(p_T(X_T|\alpha_j), p_{ST}(X_{\alpha_j \cup \omega_i})), \text{KL}(p_S(X_S|\omega_i), p_{ST}(X_{\alpha_j \cup \omega_i}))))}$ 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 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 classes.

Two lists Rank_S and Rank_T are maintained in all the iterations of the cluster mapping algorithm. Rank_S is a $N \times 1$ 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 a $M \times 1$ matrix which denotes the same as of Rank_S but from the target to the source domain.

All 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 Figure 3. These steps are detailed below.

The **identification of the certain pairs of source domain classes and the target domain clusters from G** requires 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 :

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

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

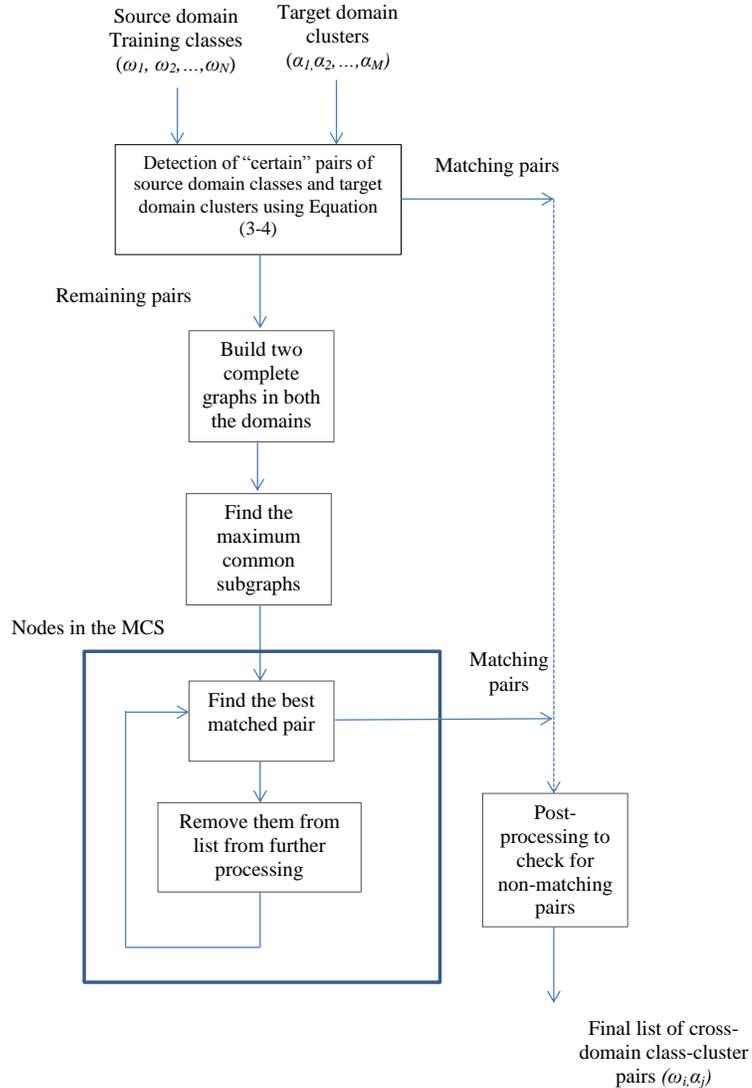


Fig. 3: Flowchart of the cluster mapping algorithm

$\nabla_{S \rightarrow T}$ 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 . $\nabla_{T \rightarrow S}$ 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'_i .

If $\nabla_{S \rightarrow T}(\omega_i) = \alpha_j$ and $\nabla_{T \rightarrow S}(\alpha_j) = \omega_i$ and $|\nabla_{S \rightarrow T}(\omega_i)| = |\nabla_{T \rightarrow S}(\alpha_j)| = 1$ (where $|\cdot|$ denotes the cardinality of a set)

then it is expected that ω_i and α_j represent the same class in both the domains. $|\nabla_{S \rightarrow T}(\omega_i)| = 1$ means that only one target domain node has ω_i as its best matched source domain counterpart. $|\nabla_{T \rightarrow S}(\alpha_j)| = 1$ 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 are considered to represent the identical land-cover class on the ground. It is also possible to find no pairs in this step in some cases if the classes undergo severe changes between the considered images. The set of all 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 non-equivalence 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 the domains separately. The next part of the proposed cluster mapping algorithm investigates the possible correspondences of these nodes. The new target domain specific nodes (CASE B,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 be disappeared in I_T (CASE C, Section II). If it is the case with the target domain subgraph, they represent the new classes appeared in I_T (CASE B).

