Graph matching for adaptation in remote sensing

Devis Tuia, *Member, IEEE*, Jordi Muñoz-Marí, *Member, IEEE*,
Luis Gómez-Chova, *Member, IEEE*, Jesus Malo, *Member, IEEE*

Abstract—We present an adaptation algorithm focused on the description of the data changes under different acquisition conditions. When considering two acquisition conditions in a source and a destination domains, the adaptation is carried out by transforming one data set to the other using an appropriate nonlinear deformation. The eventually non-linear transform is based on vector quantization and graph matching. The transfer learning mapping is defined in an unsupervised manner. Once this mapping has been defined, the samples in one domain are projected onto the other, thus allowing the application of any classifier or regressor in the transformed domain. Experiments on challenging remote sensing scenarios, such as multitemporal VHR image classification and angular effects compensation, show the validity of the proposed method to match related domains and enhance the application of cross-domains image processing techniques.

Index Terms—Multitemporal classification, Domain adaptation, Transfer learning, SVM, model portability.

I. INTRODUCTION

THE problem of adapting models to a new unseen but related dataset is one of the major challenges for future remote sensing image processing. For instance, the increase in temporal resolution of image acquisitions of modern sensors allows data users to acquire several scenes of the same area, thus making multitemporal analysis [1]–[3], multiangular studies [4]–[6] and accurate change detection [7]–[9] possible. However, since new generation satellites such as QuickBird or WorldView 2 acquire images at an almost daily frequency, it is not realistic to provide supervised labeled information for each image to perform accurate image classification or biophysical parameter retrieval. Users must thus often rely on single acquisition ground truth to process series of images. Although tempting, direct application of classifiers or regressors on newly acquired and similar images can lead to catastrophic results: even if the objects represented in the images are roughly the same, differences in reflectance, illumination, and atmospheric conditions may induce local changes in the probability distribution function (PDF) instead

of simple global linear changes. Therefore, finding an efficient PDF matching strategy becomes an urgent and complex task for new generation remote sensing processing chains.

In machine learning literature, the problem of model adaptation is usually referred to as *domain adaptation* or *transfer learning* [10]. Transfer learning aims at describing the changes occurring among data distribution and to adapt models to related tasks. This recent field of research attempts to resolve the deformations occurring among similar domains but acquired under different conditions: in remote sensing, this would correspond to data sets coming from similar images. This means that, in transfer learning, we are particularly interested in transferring the knowledge from one or more source domains to a target domain rather than learning all source and target domains simultaneously [11].

In remote sensing literature, generic attempts to describe changes in manifolds are hardly found. Currently, the adaptation problem has been mainly considered from the perspective of specific classifiers, i.e. improve/modify the classifier so that it can cope with the shifted distribution. In the 1970s such changes were studied under the name of 'signature extension' [12], but the field received little interest until recently and remained specific to simple models and mid resolution land-use applications [13]–[16]. Textural filters have also been considered for increasing the robustness of classifiers among acquisitions: in [17], the authors observe that the use of textural features reduces spectral variability and thus allows successful classification of unseen scenes sharing the same classes. However, this approach only describes indirect adaptation, as no direct attempt to compensate for dataset shift is done and the improvement in performance is related to the well known decrease of intraclass variability related to the use of contextual features.

The advent of transfer learning theory renewed the interest of the community for these problems, but in most cases the effort is mainly focused on modifying the classifier while the transform between the domains is not explicitly studied. In [18], the samples in the new domain are used to re-estimate the classifier based on a Gaussian Mixture model. This way, the Gaussian clusters optimized for the first domain are expected to match the data observed in the second. In case of strong shifts there is no guarantee of appropriate correspondence between the identified clusters. In [19], randomly generated classifiers are applied in the destination domain. A measure of the diversity of the predictions is used to prune the classifiers ensemble. In [20], a support vector machine (SVM) classifier is modified by adding and removing support vectors from both domains in an iterative fashion: the model discards contradictory old training samples and uses the distribution of the new image to adapt the model to the new conditions.

DT was with the Image Processing Laboratory (IPL), Universitat de València, València, Spain. He is now with the Laboratoire des Systèmes d'Information Géographique (LaSIG), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. devis.tuia@epfl.ch, http://devis.tuia.googlepages.com, Phone: +41-216935785, Fax : +41-216935790.
JMM, LGC and JM are with the Image Processing Laboratory (IPL), Universitat de València, València, Spain. E-mail: jordi.munoz, luis.gomezchova, jesus.malo@uv.es, http://isp.uv.es, Phone: +34-963544021, Fax: +34-963544353.

In [21], matching of the first order statistics of the data clusters in source and destination domains is performed in a kernel space. Two kernels accounting for similarities among samples and clusters are combined and the SVM weights are computed using the labels from samples in the first domain. Finally, in [22] the shifts are compensated using a small set of examples sampled on the new distribution: the algorithm learns the shifts among domains iteratively and active learning queries [23], [24] are used to maximize the ratio between samples and increase in performance. Although adaptable to any classifier, the above methods consider data set shift in an indirect way since they are focused in the modifications to be done in the classifier.

Recently, some remote sensing researches have explicitly considered the distortions occurring between data manifolds: In [25], multitemporal sequences for each pixel are aligned based on a measure of similarity between sequences barycenters, thus consisting into a global measure of alignment. In [26], pixel's spectra are spatially detrended using Gaussian processes in order to avoid shifts related to geometrical differences or to localized class variability. Even if very efficient when applied to subregions of the same area, the approach can difficultly be extended to spatially disconnected areas. Finally, in [27], a kernel machine is regularized iteratively to ensure matching between the original and test distributions. Also, the use of manifold learning [28] techniques based on graph-Laplacian have been shown very accurate in problems with a reduced number of samples. Approximate methods for out-of-sample prediction [29] are required to apply this last approach to large-scale problems.

In this paper, we also study the possibility of using local manifold techniques for adaptation. This is justified by the generally nonlinear and low-dimensional structure of remote sensing data [30], [31]. To propose a simple and easily scalable solution, we focus on the description of the changes in the manifold by defining a non-linear transform based on *vector quantization* and *graph matching*. Vector quantization is used to retrieve local properties of the data clouds. A clustering algorithm is used to define a proximity graph in each domain, where the nodes of the graph are the cluster centroids (or *codebook vectors*). Graph matching has been strongly used in pattern recognition and computer vision to study similarities among data structures representing objects [32]–[34]. The need of matching procedures relying to other criteria than Euclidean distances has generated a wide field of research about the definition of similarity measures between graphs [32], [35], optimization techniques [36]–[38], and structured estimations [39]. In the proposed algorithm, the graphs of the two domains are matched using a procedure aiming at maximizing their similarity, while at the same time preserving the structure of the transformed graph. In our case, the number of nodes remains unchanged during the procedure and no supervised information (e.g. class labels) about the nodes is required.

Once the non-linear transform has been defined, the training samples in the source are matched to the data distribution of the target domain, or inversely depending on the needs.

The proposed adaptation approach (i.e. matching cluster centroids linked through a topographical structure) could be taken using alternative data representations. In the case of using local PCA or local ICA models [40], [41], the problem is that the local models are not linked in any way so there is no guarantee for a proper identification of the correspondence under strong deformations. Related methods that integrate the set of local representations into a single global non-linear representation [42], [43] solve this identification problem and have been used in adaptation [44]. However, they are hard to use in manifolds including bifurcations. The same is true for vector quantization techniques that assume an underlying topology, such as Self Organizing Maps [45]. As opposed to [42]–[45], the graph structure proposed here is more flexible than a d-dimensional reticle.

The proposed transform can be used in many ways: in a supervised context, labeled pixels in the source domain can be transformed and then used in the destination domain for training a supervised classifier. In the reverse direction, the image data in the destination domain can be transformed and then the model built for the source can be directly used on the transformed image without model retraining. Matching of the images can be also used for unsupervised tasks such as the correction of local artifacts [46], and illumination or angular effects [47], [48], such as the hot-spot example [49] used in the experiments.

The remainder of the paper is organized as follows: Section II presents the graph-based manifold matching procedure proposed, as well as the regularization on the matching sequence based on ensembles. Section III presents the datasets used in the experiments illustrated and discussed in Section IV. Section V concludes the paper.

II. GRAPH-BASED MANIFOLD MATCHING

This section presents the graph matching algorithm proposed. The manifold deformation problem appears in situations such as the one illustrated in Fig. 1. Labeled samples from objects in Fig. 1 (top row) can be obtained in particular acquisition conditions giving rise to data $(\mathbf{X}_l^S, \mathbf{Y}_l^S)$ for the source image. Data from the same objects can be obtained at a different time, as in Fig. 1 (middle), or at a different time and location, as in Fig. 1 (bottom). In general, the samples in the destination image are affected by seasonal changes and different illumination conditions, giving rise to different samples $\mathbf{X}^D$ for which no labeled information is available. The source and destination images do not have to be registered not even have the same number of pixels. The scope of the proposed method is to find a local transform $R(\mathbf{X}^S)$ that adapts $\mathbf{X}^S$ into a new set $\mathbf{X}^*$ that matches the PDF found in the destination domain $\mathbf{X}^D$. Even though Fig. 1 shows labels in each acquisition condition for the reader convenience, the labels are *never* used in the proposed adaptation transform.

A. Vector quantization

Rather than looking for a unique global mapping of all samples, we prefer to match a small number of representers obtained by vector quantization. Therefore, we reduce the n pixels of $\mathbf{X}^S$ to a series of centroids $\mathbf{c}^S$ using a quantization (or clustering) algorithm. The same process is applied to the

$\mathbf{X}^S$ Available

$\mathbf{Y}_l^S$ Available

$\mathbf{X}^{D_1}$ Available

$\mathbf{Y}_l^{D_1}$ Not available

$\mathbf{X}^{D_2}$ Available
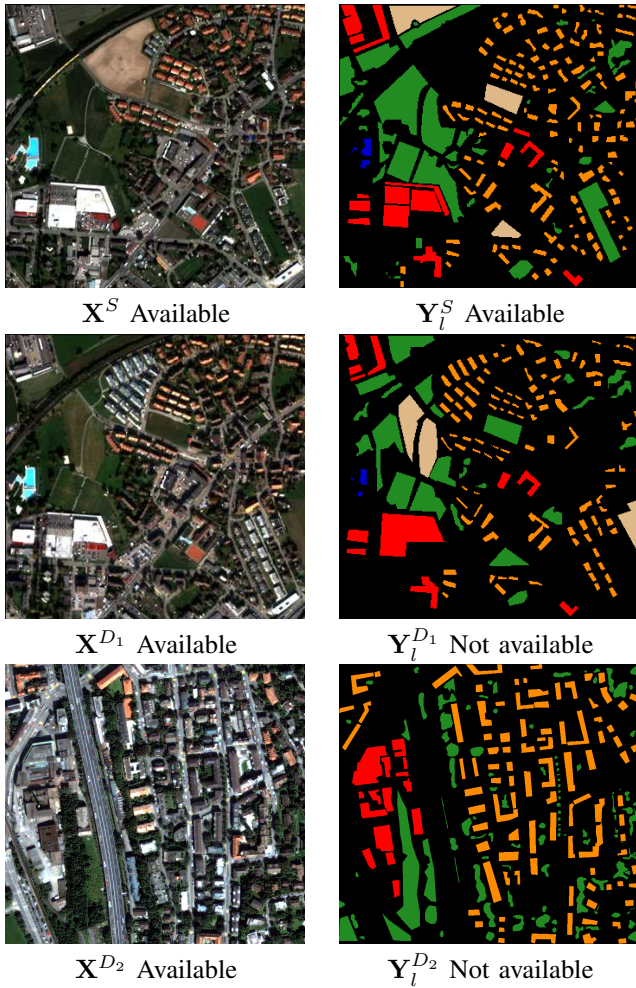
$\mathbf{Y}_l^{D_2}$ Not available

Fig. 1. Illustration of the domain adaptation problem. Source image $\mathbf{X}^S$ (top) and the same kind of objects, destination images, taken at a different time ($\mathbf{X}^{D_1}$ middle) and at a different time and location ($\mathbf{X}_2^D$ bottom). The right column shows the respective labeled pixels (colors are detailed in Table I). Note that $\mathbf{Y}_l^{D_1}$ and $\mathbf{Y}_l^{D_2}$ are usually not available and are used here only for validation purposes. Details on the images are given in Section III-B.

$\mathbf{c}^S$

$\mathbf{c}^D$

Fig. 2. Examples of centroids $\mathbf{c}^S$ and $\mathbf{c}^D$ obtained from the images $\mathbf{X}^S$ (top row of Fig. 1) and $\mathbf{X}^{D_1}$ (middle row of Fig 1), represented in the space defined by the green (G), red (R), and infrared (IR) bands of the QuickBird image. In this example, the centroids neighborhood relations are depicted by an undirected 2-nearest neighbor graph. Even though just three components are shown in these plots for visualization purposes, note that the centroids are computed using the full 4-dimensional vectors.

m pixels of $\mathbf{X}^D$ to obtain the representers $\mathbf{c}^D$. The number of centroids is a trade-off between flexibility and simplicity. Also note that the number of centroids can differ for the two domains. The important fact is that this number must be high enough to capture the nonlinear shape of the manifold. However, increasing the number of centroids beyond certain limit does not provide additional structural information but just increases the computational burden. Figure 2 shows $k = 100$ centroids obtained for the images $\mathbf{X}^S$ and $\mathbf{X}^{D_1}$ of Fig. 1, respectively. Note that differences in acquisition conditions result in local manifold deformations, but that the overall shape of the manifolds (and the structure of the associated graphs) remains similar for the two acquisitions. These properties (similar structure and smooth deformations of limited extent) are important conditions to have an efficient matching.

B. Graph matching

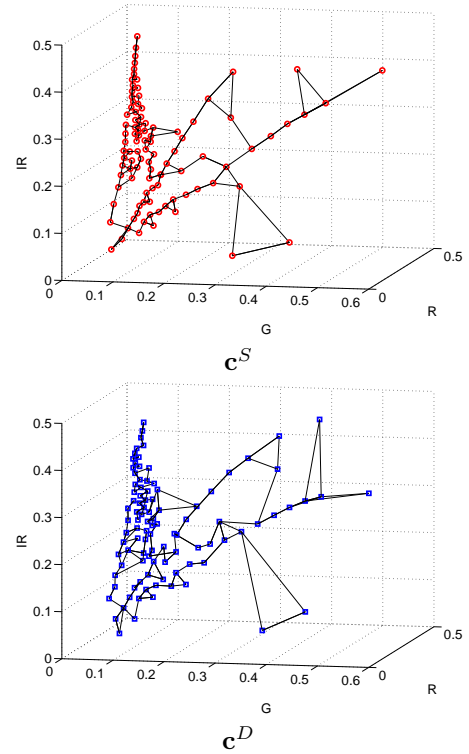We would like to match the most similar centroids in both domains. An Euclidean match, i.e. assigning each centroid to its nearest neighbor in the new domain, is not recommendable, since Euclidean distances between the two structures may not represent the manifold distances [28]. The concept is schematized in Fig. 3 for two hypothetical datasets. Euclidean matching of the centroids –illustrated in the middle panel of the second row– changes the structure of the transformed data, that loses the difference between the two extremities of the right part. The correct match should preserve such structure, as illustrated in the bottom-right panel of the same figure. The algorithm proposed in this paper enforces this desirable property.

We propose to consider the set of centroids as a *graph*. A graph is a set composed of nodes connected by edges when they are considered similar. Common criteria to establish such similarity are nearest neighbors rules or local ϵ-balls. Using a graph, only local distances make sense and geodesic distances following the manifold structure can be computed [28].

A compact representation of a graph is given by the adjacency matrix $\mathbf{W}$. Using a k-nearest neighbors rule, this matrix has elements

$$W(i,j) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are neighbors} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Graphs for domains $\mathbf{X}^S$ and $\mathbf{X}^{D_1}$ are illustrated in Fig. 2. Each edge connecting two centroids represents a local neighborhood relationship.

Our approach to graph matching for manifold deformation description can be seen as the joint optimization of two criteria: (1) minimization of centroid displacement, and (2) structure preservation. The rationale behind the first criterion is that transitions in the manifolds are generally smooth so large displacements of the destination centroids, $\mathbf{c}^*$, from their initial position, $\mathbf{c}^S$, should be avoided. The second criterion refers to the fact that local nearest-neighbor structures are generally preserved in manifold deformations. According to this, large modifications in the graph structure, as encoded in the $\mathbf{W^S}$ matrix, should be avoided. These two criteria are combined in the following cost function to be minimized:

$$\mathbf{c}^*_{opt} = \min_{\mathbf{c}^*} \left\{ \|\mathbf{c}^S - \mathbf{c}^*\|_2 + \|\mathbf{W}^S - \mathbf{W}^*\|_1 \right\} \quad (2)$$

The first term aims at minimizing the deformation, while the second aims at structure preservation by comparing the initial adjacency matrix $\mathbf{W^S}$ with the adjacency matrix after displacement $\mathbf{W}^*$. In Fig. 4, three undirected graphs with five nodes and their corresponding adjacency matrices are represented. They illustrate the structure criterion for two possible modifications of the nodes.

Graph matching optimization routines are usually NP-hard problems [33], [39]. To make the problem solvable, we used a reduced algorithm considering cross-domains relationships: since transitions among remote sensing acquisition are generally smooth, we limit the number of allowed displacements to those between nearest neighbors between graphs, as illustrated in Fig. 5(a). We will refer to this displacements graph as the *cross-domains graph*.

Once these possible destinations have been defined, the centroids are considered sequentially for matching. At each step, a centroid and its neighbors in $\mathbf{c}^S$ are considered for being moved towards their respective possible destinations. In this way, each centroid is displaced smoothly with its neighboring centroids and only a specific section of the graph is considered at each displacement. Only combinations where the scene centroids do not move to the same centroid in $\mathbf{c}^D$ are considered admissible (*local injectivity* constraint). The result of the sequential procedure optimizing Eq. (2) for each centroid is illustrated in Fig. 5(b). Note that this does not

Fig. 4. Graph regularization. Top row: three 5-nodes graphs. Bottom row: representation of the corresponding adjacency matrices and result of the right term of Eq. (2) assessing similarity of $W1$ and $W2$ with respect to the original $W0$.

(a)

(b)
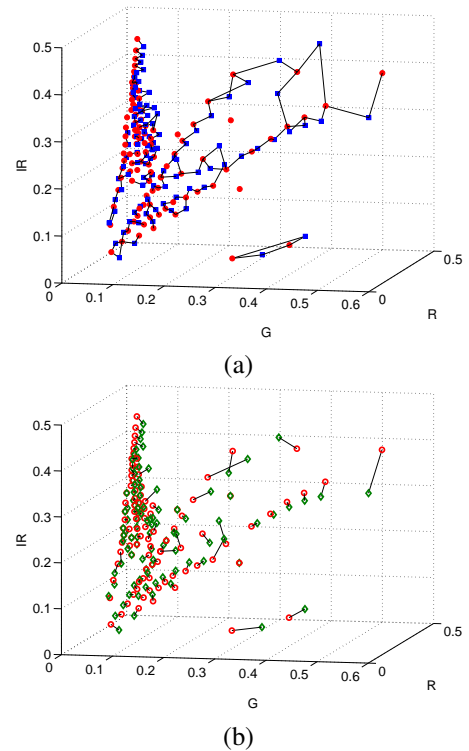
Fig. 5. Graph matching procedure. (a) Displacements considered for adaptation (cross-domains graph): in this case, each centroid in $\mathbf{c}^D$ (blue squares ■) can attract two centroids in $\mathbf{c}^S$ (red circles ●). (b) Selected displacements: each centroid of $\mathbf{c}^S$ (red circles ○) is displaced towards its new position $\mathbf{c}^*$ (green diamonds ◊). The plot (b) shows the local transformations applied to the training data.
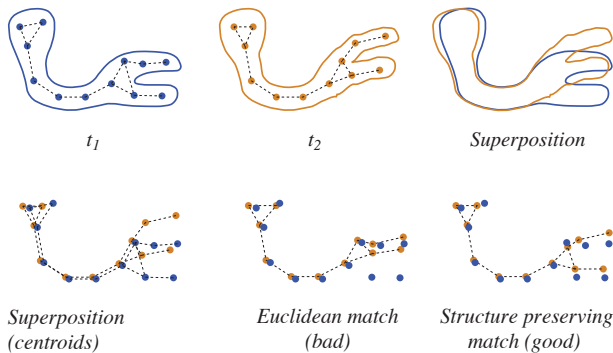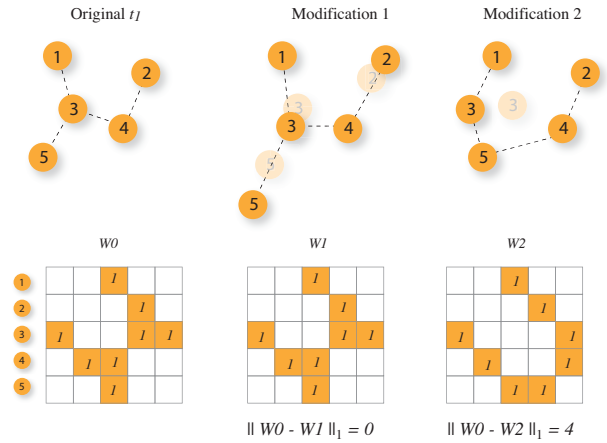