Two individual intra-domain 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 the graphs. The MCS are found from G_S and G_T according to the following analysis. If $|V_S| \leq |V_T|$ then the MCS's contain $|V_S|$ nodes else they contain $|V_T|$ nodes as both the 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 absolute difference between the cumulative edge weights of the subgraphs and
- The sum of the absolute difference between the corresponding maximum Eigenvalues of the graph adjacency matrix of both the 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 term 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 the subgraphs. The remaining source domain or the target domain nodes, which are not part of the MCS, represent CASE B-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 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 follows the same set of steps until R_S and R_T become empty (see Algorithm 2).

Once all 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

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** $R_S \neq \emptyset$ or $R_T \neq \emptyset$ **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 the 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 the matching pairs and the corresponding JS divergence between the members of each pair obtained in this step are stored for future references.
-

tend to change at every iteration of the cluster mapping algorithm. These two properties are combined to check whether a pair is non-equivalent, i.e., though 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 non-matching.

The set of common land-cover classes which have undergone minimum or no changes between the acquisitions tend 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 term 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 till 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 post-processing stage whereas during the initial iterations, only very similar classes and clusters are mapped.

After this step, all 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

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

- Define a set of training samples per cluster of I_T . For all the matching pairs, an algorithm is proposed to highlight a set of target specific training samples from the information of both the domains. For the new target domain classes, the corresponding cluster information is used.

Algorithm 3 Input: Set of all the matching pairs produced by the cluster mapping stage, the JS divergence of the candidates of each pair and the lists of all 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:** $JS(\omega_i, \alpha_j) \geq v \times m$ $\{m$ is the minimum JS divergence of the pairs produced by the cluster mapping algorithm. v 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 non-matching one. It is dependent on m and is directly proportional to the absolute difference between the average divergence of all 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 a inconsistent pair and hence, removed.
-

- The adaptation of source domain properties to I_T ones is achieved in this step through the iterative EM algorithm.
- The final land-cover map of I_T is generated by a ML classifier with the updated parameter set.

Algorithm 4 describes the steps for defining the set of training samples for all 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 samples selection algorithm over-clusters both the ω_i and α_j of a given matching pair of (ω_i, α_j) . The particular sub-cluster of α_j which is closest to ω_i in term 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:

$$x_{Tl} \in \alpha_k \Leftrightarrow \underset{\alpha_j \in I_T}{\operatorname{argmax}} \{P_T(\alpha_j)p_T(x_{Tl}|\alpha_j)\} \quad (5)$$

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 associate with I_T . It is worth noting that the subscript T has been used here to stress the dependencies of both the 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. Despite in this work classes are assumed to be Gaussian distributed, the same logic can be

applied for any other model.

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 to be applied for adaptation of the statistical parameters. The true statistical terms $\theta_{Tj} = \{\mu_{Tj}, \Sigma_{Tj}, P_T(\alpha_j)\}$ associated with each cluster α_j of I_T are computed as follows

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

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

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

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.

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: $\omega_i = \{\omega_{i1}, \omega_{i2}, \dots, \omega_{il}\}$.
 - 4: $\alpha_j = \{\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{jl}\}$.
 - 5: Mat is a $l \times l$ matrix which stores the pairwise JS divergence of all the clusters of ω_i and α_j . $\{\omega$'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 term of the JS distance measure.
 - 7: The training set for the j^{th} target domain class is defined as $\alpha_j^{\text{Tr}} = (\omega_i \cup \alpha_{jm})$.
 - 8: **end for**
-

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

- **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 the domains. In this situation, the initialization of the statistical terms are based on the training set obtained by Algorithm 4. Let θ_T^{New} denotes 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 followed hence after.
- **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 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

$$\theta_T^0 = \theta_T^{\text{New}} \cup \{\mu_u^0, \Sigma_u^0, P_T^0(\alpha_u)\} \quad (9)$$

The prior probabilities are scaled based on the newly defined training set for I_T .

- **CASE C:** $\widehat{\Omega}_T = \Omega_S \setminus \{\omega_u\}$: This is the case where some classes are disappeared between between I_S and I_T . For all 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

$$\theta_T^0 = \theta_T^{\text{New}} \setminus \{\mu_u^0, \Sigma_u^0, P_T^0(\omega_u)\} \quad (10)$$

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

$$\theta_T^0 = \theta_T^{\text{New}} \cup \{\mu_k^0, \Sigma_k^0, P_T^0(\alpha_k)\} \setminus \{\mu_u^0, \Sigma_u^0, P_T^0(\omega_u)\} \quad (11)$$

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 multi-spectral and one hyper-spectral datasets were conducted. Experiments have been designed at three different levels of the proposed framework as follows:

- 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 maximum Silhouette measure within $[N - n, N + n]$. The γ value obtained for $k = N - n$ has been used for the remaining cases. $n = \lceil \frac{N}{2} \rceil$ has been considered for all the datasets. It has been observed that kernel k-means along with the γ found in this way has produced good clustering results for all 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.

- The cross-domain cluster mapping step outputs pairs of the similar classes shared by both the 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 Co-Variances (GMCV), estimated from the covariance matrices of the data from both the domains. It further assumes that 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.
- Two different strategies have been followed to evaluate the performance of the proposed EM+ML based classifier system in producing reliable land-cover map the target domain data.

- 1) 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 a 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.
- 2) 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 training set associated with the image considered as the source were 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 single image in both the 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 multi-temporal images, the spatial correlation between the images is of no use for the proposed technique considering that the algorithm works only with spectral features. 5 independent realizations of all the experiments have been conducted and the average results are reported here.

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

B. Medium resolution Sardinia dataset

The first study area consists of two co-registered multi-spectral images acquired by the Thematic Mapper (TM) sensor of the Landsat-5 satellite. The images consist of 6 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. Figure 4(a-b) depicts band 4 of both the images. The images share 5 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. Figure 4(c) shows the band 4 of September, 1995 image with the Burned area class (Y'). Table II lists the class-wise distributions of training and test samples of all the images used for experimental purposes. All the images have been used alternatively as source and target domain. Figure 5 and 6 depict the scatter plot of the test sets of (Z) and (Y') of band 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 significant shift while remaining classes are stable in terms of the statistical properties.

TABLE II: Cardinality ($|\cdot|$) of training (Tr) and test (Ts) sets per class available for Sardinia dataset

Land Cover	Z (July '96)		Y' (September '95)		Y (September '95)	
	$ 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	-	-

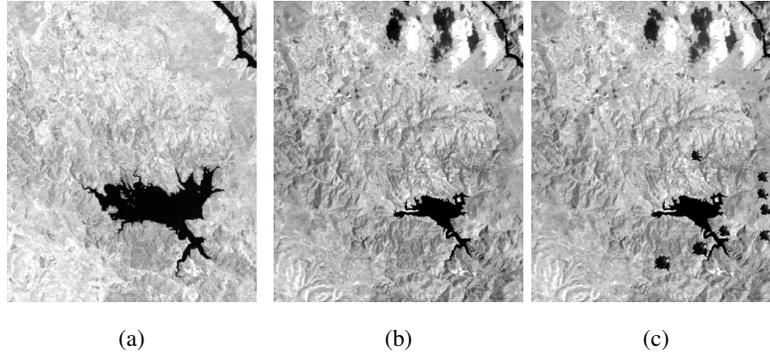


Fig. 4: The band 4 of the (a) July 96 image (Z) (b) September 95 image (Y) (c) The simulated September 95 image with Burned area (Y')

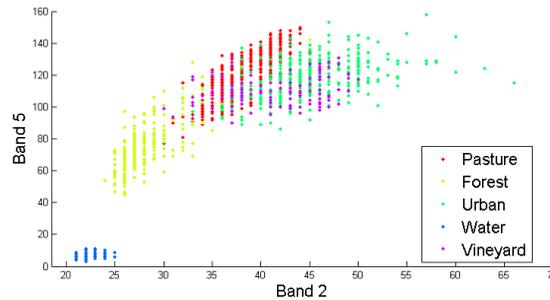


Fig. 5: The scatter plots of the bands 2 and 5 of the July 96 image (Z)

9 experimental cases have been considered for the Sardinia dataset 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 5 classes each with one common class shared between them.