Fig. 3. Examples of Euclidean and structure preserving matching. Top row: two similar artificial data sets are quantized and their superposition shows a local difference in the right part of the PDF. Bottom row: examples of Euclidean (middle) and structure preserving (right) matching.

provide a globally injective result (a result where for each centroid a different destination is chosen): the local injectivity constraint is applied to each centroid and to its neighbors for a given iteration in the sequence, but not among sequences. Consider a centroid in the source $\mathbf{c}^S_1$ and its neighbors $\mathbf{c}^S_{nn}$: if centroid $\mathbf{c}^S_1$ moves towards $\mathbf{c}^D_1$, the $\mathbf{c}^S_{nn}$ won't be allowed to move towards $\mathbf{c}^D_1$. On the contrary, there is nothing avoiding the following centroid in the sequence, $\mathbf{c}^S_2$, or its neighbors to
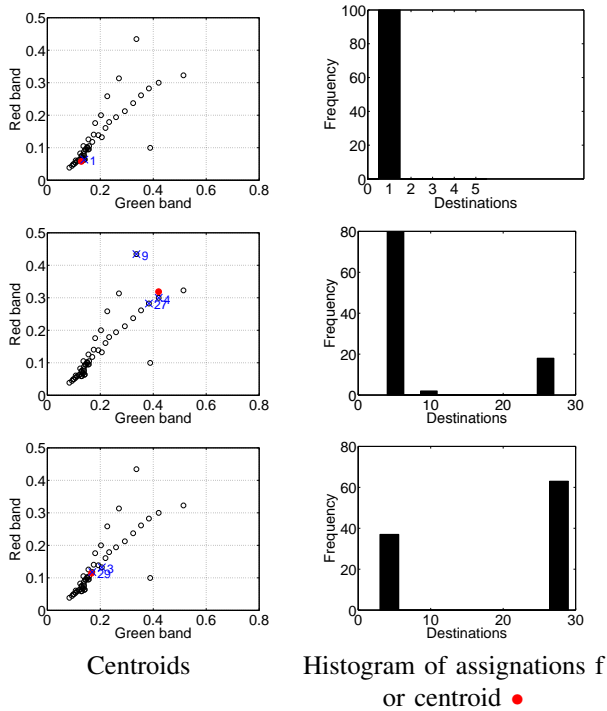
Fig. 6. Histograms of selected destinations for three centroids over 100 sequences: (top) clear assignment, $H(\mathbf{c}_1^S) = 0$; (middle) uncertain assignment on a single matching result $H(\mathbf{c}_2^S) = 0.20$; (bottom) uncertain assignment on a many matching results $H(\mathbf{c}_3^S) = 0.41$. Red filled dots ● are the origin centroid and blue crosses × are destinations selected by at least one member of the committee among the possible destinations ○. The numbers relative to each × correspond to the horizontal axis of the corresponding histogram.

move towards $\mathbf{c}_1^D$. We added this local injectivity constraint 1) to avoid local structures to collapse into a single destination, and 2) to reduce computational complexity, as we are reducing the combinatory of possible source-destination mappings.

C. Dependence on matching sequence and ensemble methods

Since adaptation is performed locally and following a given sequence, i.e., each centroid is considered with its neighborhood sequentially, the final matching result may depend on the order of consideration of the centroids adopted by the algorithm. Therefore, relying on a single matching result could lead into a local minimum, as it depends on the sequence of consideration of the centroids. To cope with this risk, we considered ensemble methods [50], that have shown to increase robustness when considering sequential algorithms that approach optimal solutions in remote sensing problems [51]–[53]. In our case, we build a committee of experts by running the proposed matching algorithm several times varying the sequence of consideration of the centroids. The sequences are randomly drawn. The final assignment between nodes is determined by majority vote on the committee.

Figure 6 illustrates the variability or uncertainty in the committee of graph matching solutions. As an example, we used the graphs of Figures 2 and 5. The uncertainty of committee vote is measured in terms of normalized entropy of the frequency that a node $\mathbf{c}_i^S$ is matched to a node $\mathbf{c}_j^D$. Such quantity is computed as $H(\mathbf{c}_i^S) = \sum_j^k -p(\mathbf{c}_i^S \rightarrow$

$\mathbf{c}_j^D) \log(p(\mathbf{c}_i^S \rightarrow \mathbf{c}_j^D))/H^+$, where H^+ is a normalization constant given by the maximum possible entropy obtained when all destinations are equiprobable ($H(\mathbf{c}) \in [0, 1]$). In analyzing the result of possible assignments given by the committee one can find three situations depending on the agreement of the committee:

- Total agreement. The centroid is systematically attributed to the same destination $\mathbf{c}^D$ and show an entropy of assignation of 0 (top row of Fig. 6).
- Different assignments are made with very different probabilities (middle row of Fig. 6). In this case, related to small entropy values ($0 < H(\mathbf{c}_i^S) < 0.3$), the effect of the committee is important, because it fixes single wrong assignments; in the example reported, a wrong attribution to centroid #9 is corrected by the rest of the committee.
- Many equiprobable assignments (bottom row of Fig. 6). In this case, related to high entropy values ($H(\mathbf{c}_i^S) > 0.3$), the equiprobable assignations correspond to centroids that are very close between each other, thus making the choice of the specific destination less important; in the example reported, centroids #3 and #29 are extremely close. Please note that this scenario contains a limit case where the most probable destination is 'non assignation': in this case, the algorithm decides not to move the centroid, even if there is a strong second best option.

III. DATA AND SETUP

This section presents the datasets used and the setup of the experiments performed. In particular we consider (1) a synthetic example in order to assess the relevance of the structural term of the proposed cost functional and the difference with Euclidean matching; (2) two classification examples in which the proposed local graph matching is compared to Euclidean Matching, global histogram matching [54], and to the no-adaptation baseline; (3) a radiometric correction example.

A. Toy dataset

From a data processing point of view, we can distinguish two types of deformations: global (such as translations, scaling and rotations of the feature space) and local (same, but locally). The first can be corrected with known methods such as a common PCA or histogram matching [54], but the second requires a specialized algorithm, such as the graph matching procedure proposed here. Regarding local transformations, we studied which amount of deformation could be managed by the proposed algorithm. To study this important point, we consider a synthetic dataset consisting of two dimensional series of centroids and we try to reproduce the qualitative behavior found in remote sensing data with a local rotation as the one shown in Fig. 3. This dataset will be used to study convergence properties of the proposed graph matching technique and to compare it to different types of Euclidean matching. Three increasing levels of distortion are considered, as illustrated in Fig. 7. The three levels of distortion basically impact the upper right part of the $\mathbf{X}^S$ graph (red ●) only and convert the adaptation problem from a relatively simple one (low
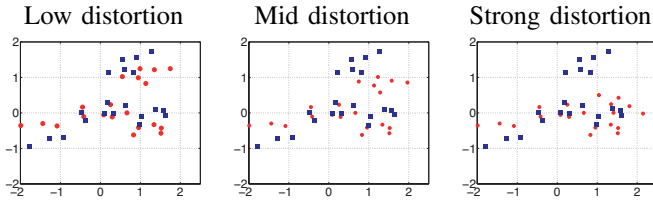
Fig. 7. 2-D toy dataset used in the first experiment for three levels of local distortion. $\mathbf{c}^S \bullet \rightarrow \mathbf{c}^D \blacksquare$.

distortion, left column) to a hard one (strong distortion, right column). We compare the proposed graph matching procedure with three progressively simplified local matching algorithms:

- *Graph matching without structural constraint*: the proposed graph matching, but without the graph structure constraint. With this setting, only the constraint on the movements of the centroids is considered. This setting is used to assess the importance of the structural constraint on the final result.
- *Injective Euclidean matching ('E-Inj.')*, where each centroid is mapped to its closest neighbor, which is then removed from the list of possible destinations.
- *Euclidean matching ('Eucl.')*, where each centroid is mapped to its closest neighbor in the $\mathbf{c}^D$ centroids.

B. Classification in different environments

This example considers a series of three images acquired by the QuickBird satellite, one the 2*nd* of August 2002 (image $\mathbf{X}^S$) and two the 6*th* of October 2006 (images $\mathbf{X}^{D_1}$ and $\mathbf{X}^{D_2}$). The images provide 3 bands in the visible and 1 band in the near infrared regions of the spectrum. Images $\mathbf{X}^S$ and $\mathbf{X}^{D_1}$, illustrated in top and middle rows of Fig. 1, describe a residential neighborhood of the region of Brutisellen, downtown Zurich (Switzerland). They are coregistered, even though this is not a requirement of the proposed method, and have a resolution of 2.4 m. The third image, $\mathbf{X}^{D_2}$, illustrated in the bottom row of Fig. 1, is another area of the 2006 Zurich image, near the Rieterpark on the West side of the Zurich lake.

Adaptation is required because of the differences in acquisition conditions of the images. Both images have been acquired between 10 and 11 a.m. with similar sun azimuth angles and sun zenith angles around 60 and 30 deg, respectively. Moreover, the 2002 image is taken with an off-nadir angle of 5.6 degrees, while the 2006 image with a 12.2 degrees off-nadir angle, thus enhancing the shadowing effect. Seasonal effect of vegetation cycle is also visible, as the first image is taken in summer, while the second in autumn.