- 1) **Sardinia 1** $Z3Y4 - 1$: Z and Y have 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, 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 5 classes have 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 5 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. In Y' , the Burned area class has been added while the Vineyard class has not been considered. Hence, though same number of land-cover classes are present in both the domains, but one of the pairs is non-equivalent. $\Omega_T = \Omega_S \cup \{\text{Burned - area}\} \setminus \{\text{Vineyard}\}$.
- 5) **Sardinia 5** $Z4Y'6 - 4$: Z and Y' have been used as the source and the target domain images. 4 classes have been considered in Z excluding the Urban class. In Y' , 6 classes have been considered. In addition to the 4 classes of Z , Burned area and Urban have been included in Y' , i.e., $\Omega_T = \Omega_S \cup \{\text{Burned - area}, \text{Urban}\}$.
- 6) **Sardinia 6** $Y3Z3 - 1$: Y and Z have been selected to be the source and the target domain images respectively. Pasture,

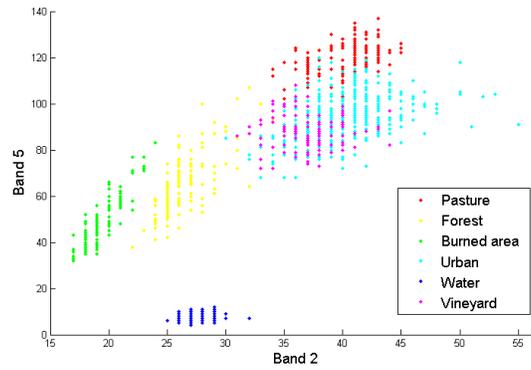


Fig. 6: The scatter plots of the bands 2 and 5 of the simulated September 95 image with Burned area (Y')

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 5 classes have been considered for both the 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 5 common classes. i.e., $\Omega_T = \Omega_S \setminus \{\text{Burned - area}\}$. 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. 4 classes have been considered in Z excluding the Urban class. In Y' , 6 classes have been considered, in addition to the 4 classes of Z , Burned area and Urban have been included in Y' , i.e., $\Omega_T = \Omega_S \setminus \{\text{Burned - area, Urban}\}$.

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

The initial clustering step of the proposed method correctly found out the optimal target domain clustering in all 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 (Equation 1) in all the cases. A typical γ value obtained for Sardinia 1 was 0.00075 which has further been used for all the remaining experimental cases. v has been set to an average value of 3.75 to compensate for the high variability of the divergence measures for all the matching pairs of the source domain classes and target domain clusters found in the cross-domain cluster mapping stage for all the experiments conducted.

In the cross domain cluster mapping stage, the Water and Forest classes are easy to map as one-to-one mapping exist in both the cases with respect to the JS divergence. Unique injective mappings are difficult to obtain for the Pasture, Vineyard and the Urban classes due to high overlapping. The inter-domain 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 the 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.

Table III and IV show the comparisons of classification accuracy of the proposed method for the experimental cases mentioned above 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 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 the acquisition of the respective images. It is reflected in the classification accuracy 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% to 90.11% for Sardinia 2-5 and 7-9). The ML classifier trained on the source domain performs poorly in all the Experimental cases with a best performance of 67.06% for Sardinia 7-8 in classifying the common set of land-cover classes shared between the domains. A sharp enhancement in term of the classification accuracy is observed when EM retraining is used along with ML classifier (79.52% to 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 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% to 99.19%) is far better than the Kernel k-means based classification results (73.98% to 98.47%) and is very close to the results of a supervised ML classifier trained on the reliable target domain training samples (92.66% to 99.43%).