For each acquisition, between 20000 and 30000 labeled pixels were extracted by accurate photo-interpretation. For the source image a series of labeled pixels is available, $\mathbf{X}_l^S = \{\mathbf{x}_i\}_{i=1}^l$ with $\mathbf{Y}_l^S = \{y_i\}_{i=1}^l$. These data generalize well the rest of the data of the source domain $\mathbf{X}_u^S = \{\mathbf{x}_i\}_{l+1}^n$. Note that in the classification experiments the labeled information in the destination domain is used only for validation purposes. Five classes of interest have been retained for the two first images (see Table I for details), while for the Rieterpark dataset only three of them were present in the scene. Changes between

image $\mathbf{X}^S$ and $\mathbf{X}^D$ are not considered as change class, since the objective of this experiment is to perform multitemporal classification, rather than change detection.

For these experiments, two quantization methods have been used, k-means and fuzzy c-means (FCM). Note that any clustering algorithm can be used for the definition of the centroids and we selected these because they are well established, simple and allowed to consider the cases of crisp and fuzzy assignments. Two quantizations, with $k = 50$ and $k = 100$ have been considered and 100 sequences have been generated to build the committee. Once the matched centroids $\mathbf{c}^*$ and the corresponding transforms have been defined, a SVM is trained for different numbers of training pixels, randomly selected but respecting the proportions per class in the ground truth: [0.5 % = 143, 1% = 286, 5% = 1430, 10% = 2860]. SVM parameters have been optimized by 4-fold cross validation. Ten realizations of the classification with independent training sets have been performed to ensure stability of the result.

In these experiments, two adaptation settings are compared:

1) Adaptation of the training samples to $\mathbf{X}^D$ ($\mathbf{X}_l^S \rightarrow \mathbf{X}_l^* \rightarrow \mathbf{X}^D$): once the transform has been determined for each centroid, the training samples undergo the same transform as the centroid they have been assigned to by the clustering algorithm. The model is retrained with the adapted training samples, $\mathbf{X}_l^* = \{\mathbf{x}_i^*\}_{i=1}^l$ with $\mathbf{Y}_l^S = \{y_i\}_{i=1}^l$, and used to predict the data in the destination domain $\mathbf{X}^D$.

2) Adaptation of the image in the destination domain ($\mathbf{X}_l^S \leftarrow \mathbf{X}^* \leftarrow \mathbf{X}^D$): the whole PDF in $\mathbf{X}^D$ can be adapted backwards to the distribution of $\mathbf{X}^S$. This solution has the advantage to allow the use of the same model developed for $\mathbf{X}^S$ without retraining, since the labeled pixels $\mathbf{X}_l^S$ have not been moved and keep their mutual relationships.

C. Hot spot correction

The last adaptation problem considers the correction of angular effects. To this end, we studied a set of two HyMap hyperspectral images taken over the Barrax test site in 1999 during the DAISEX99 ESA campaign. The flight was designed to enhance angular effects, so that two overpasses were acquired over the same site at noon: one in North-South direction along the solar principal plane and the other in East-West direction along the orthogonal plane (Fig. 8d). This flight configuration allowed the observation of an interesting

TABLE I
LABELED POINTS $\mathbf{X}_l$ AVAILABLE FOR THE SOURCE AND DESTINATION DOMAINS FOR THE ZURICH IMAGES.

Class	2002 $\mathbf{X}_l^S$	2006 $\mathbf{X}_l^D$ (same area)	$\mathbf{X}_l^{D_2}$ ($\neq$ area)	Legend color
Residential	6746	5245	13481	Orange
Commercial	5277	4992	4402	Red
Vegetation	14218	9454	11150	Green
Harvested vegetation	2523	2883	-	Brown
Water	269	165	-	Blue
Total	29033	22739	29033	

(a) Hot spot image $\mathbf{X}^S$ (b) Destination image $\mathbf{X}^D$

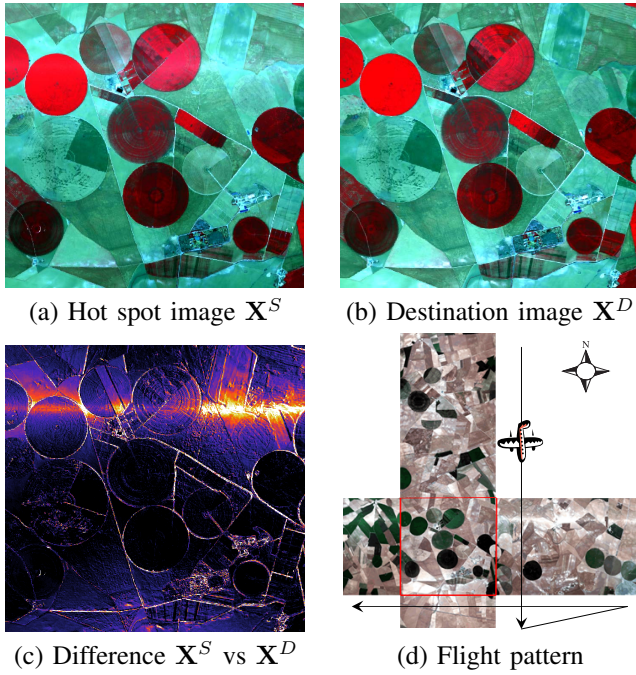(c) Difference $\mathbf{X}^S$ vs $\mathbf{X}^D$ (d) Flight pattern

Fig. 8. DAISEX dataset. Top left: acquisition perpendicular to the principal plane, with hot spot maximizing angular effects. Top right: acquisition parallel to the principal plane, without hot spot. Bottom left: differences between $\mathbf{X}^S$ and $\mathbf{X}^D$, corresponding to angular hot spot. Bottom right: flight pattern followed for DAISEX99 at Barrax (Spain).

(a) $\mathbf{X}^S$ (b) $\mathbf{X}^D$

Fig. 9. Data manifolds for the images with [$\mathbf{X}^S$ (a)] and without [$\mathbf{X}^D$ (b)] hot spot. In light blue, pixels corresponding to the hot spot area, which are those that should be adapted between the domains.

directional anisotropy effect called *hot spot* in the east-west flight line at noon. The hot spot is characterized by a sharp increase in reflectance when the surface is viewed in the same direction as it is illuminated by the Sun. This phenomenon, although only appearing under unusual angular configurations, provides an excellent opportunity to test the proposed adaptation method in a challenging scenario where a high reflectance peak appears in the affected region (see Fig. 8a-c). In a previous publication [49], a method for compensating for the angular effects based on bidirectional reflectance distribution function (BRDF) modeling was proposed. However, such a compensation also requires knowledge of the land covers at the surface, that were estimated using a classification result. In this paper, we propose to perform this correction *without* prior knowledge of the surfaces on the ground and their BRDF characterization.

Figure 9 illustrates the manifold of the two acquisitions. Globally, the main manifold shape is similar, as illustrated in Figs. 9a and 9b respectively. However, pixels located in the hot spot region in Fig. 8 behave differently: their values observed in Fig. 9 (highlighted in light blue) show 1) that the hot spot corresponds to an increase of the reflectivity values, and 2) that the hot spot is localized in the spectral space, thus allowing to consider a compensation strategy based on vector quantization as the one proposed. For this experiment, we considered a single adaptation setting, where the image with the hot spot ($\mathbf{X}^S$) is adapted to the one without it ($\mathbf{X}^D$), i.e. $\mathbf{X}^S \rightarrow \mathbf{X}^*$. Classification is not the aim of this example, so no labeled information is used for these experiments. Quality of the transform will be assessed by comparing the radiometric and spectral differences with respect to the $\mathbf{X}^D$ image.
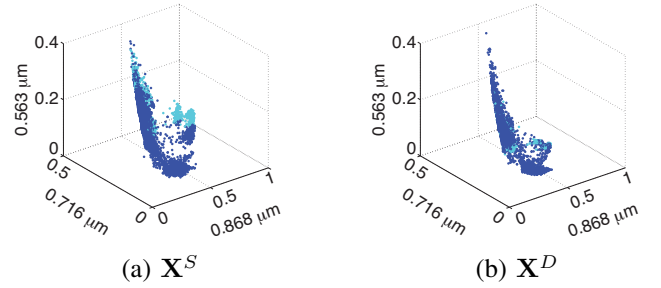
IV. RESULTS AND DISCUSSION

A. Toy dataset

Results for the toy dataset are shown in Fig. 10: in almost all the situations, the proposed graph matching procedure (second column of Fig. 10) finds a suitable match for each centroid (compare with the expected assignments in the first column), given that a correct displacement is present in the possible movements. The only exception is observed when some centroids are not attributed: this case occurs when the current centroid position is more suitable than any possible movement for a given set of centroids. The committee, alternating the sequences, corrects only partially this behavior, since for peripheral centroids the amount of non-attributions is higher than the amount of corrected attributions (see the discussion at the end of Section II-C). As an example, the centroid that remains still (upper right side for the low distortion case) is attributed correctly by about 40% of the committee members, while in 60% of the cases it does not move. This behavior could be handled by adding specific cases regarding non-attributed centroids, but it will result in an ad-hoc solution which is out of the scope of this paper. The other matching strategies are illustrated in the third to fifth columns of Fig. 10:

- *Graph matching without structural constraint* (third column of Fig. 10): removing the structural constraint on the $\mathbf{W}$ matrices has little effect in the simplest case of low distortion, where the Euclidean matching can also provide a good solution (see discussion below). On the contrary, when the distortion become stronger and Euclidean distances will lead to wrong assignations, the algorithm become more and more sensitive to the level of distortion. Some correct matches are still found, but this is related to the fact that the possible movements are limited by the cross-domains graph.

- *Injective Euclidean matching ('E-Inj.', fourth column of Fig. 10): this case is injective, in the sense that each centroid in $\mathbf{c}^S$ must be mapped to a different centroid in $\mathbf{c}^D$. Adding such a constraint implies that when all proximal centroids have been attributed, centroids start to be mapped to centroids of the destination domain that may be too far away (and become the nearest neighbors, just because the others are already occupied). This results in wrong assignations, as those observed in the bottom row, where centroids of the lower structure of $\mathbf{c}^S$ are
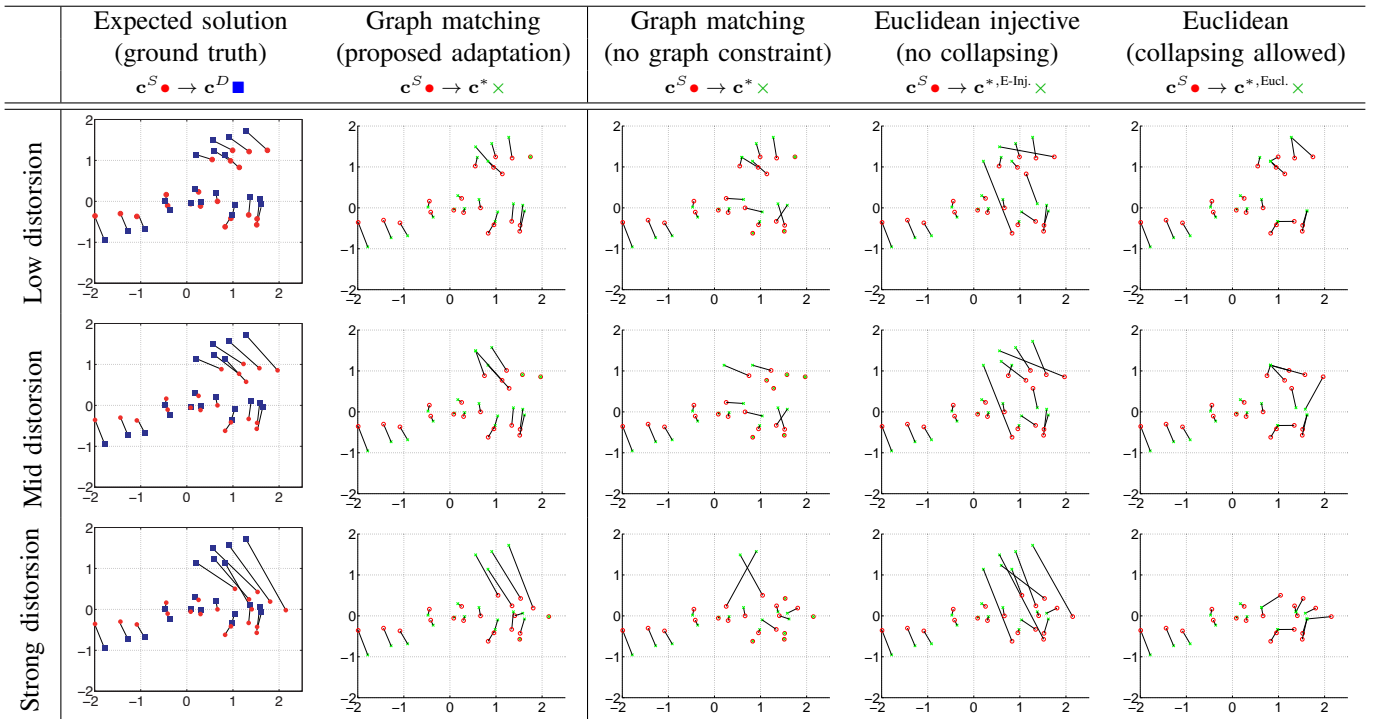
Fig. 10.   Graph matching adaptation with increasing distortions on a 2D toy example and $k = 4$ nearest neighbors in the cross-domains graph. Euclidean matches to the closest centroids in a single sequence, Euclidean injective matches to the closest unmatched centroid.

mapped to centroids of the upper structure of $\mathbf{c}^D$. This case is very dangerous, since it maps to centroids that i) are far away and ii) do not respect the graph structure.

- *Euclidean matching ('Eucl.', fifth column of Fig. 10)*: in this case, each centroid is mapped to its nearest neighbor in the $\mathbf{c}^D$ graph. Even if the solution retrieved is correct for small levels of distortion (top row), an Euclidean match ends up providing wrong matches as soon as the lower part of the $\mathbf{c}^D$ graph become closer than the upper part. In this matching, being non-injective (i.e., it allowing more than one centroid to be mapped to the same centroid in $\mathbf{c}^D$), few errors other than errors of proximity are observed, but overall, the algorithm cannot find the correspondences.

The observed behavior shows the power, but also the limitations of the proposed matching system proposed: the graph matching strategy works well for low to mid distortions, but when the distortion becomes strong, a rich enough cross-domains graph is mandatory to allow the matching procedure to find good correspondences. The structure criterion is robust enough to discard wrong assignments, but only if the correct ones are present in the list of possible attributions.

B. Classification in different environments

Table II summarizes the numerical results obtained while considering adaptation from $\mathbf{X}^S$ to $\mathbf{X}^{D_1}$, i.e. between coregistered images. The last two rows show the results obtained by a model without adaptation (lower bound on performance), as well as those by a model using labeled pixels in the $\mathbf{X}^{D_1}$ image (upper bound on performance). We observe a that the difference in performance is between 5 and 10%, depending

on the number of labeled examples. It is worth noting that for low numbers of training samples, results for adapted models are even better than when using labeled samples from the destination domains, however, σ in these cases is higher and the differences are thus not statistically different.

A general trend observed in the proposed adaptation models is that all experiments improve the results of classifiers without adaptation by about 7% both in overall accuracy and kappa statistics. This confirms that the shifts between the acquisitions are present and that the proposed adaptation strategy corrects the local distortions. Secondly, the proposed graph matching procedure always outperforms global histogram matching [54] by 2-5%, thus showing the interest of a local correction method maintaining the local data structures instead of a correction method based on global statistics of the images. Finally, with respect to other local methods such as Euclidean matching and Euclidean injective, we can still observe an improvement around 2% and 8% in classification, respectively. As for the toy example in the previous section, the Euclidean injective matching leads to a deterioration of the solution resulting in lower accuracies than the case without adaptation.

Backward adaptation (displacement of destination domain $\mathbf{X}_l^S \leftarrow \mathbf{X}^* \leftarrow \mathbf{X}^{D_2}$) provides slightly lower results than forward adaptation. However, this setting has the advantage of keeping the relationships among training samples unchanged, thus avoiding re-training of the SVM. The robust solutions provided by backward adaptation show that the destination domain can be adapted efficiently to the source domain, thus opening interesting opportunities for multitemporal classification of large image time series.

Regarding the size of the committee, a minimal size of 50 members seems necessary to minimize the average entropy

TABLE II

AVERAGE CLASSIFICATION RESULTS FOR THE BRUTISELLEN IMAGE $\mathbf{X}^{D_1}$ USING TWO QUANTIZATION ALGORITHMS AND DIFFERENT PARAMETERS SETTINGS. RESULTS USING COMMITTEES OF 100 MODELS ARE REPORTED FOR THE GRAPH MATCHING ALGORITHM. IN BOLD, THE BEST RESULT FOR EACH ADAPTATION SETTING. (EUCL = EUCLIDEAN MATCHING; E-INJ. = EUCLIDEAN INJECTIVE MATCHING; HM = HISTOGRAM MATCHING).

|	Clustering	k	# training from $\mathbf{X}_l^S$	Overall Accuracy (OA)								Kappa statistics (κ)							
				143		286		1430		2860		143		286		1430		2860	
				μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Graph matching	k-means	50	$\mathbf{X}_l^S \rightarrow \mathbf{X}_l^* \rightarrow \mathbf{X}^{D_1}$	86.26	2.42	**88.37**	2.29	**89.46**	1.37	**89.97**	0.55	0.802	0.035	**0.833**	0.034	**0.848**	0.020	**0.856**	0.008
		100		86.94	2.33	86.98	2.43	88.79	1.03	89.19	0.61	0.811	0.035	0.812	0.035	0.839	0.015	0.844	0.009
	FCM	50		85.56	1.68	87.30	1.80	88.06	0.54	88.09	1.08	0.790	0.025	0.816	0.026	0.827	0.008	0.828	0016
		100		**87.41**	2.29	88.25	1.87	87.25	0.74	88.42	0.80	**0.819**	0.033	0.832	0.027	0.818	0.011	0.835	0.012
	k-means	50	$\mathbf{X}_l^S \leftarrow \mathbf{X}^* \leftarrow \mathbf{X}^{D_1}$	85.72	1.61	87.44	1.20	87.82	0.81	88.39	0.52	0.793	0.024	0.819	0.017	0.824	0.012	0.832	0.008
		100		86.36	1.22	87.29	1.40	87.12	0.79	87.72	0.49	0.803	0.018	0.817	0.019	0.815	0.011	0.823	0.007
	FCM	50		85.08	1.74	86.80	0.99	87.39	0.68	88.31	0.66	0.784	0.026	0.810	0.014	0.819	0.010	0.832	0.010
		100		85.87	1.46	87.65	1.59	86.97	0.79	88.40	0.64	0.797	0.022	0.824	0.022	0.815	0.011	0.835	0.010
Eucl.	k-means	100	$\mathbf{X}_l^S \rightarrow \mathbf{X}_l^{*,\text{Eucl}} \rightarrow \mathbf{X}^{D_1}$	85.64	2.53	86.81	1.68	87.47	0.79	87.72	1.06	0.796	0.037	0.814	0.024	0.824	0.011	0.827	0.015
	k-means	100	$\mathbf{X}_l^S \rightarrow \mathbf{X}_l^{*,\text{E-Inj.}} \rightarrow \mathbf{X}^{D_1}$	79.80	3.93	79.72	2.50	80.59	1.02	81.46	0.43	0.704	0.062	0.705	0.039	0.719	0.015	0.732	0.006
	Hist. matching		$\mathbf{X}_l^S \rightarrow \mathbf{X}^{*,\text{HM}} \rightarrow \mathbf{X}^{D_1}$	84.94	1.59	86.45	2.28	84.13	1.30	85.36	1.59	0.785	0.023	0.807	0.032	0.776	0.018	0.793	0.022
	No adaptation		$\mathbf{X}_l^S \rightarrow \mathbf{X}^{D_1}$	80.84	1.37	81.96	1.94	80.67	0.97	81.85	1.00	0.730	0.019	0.747	0.027	0.730	0.013	0.746	0.014
	Training from $\mathbf{X}^{D_1}$		$\mathbf{X}_l^{D_1} \rightarrow \mathbf{X}^{D_1}$	85.43	1.42	87.31	0.84	89.65	0.37	90.12	0.29	0.794	0.020	0.826	0.012	0.854	0.005	0.860	0.004
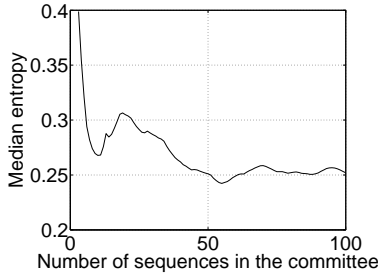
Fig. 11. Dependence of the uncertainty of the committee (median of the normalized entropy) as a function of the committee size (number of assignment sequences) for the Brutisellen dataset.