TABLE III: Average overall classification accuracies (in %) computed on the target domain for the common set of classes over 5 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 %) computed on all the target domain classes over 5 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

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 class 1-4) classes have been selected whereas Pasture, Forest, Burned area and Water (Target class 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 the 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, 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 the domains using the remaining classes independently and the MCS of both the 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 the Class 2 of both the domains is 0.72 which is also the minimum JS divergence between all the pairs. On the other hand, the distance is 3.09 between Class 3 of both the domains which is greater than 3.75×0.72 (Algorithm 3). It indicates that Class 3 of both the domains are not pointing towards 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 post-processing step.

C. Very High Resolution QuickBird dataset

The second data set is made up of two co-registered and pan-sharpened multi-spectral Very High geometrical Resolution (VHR) images acquired by the QuickBird satellite. All the 4 spectral bands of QuickBird have 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 have been used to build a training set and a test set for each image. Four main land-cover classes common to both the 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. The Band 1 of both the images are shown in Figure 7. Table VI lists the class-wise number of training and test samples for both the images used for the experiments along with the JS divergences between the pairs of common training classes. It is evident from the class-wise JS divergences of the training samples of both the images that all the classes have undergone 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 the images, the minimum divergence (m) is comparatively higher while the absolute difference between the average and the minimum divergence of the matching pairs is lesser in comparison to the Sardinia dataset. Accordingly, v has been set to an average value of 2.

The classes have undergone 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 as follows with the same naming convention as of the experimental cases of the Sardinia dataset 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 3 land-cover classes common to them, i.e. Water-body, Red-roof and Asphalt. $\Omega_S = |\Omega_T| \cup \{\text{Field}\}$.

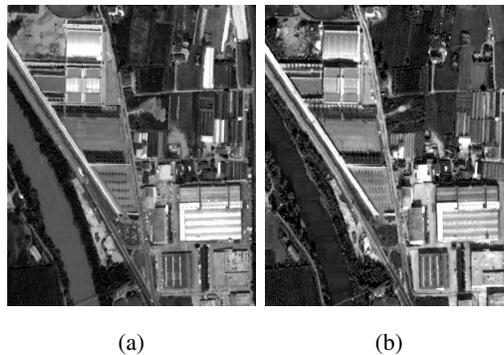


Fig. 7: The band 4 of the (a) July 2006 image (Z) (b) October 2005 image (Y)

TABLE VI: Cardinality ($|\cdot|$) of training (T_r) and test (T_s) sets per class available for QuickBird dataset

Land Cover	Oct. 2005' image (Y)		July 2006' image (Z)		JS divergence between training samples
	$ T_r $	$ T_s $	$ T_r $	$ T_s $	
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

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 the 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.

Table 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 the classes are almost properly detected by all the techniques (classification accuracy $\geq 99\%$) except the ML classifier trained on the source domain samples (49% to 50%). The Water-body class has undergone 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 by incorporating the target domain samples in the classification process.

TABLE VII: Average overall classification accuracies (in %) computed on the common set of target domain classes over 5 iterations, QuickBird experiments 1-2

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 %) computed on all the target domain classes over 5 iterations, QuickBird experiments 1-2

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

D. Hyper-spectral dataset

The third study area considered is represented by a pair of hyper-spectral 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 30m thus it covers a large strip of $7.7 \times 44.3km^2$ 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 dataset can be obtained in [30]. 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 post-processing for land-cover map refinement. Thus a subset of 9 classes have been considered such that the clustering works at least to a certain extent. These 9 classes have initially been divided between the domains with each domain containing 5 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 the 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 non-discriminant 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 the areas are 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 divergence of 1.59, 1.60 and 1.52 and 1.60 respectively), while the remaining classes are more stable.

TABLE IX: Cardinality ($|\cdot|$) of training (Tr) and test (Ts) sets per class available for Hyper-spectral dataset

Land Cover	Area 1 (Y)		Area 2 (Z)		JS div. for training data
	$ Tr $	$ Ts $	$ Tr $	$ Ts $	
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

Two distinct sets of experiments totaling 12 have been performed for the hyper-spectral dataset considering Y and Z

alternatively the as source and target domains. Initially, 5 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 the domain share a common set of 9 aforementioned land-cover classes (Hyper 7 and 12).

For Hyper 1, the target domain data has been found to be clustered into 5 classes while maximizing F (Equation 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 has been clustered into 5 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. v has been set to a mean value of 3.75 similar to the Sardinia dataset. Only a valid one-to-one mapping has been obtained for the Water class which has also been represented by the pair with smallest JS divergence (0.76) between all the corresponding source domain classes and the target domain clusters. For the remaining set of pairs, high JS divergence between the corresponding members (> 2.80) has confirmed the absence of any further matching pairs. For Hyper 1, [29] has performed 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 kernel k-mean used in the proposed setup as indicated by the JS divergence measures in both the cases.

Hyper 2-5 have been carried out in 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 over-lapping 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 have successfully performed optimal target domain clustering followed the cross-domain cluster mapping without any false matching.

Once the common set of classes and clusters are 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 the domains, thus, allowing the classifier to be adopted to the target domain data significantly with more direct target domain oriented training samples. Table X and XI mention the overall classification accuracy of the proposed classification system for the common set of land-cover classes shared by both the 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 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% to 93.82% according to Table XI). ML classifier trained on the training samples of Y consistently exhibits poor generalization accuracy (72.11% to 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% to .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 9 land-cover classes. In Hyper 5, the source domain contains 5 classes whereas it contains 9 classes in Hyper 6. The performance of kernel k-means is almost identical in both the cases (72.45% and (72.62%) in classifying the common set of classes which signifies that kernel k-means detects the 4 target domain classes of Hyper 6 with high accuracy. However, the proposed classifier shows extremely impressive performance in both the 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% to 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 Table XII and XIII respectively. Similar trends of Hyper 1-6 are also followed here. The proposed classifier system (91.53% to 94.90% for Hyper 9-12) outperforms the ML classifier trained on the the training samples of Z (75.28% to 86.01%) and the ML classifier with EM based re-training (76.66% to 84.17%) for the common set of land-cover classes. The overall target domain generalization performance of the proposed classifier system (87.71% to 93.94%) is consistent and is comparable to the performance of a supervised ML classifier training in Y (96.39% to 100%).

TABLE X: Average overall classification accuracies (in %) computed on the common set of target domain classes over 5 iterations, Hyper-spectral 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

TABLE XI: Average overall classification accuracies (in %) computed on all the target domain classes over 5 iterations, Hyper-spectral 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

V. CONCLUSION

A novel unsupervised domain adaptation 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 the target domain

TABLE XII: Average overall classification accuracies (in %) computed on the common set of target domain classes over 5 iterations, Hyper-spectral 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 %) computed on all the target domain classes over 5 iterations, Hyper-spectral 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

clusters having correspondences to the identical source domain classes. The proposed cluster mapping method is non parametric and demonstrated to be effective in highlighting new classes 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 the properties of the target domain land-cover classes. A post processing 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 able to handle the situation in which target and source domain do not share all the classes. However any other classifier designed to deal with such a situation can be used. Experimental results demonstrate the robustness of the proposed framework in addition/deletion of maximum of two classes. The algorithm demonstrated to be scalable, i.e., able to handle more complex situations involving simultaneous addition/deletion of classes. Despite testing has been carried out on multi-temporal data, the method can be applied without any restriction to any domain adaptation problem. The current mode of research is to explore the possibility of over-clustering as a measure which can further simplify the proposed framework. The properties of some gradient based non-parametric kernel clustering techniques like Mean-Shift (MS) 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 work was carried out under the framework of India Trento Program for Advance Research (ITPAR) during a stay of Biplob Banerjee at the University of Trento (Italy). The authors are grateful to Prof. Melba Crawford of Purdue University for sharing the hyper-spectral dataset.