(so, the number of equiprobable predictions per centroid), as illustrated by the curve in Fig. 11. As computational complexity grows linearly with the number of sequences considered, the evaluation of a limited number of sequences remains reasonable in terms of processing time.

Results obtained when predicting a spatially disjoint images are reported in Table III. When considering a different scene such as the Rieterpark image ($\mathbf{X}^{D_2}$) shown in last row of Fig. 1, we observe that the direct application of the SVM trained on $\mathbf{X}^S$ does not provide reliable classification of the scene (accuracy of 75%, while training with labeled data from $\mathbf{X}^{D_2}$ results in 85%). This can be explained by two factors, also visible in the left confusion matrix of Table IV: first, since the source training data consider five classes, while the destination domain only presents three, there are errors due to the classification of pixels in one of the missing classes; second, the manifold structure is different, thus implying the domain adaptation problem that we study in this paper.

The application of the proposed approach results into an improvement of the general classification performance between 5 and 6% in overall accuracy and about 9% in kappa with respect to the 'No adaptation' scenario. In this case, the performance of the model considering only three classes (last row of Table III) is not reached, as there is still some misclassifications in the missing class 'harvested vegetation'

(see the central confusion matrix in Table IV), but the number of misclassifications is strongly reduced by the adaptation procedure. In this case, the adaptation procedure does not resolve the conflict between residential and commercial buildings, since in $\mathbf{X}^S$ commercial buildings have mainly bright roofs, while in $\mathbf{X}^{D_2}$ they have dark roofs, but increases the user's accuracy by strongly reducing the classification of pixels into a missing class. In order to resolve the confusions between buildings, one should consider applying contextual filters, as in [55], [56]. As in the previous case study, comparison with Euclidean graph matching and Histogram Matching give rise to poorer results, but with a stronger deterioration of the solution when using the Euclidean injective setting.

A final observation can be done regarding the stability of the predictions reported in Table III: without adaptation, the standard deviation of the result is higher (between 3 and 5 %), while when adapting the manifold prior to classification, the standard deviation of the accuracy is between 0.3 and 1.4 %. This is a consequence of the use of committees of graph matching procedures, that helps the method to converge to a stable solution.

Also, in these experiments, no significant differences can be appreciated between the two analyzed clustering algorithms. However, k-means algorithm provides slightly better results than its fuzzy version. This can be explained since we are evaluating the algorithms by the performance of a sparse classifier trained on a matched graph. This means that a rough displacement of the centroids (and of the corresponding training pixels) can be acceptable, since the centroids represents all the pixels in the image and the training pixels will be within these clusters. Moreover, the clusters are supposed to belong to single classes (the cluster assumption states that classes are separated by low-density regions). So, when using k-means, the training samples of different classes are 'pushed apart', thus giving a consistent advantage for classification. On the contrary, for FCM, points at the interface between clusters tend to move jointly with the closest centroids, which usually 'pull' them in two opposite directions. As a result, some training

TABLE III
AVERAGE CLASSIFICATION RESULTS WHEN PREDICTING THE RIETERPARK ZURICH DATASET $\mathbf{X}^{D_2}$ USING TWO QUANTIZATION ALGORITHMS AND DIFFERENT PARAMETERS SETTINGS. ALL EXPERIMENTS AVERAGE COMMITTEES OF 100 MATCHING PROCEDURES. IN BOLD, THE BEST RESULT FOR EACH ADAPTATION SETTING. (EUCL = EUCLIDEAN MATCHING; E-INJ. = EUCLIDEAN INJECTIVE MATCHING; HM = HISTOGRAM MATCHING)

	Clustering	k	# training from $\mathbf{X}_l^S$	Overall Accuracy (OA)				Kappa statistics (κ)			
				1430		2860		1430		2860	
				μ	σ	μ	σ	μ	σ	μ	σ
Graph matching	k-means	50	$\mathbf{X}_l^S \rightarrow \mathbf{X}_l^* \rightarrow \mathbf{X}^{D_2}$	**80.72**	1.16	**81.35**	0.57	**0.685**	0.02	**0.695**	0.01
		100		78.49	1.38	78.99	0.88	0.647	0.02	0.655	0.01
	FCM	50		78.36	0.79	78.03	0.49	0.657	0.01	0.653	0.01
		100		78.51	1.00	78.06	0.32	0.652	0.01	0.647	0.01
	k-means	50	$\mathbf{X}_l^S \leftarrow \mathbf{X}^* \leftarrow \mathbf{X}^{D_2}$	78.39	1.68	79.82	0.85	0.654	0.02	0.674	0.01
		100		77.83	0.94	77.96	0.56	0.642	0.01	0.644	0.01
	FCM	50		77.18	0.86	77.54	0.36	0.640	0.01	0.645	0.01
		100		77.36	1.18	77.69	0.50	0.633	0.02	0.638	0.01
Eucl.	k-means	100	$\mathbf{X}_l^S \rightarrow \mathbf{X}_l^{*,\text{Eucl}} \rightarrow \mathbf{X}^{D_1}$	79.06	1.04	78.67	0.75	0.660	0.01	0.657	0.01
	k-means	100	$\mathbf{X}_l^S \rightarrow \mathbf{X}_l^{*,\text{E-Inj.}} \rightarrow \mathbf{X}^{D_1}$	70.84	1.11	71.33	0.87	0.545	0.02	0.550	0.01
	Histogr. matching		$\mathbf{X}_l^S \rightarrow \mathbf{X}_l^{*,\text{HM}} \rightarrow \mathbf{X}^{D_1}$	75.70	3.63	75.42	4.05	0.600	0.05	0.590	0.06
	No adaptation		$\mathbf{X}_l^S \rightarrow \mathbf{X}^{D_2}$	74.29	5.20	75.48	2.45	0.602	0.06	0.608	0.04
	Training from $\mathbf{X}^{D_2}$		$\mathbf{X}_l^{D_2} \rightarrow \mathbf{X}^{D_2}$	85.06	0.88	86.21	0.43	0.761	0.01	0.779	0.01

TABLE IV
RIETERPARK DATASET. CONFUSION MATRICES FOR A SINGLE RUN USING k-MEANS CLUSTERING, 1430 TRAINING PIXELS AND 50 CENTROIDS. (B = BUILDINGS; C = COMMERCIAL BUILDINGS; V = VEGETATION; H = HARVESTED VEGETATION; W = WATER).

(a) No adaptation

Predicted	True B	C	V	%PA
B	8247	2274	258	76.51
C	2041	1000	8	32.80
V	84	72	10460	98.53
H	2193	908	38	-
W	0	0	0	-
%UA	65.63	23.51	97.18	**71.44**

(b) Proposed

Predicted	True B	C	V	%PA
B	10377	2382	346	79.18
C	1981	1622	4	44.97
V	61	163	10399	97.89
H	146	87	15	-
W	0	0	0	-
%UA	82.59	38.13	96.61	**81.20**

(c) Training from $\mathbf{X}^{D_2}$

Predicted	True B	C	V	%PA
B	11587	2632	51	81.20
C	892	1594	19	63.63
V	86	28	10694	98.95
H	0	0	0	-
W	0	0	0	-
%UA	92.22	37.47	99.35	**86.56**

samples remain at the same relative position in the boundary between classes, and this can decrease the efficiency of a classifier based on support vectors and large margin.

In the next experiment, the effect of the clustering algorithm is illustrated in a continuous adaptation problem.

C. Hot spot correction

Figure 12 illustrates spectra in the Barrax image averaged for different types of crops: sugar beet (sparse vegetation), barley (senescent vegetation), alfalfa (dense vegetation), and bare soil. Panel (a) reports pixels out of the hot spot area, while panel (b) considers pixels within it. From these curves, the spatial component of the data shift effect can be appreciated, since the spectra vary much more in the hot spot area (differences between dashed and solid lines in the bottom panel). This created the necessity for a different quantization approach based on spatially constrained clustering: a compensation based solely on clustering in the spectral space would have been suboptimal, because certain angular effects between the two acquisitions are not related to the hot spot and must not be compensated in the same way as for the hot spot area. For instance, consider the bottom left corner of the images, where the bare soil fields are much more reflective in the orthogonal acquisition than in the *hot-spot* acquisition (which is a normal behavior of soil BRDF) showing a reversed effect with respect to the one we are trying to correct. For this reason, we applied a spectral/spatial approach, where the quantization was constrained to increase similarity in case of geographical proximity. To do that, geographical distance has been added to the spectral similarity matrix of the k-means (or FCM).

Results of the compensation are reported in Fig. 13: the first column shows a mosaic of the desired image $\mathbf{X}^D$ and of the image corrected with the proposed algorithm $\mathbf{X}^*$: on these mosaics[1], we can observe that most of the hot spot has been corrected by the graph matching procedure, especially when using the FCM algorithm. The k-means results show more pronounced differences in the corrected image, since all the pixels attributed to a same centroid are displaced with the same transform; on the contrary, the use of FCM allows a more gradual displacement, thus returning a smoother transform, where the pixels of the artifact are strongly corrected, while the correction strength decreases smoothly as long as we move further away. This can be observed in the central column of Fig. 13, where the difference between the hot spot image $\mathbf{X}^S$ and $\mathbf{X}^*$ is reported. Finally, in the rightmost column of Fig. 13 the per-pixel difference between $\mathbf{X}^D$ and $\mathbf{X}^*$ is illustrated: by comparing the resulting difference maps with the difference between the original acquisitions (in the right panel of top row), we can see that the graph matching procedure strongly reduces the angular effects, reducing the RMSE between the target domain $\mathbf{X}^D$ and the corrected image $\mathbf{X}^*$ form 5.98 (no adaptation) to 4.55 (k-means) and 4.19 (FCM).