REFERENCES

- [1] G. Camps-Valls and L. Bruzzone, "Kernel methods for remote sensing data analysis," vol. 26, 2009.
- [2] H. Daumé III and D. Marcu, "Domain adaptation for statistical classifiers." *J. Artif. Intell. Res.(JAIR)*, vol. 26, pp. 101–126, 2006.
- [3] J. Jiang, "A literature survey on domain adaptation of statistical classifiers," URL: <http://sifaka.cs.uiuc.edu/jiang4/domainadaptation/survey>, 2008.
- [4] S. J. Pan and Q. Yang, "A survey on transfer learning," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [5] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A dasvm classification technique and a circular validation strategy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 5, pp. 770–787, 2010.
- [6] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 39, no. 2, pp. 456–460, 2001.
- [7] D. Tuia, J. Munoz-Mari, L. Gomez-Chova, and J. Malo, "Graph matching for adaptation in remote sensing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 1, pp. 329–341, 2013.
- [8] D. Tuia, E. Pasolli, and W. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sensing of Environment*, vol. 115, no. 9, pp. 2232–2242, 2011.
- [9] C. Persello and L. Bruzzone, "A novel active learning strategy for domain adaptation in the classification of remote sensing images," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*. IEEE, 2011, pp. 3720–3723.
- [10] B. Demir, F. Bovolo, and L. Bruzzone, "Classification of time series of multispectral images with limited training data," *Image Processing, IEEE Transactions on*, vol. 22, no. 8, pp. 3219–3233, 2013.
- [11] G. Jun and J. Ghosh, "Spatially adaptive classification of land cover with remote sensing data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 7, pp. 2662–2673, 2011.
- [12] W. Kim and M. M. Crawford, "Adaptive classification for hyperspectral image data using manifold regularization kernel machines," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 11, pp. 4110–4121, 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," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 7, pp. 2810–2826, 2012.
- [14] L. J. Grady *et al.*, "Space-variant computer vision: a graph-theoretic approach," Ph.D. dissertation, 2004.
- [15] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento, "A (sub) graph isomorphism algorithm for matching large graphs," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 10, pp. 1367–1372, 2004.
- [16] M. Volpi, D. Tuia, G. Camps-Valls, and M. Kanevski, "Unsupervised change detection by kernel clustering," in *Remote Sensing, International Society for Optics and Photonics*, 2010, pp. 78 300V–78 300V.
- [17] J. Goldberger, S. Gordon, and H. Greenspan, "An efficient image similarity measure based on approximations of kl-divergence between two gaussian mixtures," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE, 2003, pp. 487–493.
- [18] D. Conte, P. Foggia, C. Sansone, and M. Vento, "Thirty years of graph matching in pattern recognition," *International journal of pattern recognition and artificial intelligence*, vol. 18, no. 03, pp. 265–298, 2004.
- [19] R. Zhang and A. I. Rudnicky, "A large scale clustering scheme for kernel k-means," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 4. IEEE, 2002, pp. 289–292.
- [20] I. Dhillon, Y. Guan, and B. Kulis, *A unified view of kernel k-means, spectral clustering and graph cuts*. Citeseer, 2004.
- [21] M.-S. Yang, "A survey of fuzzy clustering," *Mathematical and Computer modelling*, vol. 18, no. 11, pp. 1–16, 1993.
- [22] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*. Wiley.com, 2009, vol. 383.
- [23] J. He, Y. Liu, and R. Lawrence, "Graph-based transfer learning," in *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, 2009, pp. 937–946.
- [24] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," *Information Theory, IEEE Transactions on*, vol. 49, no. 7, pp. 1858–1860, 2003.
- [25] L. A. Zager and G. C. Verghese, "Graph similarity scoring and matching," *Applied mathematics letters*, vol. 21, no. 1, pp. 86–94, 2008.
- [26] J. T. Tou and R. C. Gonzalez, *Pattern recognition principles*. Wiley Online Library, 1974.

- [27] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.
- [28] S. Petrovic, "A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters," in *Proceedings of the 11th Nordic Workshop of Secure IT Systems*, 2006, pp. 53–64.
- [29] S. Samanta and S. Das, "Inter-domain cluster mapping and gmvc based transformation for domain adaptation," in *Pattern Recognition and Machine Intelligence*. Springer, 2013, pp. 74–81.
- [30] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 4, pp. 1231–1242, 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," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 9, pp. 3180–3191, 2009.



Biplab Banerjee is currently a research scholar at the Satellite Image Analysis lab, CSRE, IIT Bombay. He has obtained a M.E. degree in Computer Science from Jadavpur University, Kolkata. He has completed a 3-months research stay at the Signal and Image Processing Department, Telecom ParisTech, France for the duration April, 2012 to June, 2012. He was a visiting researcher at the Remote Sensing Lab, University of 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 (S05M07-SM13) received the Laurea (B.S.), the Laurea Specialistica (M.S.) degrees in telecommunication engineering (summa cum laude) and the Ph.D. in Communication and Information Technologies from the University of Trento, Italy, in 2001, 2003 and 2006, respectively where she has been a research fellow until June 2013. She is the founder and the head of the Remote Sensing for Digital Earth unit at Fondazione Bruno Kessler (FBK), Trento, Italy and a member of the Remote Sensing Laboratory (RSLab) in Trento. 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 SAR 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. She is a referee for several international journals. She is one of the Co-Investigatortors of the Radara for Icy Moon Exploration (RIME) instrument of the ESA Jupiter Icy moons Explorer (JUICE). Dr. Bovolo ranked 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 is an associate editor of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. She has been 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 has served on the Scientific Committee of the SPIE International Conference on Signal and Image Processing for Remote Sensing. Since 2014 she is co-chair of the same conference. She is the publication chair for International Geoscience and Remote Sensing Symposium 2015. Since 2012 she is a member of the international program committee of the conference on Pattern Recognition Applications and Methods. She has served 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 Avik Bhattacharya (M08) received his PhD degree from the Department of Signal and Images, Telecom ParisTech, Paris, France and the Ariana Research Group at the Institut National de Recherche en Informatique et en Automatique (INRIA), Sophia Antipolis, Nice, France. He has an integrated M.Sc degree in Mathematics from Indian Institute of Technology, Kharagpur. He is currently an Assistant Professor at the Centre of Studies In Resources Engineering, Indian Institute of Technology Bombay (IITB). Prior to joining IITB, he was a Canadian government research fellow at the Canadian Centre for Remote Sensing (CCRS) in Ottawa, Canada. He has received the Natural Sciences and Engineering Research Council of Canada (NSERC) visiting scientist fellowship at the Canadian national laboratories from 2008 to 2011. His current research includes SAR polarimetric application to cryosphere and planetary exploration, Statistical analysis of polarimetric SAR images, Machine learning and Pattern Recognition.



Lorenzo Bruzzone received the Laurea (M.S.) degree in electronic engineering (summa cum laude) and the Ph.D. degree in telecommunications from the University of Genoa, Italy, in 1993 and 1998, respectively. He is currently a Full Professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, pattern recognition, and electrical communications. Dr. Bruzzone is the founder and the director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His current research interests are in the areas of remote sensing, radar and SAR, 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. 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 editor/co-editor of 15 books/conference proceedings and 1 scientific book. His papers are highly cited, as proven from the total number of citations (more than 11200) and the value of the h-index (55) (source: Google Scholar). He was invited as keynote speaker in 24 international conferences and workshops. Since 2009 he is a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society. Dr. Bruzzone ranked 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 was recipient of many international and national honors and awards. Dr. Bruzzone was a Guest Co-Editor of different Special Issues of international journals. He is the co-founder 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. Currently he is an Associate Editor for the IEEE Transactions on Geoscience and Remote Sensing and the Canadian Journal of Remote Sensing. Since 2012 he has been appointed Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society. Dr. Bruzzone is IEEE Fellow.



Subhasis Chaudhuri (M 99SM 02F 11) is currently serving as the KN Bajaj Chair Professor with the Department of Electrical Engineering and the Deputy Director of Indian Institute of Technology Bombay, Mumbai, India. During 2005-2008, he served as the Head of the Department and during 2009-2013, as the Dean (International Relations). 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 is the recipient of Bhatnagar Prize in engineering sciences for the year 2004 and GD Birla Award in 2011.



Krishna Mohan Buddhiraju Prof. B. Krishna Mohan received his PhD in Electrical Engineering from IIT Bombay in 1991 specializing in Digital Image Processing and Analysis. He is a Professor and present Head of the Centre of Studies in Resources Engineering at IIT Bombay. His current interests are high resolution remote sensing, machine learning, and multimedia educational content development for satellite image processing. Currently he is supervising six PhD students and he has over 100 publications in journals, book chapters and conference proceedings. He is recipient of 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. He is Life member of Indian Society of Remote Sensing, Indian Society of Geomatics and member of IEEE.