---

[1]The mosaics were produced by creating a chessboard pattern showing $\mathbf{X}^D$ patches on black squares and $\mathbf{X}^*$ on white squares.
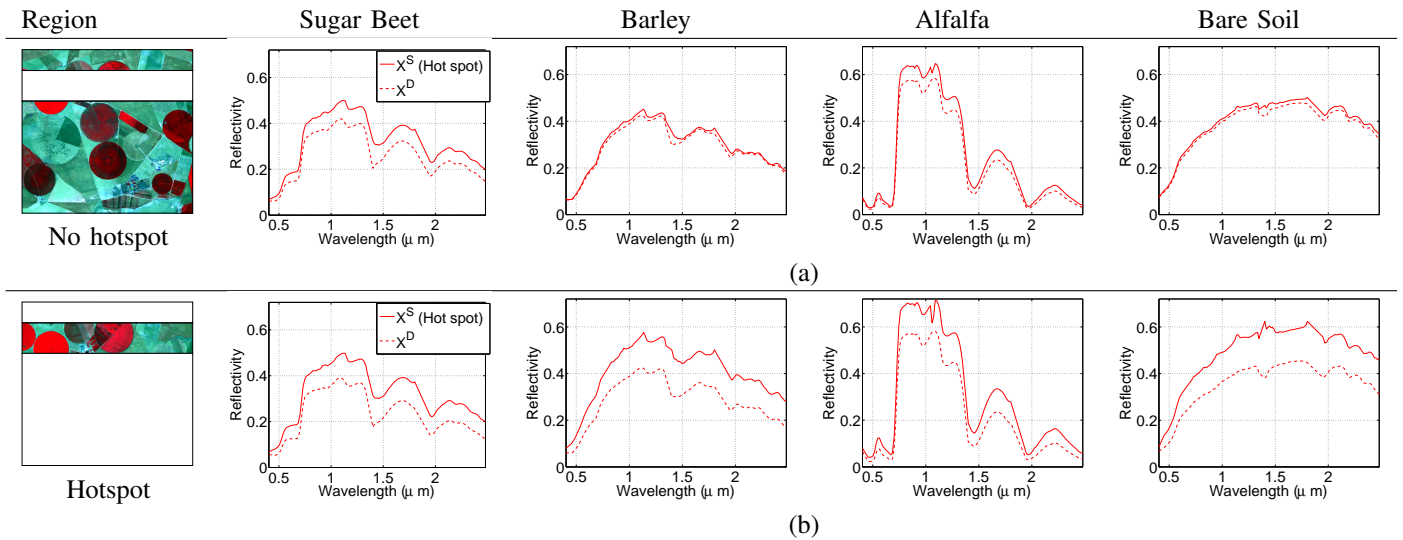
Fig. 12. Average spectra for different types of crops out of the hot spot area (a) and within (b). Solid lines represent averages for spectra in the image with hot spot $\mathbf{X}^S$ and dashed lines for the image without hot spot $\mathbf{X}^D$. In panel (b), differences between the two curves correspond to the shift to be corrected.

V. CONCLUSIONS

In this paper, we have presented a method based on graph matching for manifold adaptation and compensation of changes in acquisition conditions. The method matches similarity graphs, where the nodes are codebook vectors of the data distribution obtained by vector quantization. A cost function accounting for centroids displacements and graph deformation is proposed to ensure smooth and structure-preserving transforms. Experiments in a synthetic, controlled scenario, and in three challenging remote sensing scenarios, proved the interest of the method. In different settings for classification (multitemporal and spatially disconnected areas), the method proved its value as an efficient correction method for changes in shadowing, seasonal, and acquisition angle. It was also applied for the correction of angular effects with an example characterized by a strong hot spot during an acquisition orthogonal to the principal plane, where the method compensated efficiently the directional reflectance changes and spectral deformations occurred.

Despite the good performances observed, the method in its current version must be handled with care: the definition of the cross-domains graph is crucial, since it must contain the correct matching sequence among the possible matches. When those movements are not present, the algorithm cannot find a good adaptation mapping. A solution would be to use a fully connected cross-domains graph, but this would be done at the price of a high computational cost.

The presented examples illustrate the application of the proposed method and the use of graph matching in an elegant way to allow local compensation of manifold deformations. Future research will consider semisupervised extensions and displacements taking into account local properties of the destination manifold as curvature of principal directions.

ACKNOWLEDGEMENTS

REFERENCES

[1] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. Luis Rojo-Álvarez, and M. Martínez-Ramón, "Kernel-based framework for multi-temporal and multi-source remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, 2008.

[2] A. Mahtab, V.N. Sridhar, and R.R. Navalgund, "Impact of surface anisotropy on classification accuracy of selected vegetation classes: An evaluation using multidate multiangular MISR data over parts of Madhya Pradesh, India," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 250–258, 2008.

[3] G. Yang, C. Zhao, Q. Liu, W. Huang, and J. Wang, "Inversion of a radiative transfer model for estimating forest LAI from multisource and multiangular optical remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 988–1000, 2011.

[4] R. Duca and F. Del Frate, "Hyperspectral and multiangle CHRIS-PROBA images for the generation of land cover maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2857 – 2866, 2008.

[5] N. Longbotham, C. Chaapel, L. Bleiler, C. Padwick, W. J. Emery, and F. Pacifici, "Very high resolution multiangle urban classification analysis," *IEEE Trans. Geosci. Remote Sens.*, in press.

[6] W. A. Dorigo, "Improving the robustness of cotton status characterisation by radiative transfer model inversion of multi-angular CHRIS/PROBA data," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 1, pp. 18–29, 2012.

[7] F. Bovolo and L. Bruzzone, "A theoretical framework for unsupervised change detection based on change vector analysis in the polar domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, 2007.

[8] F. Bovolo, L. Bruzzone, and S. Marchesi, "Analysis and adaptive estimation of the registration noise distribution in multitemporal VHR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2658–2671, 2009.

[9] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, and Q. Du, "Multi-modal change detection, application to the detection of flooded areas: outcome of the 2009-2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 1, pp. 331–342, 2012.

[10] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset Shift in Machine Learning*, MIT Press, 2009.

[11] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *23rd AAAI Conf.*, Chicago, IL, USA, 2008, pp. 677–682.

[12] M. D. Fleming, J. S. Berkebile, and R. M. Hoffer, "Computer-aided analysis of LANDSAT-I MSS data: a comparison of three approaches, including a "modified clustering" approach," LARS information note 072475, Purdue University, 1975.

[13] M. Pax-Lenney, C. E. Woodcock, S. A. Macomber, S. Gopal, and C. Song, "Forest mapping with a generalized classifier and landsat TM data," *Remote Sens. Environ.*, vol. 77, no. 3, pp. 241–250, 2001.
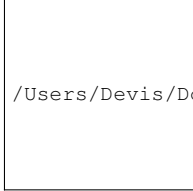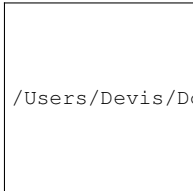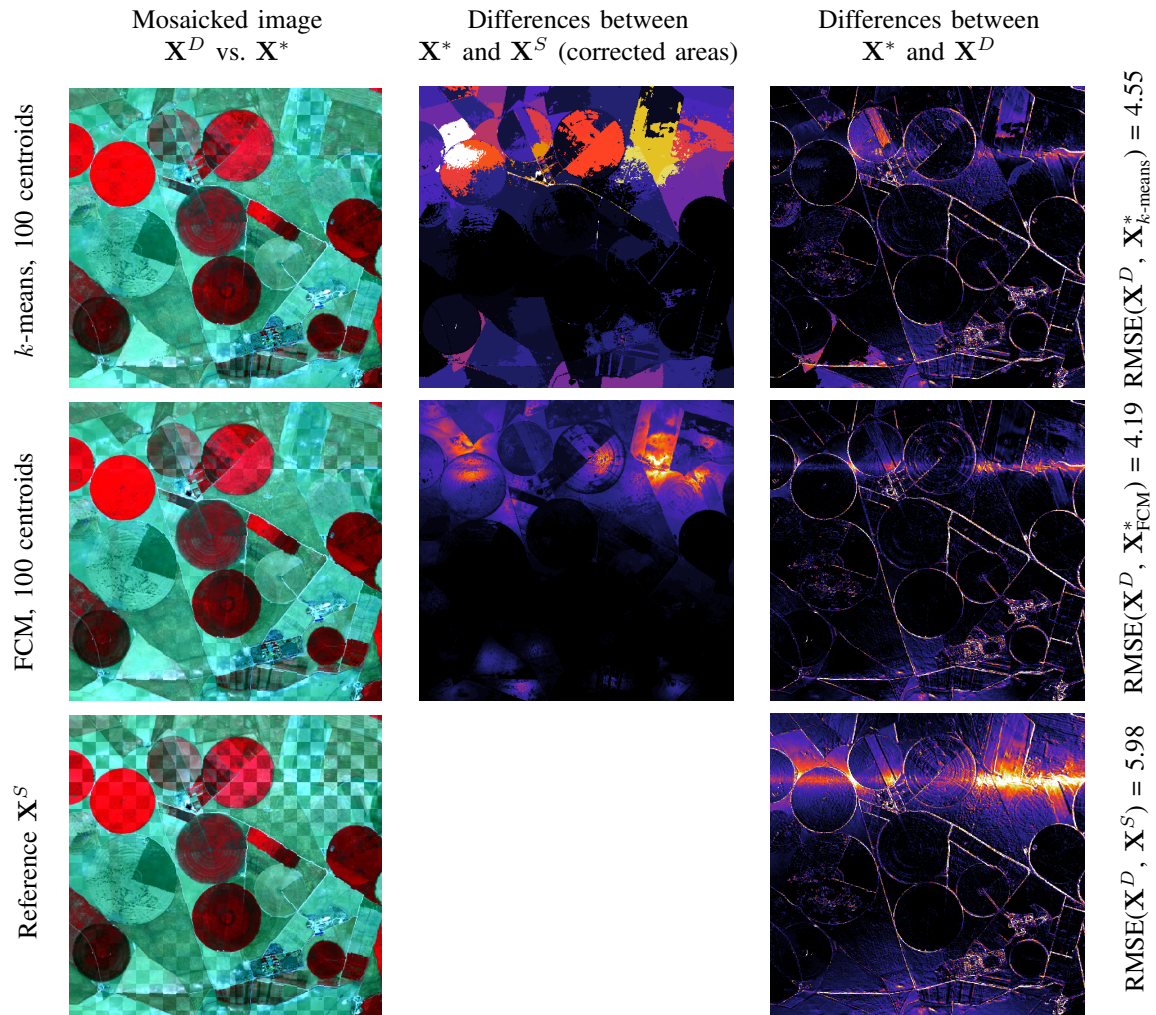
Fig. 13. Automatic correction of the DAISEX image: (left) mosaic superimposing the corrected image $\mathbf{X}^*$ and the target image $\mathbf{X}^D$; (middle) corrected areas, measured as the difference with respect to the hot spot image; (right) difference with respect to the target image.

[14] C. E. Woodcock, S. A. Macomber, M. P. Pax-Lenney, and W. B. Cohen, "Monitoring large areas for forest change using landsat: Generalization across space, time and landsat sensors," *Remote Sens. Environ.*, vol. 78, no. 1-2, pp. 194–203, 2001.

[15] X. Jia and J. A. Richards, "Cluster-space representation for hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 40 (3), pp. 593–598, 2002.

[16] G. M. Foody, D. S. Boyd, and M. E. J. Cutler, "Predictive relations of tropical forest biomass from landsat TM data and their transferability between regions," *Remote Sens. Environ.*, vol. 85, no. 5, pp. 463–474, 2003.

[17] F. Pacifici, M. Chini, and W.J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276–1292, 2009.

[18] L. Bruzzone and D. Fernandez-Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 456–460, 2001.

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

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

[21] L. Gómez-Chova, G. Camps-Valls, L. Bruzzone, and J. Calpe-Maravilla, "Mean map kernel methods for semisupervised cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 207–220, 2010.

[22] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sensing of Environment*, vol. 115, no. 9, pp. 2232–2242, 2011.

[23] B. Settles, "Active learning literature survey," Computer Sciences Technical Report 1648, University of Wisconsin-Madison, 2010.

[24] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Muñoz-Marí, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Proc.*, vol. 5, no. 3, pp. 606–617, 2011.

[25] F. Petitjean, J. Inglada, and P. Gancarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recogn.*, vol. 44, no. 3, pp. 678–693, 2011.

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

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

[28] J. A. Lee and M. Verleysen, *Nonlinear dimensionality reduction*, Springer, 2007.

[29] Y. Bengio, J. F. Paiement, P. Vincent, O. Delalleu, N. Le Roux, and M. Ouimet, "Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering," in *Proc. (NIPS)*, Cambridge, MA, 2004, MIT Press.

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

[31] C.M. Bachman, T.L. Ainsworth, and R.A. Fusina, "Improved manifold

coordinate representations of large-scale hyperspectral scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2786–2803, 2006.

[32] L. G. Shapiro and R.M. Haralick, "A metric for comparing relational descriptions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 7, no. 1, pp. 90–94, 1985.

[33] H. Bunke, "Error correcting graph matching: on the influence of the underlying cost function.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 9, pp. 917–922, 1999.

[34] B. Luo and R. Hancock, "Structural graph matching using the em algorithm and singular value decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1120–1136, 2001.

[35] A. K. C. Wong and M. You, "Entropy and distance of random graphs with application to structural pattern recognition," *IEEE Pattern Anal. Mach. Intell.*, vol. 7, no. 5, pp. 509–609, 1985.

[36] R. S. T. Lee and J. N. K. Liu, "Tropical cyclone identification and tracking system using integrated neural oscillatory elastic graph matching and hybrid RBF network track mining techniques," *IEEE Trans. Neural Net.*, vol. 11, no. 3, pp. 680–689, 2000.

[37] T. S. Caetano, T. Caelli, D. Schuurmans, and D.A.C. Barone, "Graphical models and point pattern matching," *IEEE Trans. Pattern Anal. Mach, Intell.*, vol. 28, no. 10, pp. 1646–1663, 2006.

[38] J. J. McAuley and T. S. Caetano, "Fast matching of large point sets under occlusion," *Pattern Recogn.*, in press.

[39] T. S Caetano, J. J. McAuley, L. Cheng, Q. V. Le, and A. J. Smola, "Learning graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1048–1058, 2009.

[40] N. Kambhatla and T. Leen, "Dimension reduction by local PCA," *Neur. Comp.*, vol. 9, pp. 1493–1500, 1997.

[41] J. Karhunen, S. Malaroiu, and Mika Ilmoniemi, "Local linear independent component analysis based on clustering," *Intl. J. Neur. Syst.*, vol. 10, no. 6, December 2000.

[42] J. Malo and J. Gutiérrez, "V1 non-linear properties emerge from local-to-global non-linear ICA," *Network: Computation in Neural Systems*, vol. 17, no. 1, pp. 85–102, 2006.

[43] V. Laparra, D. Tuia, S. Jimenez, G. Camps-Valls, and J. Malo, "Principal polynomial analysis for remote sensing data processing," in *IEEE Intl. Geosci. Rem. Sens. Symposium*, July 2011.

[44] V. Laparra, S. Jimenez, G. Camps, and J. Malo, "Nonlinearities and adaptation of color vision from sequential principal curves analysis," *Accepted in Neural Comput.*, 2012.

[45] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, January 1982.

[46] Luis Gómez-Chova, Luis Alonso, Luis Guanter, Gustavo Camps-Valls, Javier Calpe, and José Moreno, "Correction of systematic spatial noise in push-broom hyperspectral sensors: application to CHRIS/PROBA images," *Applied Optics*, vol. 47, no. 28, pp. F46–F60, Oct 2008.

[47] C. Simmer and S.A.W. Gerstl, "Remote sensing of angular characteristics of canopy reflectances," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-23, no. 5, pp. 648 –658, sept. 1985.

[48] F. Maignan, F.M. Bréon, and R. Lacaze, "Bidirectional reflectance of earth targets: Evaluation of analytical models using a large set of spaceborne measurements with emphasis on the hot spot," *Remote Sensing of Environment*, vol. 90, no. 2, pp. 210–220, 2004.

[49] U. Beisl, "A new method for correction of bidirectional effects in hyperspectral imagery," in *International Workshop in Spectroscopy Application in Precision Farming*, 2001.

[50] L. I. Kuncheva, *Combining pattern classifiers*, Wiley, NJ, 2004.

[51] G.J. Briem, J.A. Benediktsson, and J.R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2291–2300, 2002.

[52] J. A. Benediktsson and J. R. Sveinsson, "Multisource remote sensing data classification based on consensus and pruning," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 932–936, 2003.

[53] B. Waske, S. van der Linden, J. A. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, 2010.

[54] D. Coltuc, P. Bolon, and J-M. Chassery, "Exact histogram specification," *IEEE Trans.Image Proc.*, vol. 15, no. 5, pp. 1143–1152, 2006.

[55] D. Tuia, F. Ratle, A. Pozdnoukhov, and G. Camps-Valls, "Multi-source composite kernels for urban image classification," *IEEE Geosci. Remote Sens. Letters*, vol. 7, no. 1, pp. 88–92, 2010.

[56] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, "Supervised change detection in VHR images using contextual information and support vector machines," *Int. J. Appl. Earth Obs. Geoinf.*, in press.

Devis Tuia (SM'07, M'09) was born in Mendrisio, Switzerland, in 1980. He received a diploma in Geography at the University of Lausanne (UNIL) in 2004, a Master of Advanced Studies in Environmental Engineering at the Federal Institute of Technology of Lausanne (EPFL) in 2005 and a Ph.D in environmental sciences at UNIL in 2009. He was a postdoc researcher at both the University of Valéncia, Spain and the Univeristy of Colorado at Boulder under a Swiss National Foundation program. He is now senior research associate at the LaSIG laboratory, EPFL. His research interests include the development of algorithms for information extraction and classification of very high resolution remote sensing images using machine learning algorithms. Visit http://devis.tuia.googlepages.com/ for more information.

Jordi Muñoz-Marí (M'11) was born in València, Spain in 1970, and received a B.Sc. degree in Physics (1993), a B.Sc. degree in Electronics Engineering (1996), and a Ph.D. degree in Electronics Engineering (2003) from the Universitat de València. He is currently an Associate Professor in the Electronics Engineering Department at the Universitat de València, where teaches Electronic Circuits and, Programmable Logical Devices, Digital Electronic Systems and Microprocessor Electronic Systems. His research interests are mainly related with the development of machine learning algorithms for signal and image processing. Visit http://gpds.uv.es/~jordi/ for more information.

Luis Gómez-Chova (S'08-M'09) received the B.Sc. (with first-class honors), M.Sc., and Ph.D. degrees in electronics engineering from the University of Valencia, Valencia, Spain, in 2000, 2002 and 2008, respectively. Since 2000, he has been with the Department of Electronics Engineering, University of Valencia, first enjoying a research scholarship from the Spanish Ministry of Education and currently and currently as an Assistant Professor. He is also a researcher at the Image Processing Laboratory (IPL), where his work is mainly related to pattern recognition and machine learning applied to remote sensing multispectral images and cloud screening. Visit http://www.uv.es/chovago for more information.

Jesus Malo was born in 1970. He received the M.Sc. and Ph.D. degrees in physics from the Universitat de València, Spain, in 1995 and 1999, respectively. He worked as a Fulbright Post-Doctoral Fellow at the Vision Group, National Aeronautics and Space Administration Ames Research Center, Mountain View, CA, in 2000 and 2001 (with A. B. Watson), and the Laboratory of Computational Vision, Center for Neural Science, New York University, New York (with E. P. Simoncelli). Currently, he is with the Image Processing Laboratory, Universitat de València. His current research interests include models of low-level human vision, their relations with information theory, and their applications to image processing and vision science experimentation, Fourier, MATLAB, Equipo Crónica, Jim Jarmusch, Jordi Savall, Pixies, Manara, la Bola de Cristal, Faemino y Cansado, and beauty in general. Dr. Malo is a member of the Asociación de Mujeres Investigadoras y Tecnólogas, and an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING. He was the recipient of the Vistakon European Research Award in 1994.