Semisupervised Manifold Alignment of Multimodal Remote Sensing Images

Devis Tuia, *Member, IEEE*, Michele Volpi, *Student Member, IEEE*, Maxime Trolliet, and Gustau Camps-Valls, *Senior Member, IEEE*

Abstract

We introduce a method for manifold alignment of different modalities (or domains) of remote sensing images. The problem is recurrent when a set of multitemporal, multisource, multisensor and multiangular images is available. In these situations, images should ideally be spatially coregistred, corrected and compensated for differences in the image domains. Such procedures require the interaction of the user, involve tuning of many parameters and heuristics, and are usually applied separately. Changes of sensors and acquisition conditions translate into shifts, twists, warps and foldings of the image distributions (or manifolds). The proposed semisupervised manifold alignment (SS-MA) method aligns the images working directly on their manifolds, and is thus not restricted to images of the same resolutions, either spectral or spatial. SS–MA pulls close together samples of the same class while pushing those of different classes apart. At the same time, it preserves the geometry of each manifold along the transformation. The method builds a linear invertible transformation to a latent

Manuscript received XXXX;

This work has been partly supported by the Swiss National Science Foundation (grants PZ00P2-136827 (http://p3.snf.ch/project-136827) and P2LAP2-148432 (http://p3.snf.ch/Project-148432)) and by the spanish Ministry of Economy and Competitiveness (MINECO) under project LIFE-VISION TIN2012-38102-C03-01.

DT and MT are 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.

MV was with the Centre for Research on Terrestrial Environment, Université de Lausanne, Switzerland. He is now with the CALVIN at the Institute of Perception, Action and Behaviour, the University of Edinburgh, United Kingdom. Telephone: +44 (0) 131 650 8741, Fax: +44 (0) 131 651 5651, Email: michele.volpi@ed.ac.uk, Web: https://sites.google.com/site/michelevolpiresearch.

GCV is with the Image Processing Laboratory (IPL), Universitat de València, C/ Catedrático A. Escardino, 9 - 46980 Paterna, València (Spain). gustavo.camps@uv.es, http://isp.uv.es, Phone: +34-963544064, Fax: +34-963543261.

PREPRINT

space where all images are alike, and reduces to solving a generalized eigenproblem of moderate size. We study the performance of SS–MA in toy examples and in real multiangular, multitemporal, and multisource image classification problems. The method performs well for strong deformations and leads to accurate classification for all domains. A Matlab implementation of the proposed method is provided at http://isp.uv.es/code/ssma.htm.

Index Terms

Feature extraction, Graph-based methods, Very high resolution, Domain adaptation, Multiangular, Multitemporal, Multisource, Classification

I. Introduction

Remote sensing analysts can nowadays exploit images of unprecedented spatial resolution (e.g. higher than the meter for commercial sensors such as WorldView 2) with increasing revisit time (which will be increased with the upcoming Sentinels) [1], [2]. With these very high resolution (VHR) images, it becomes possible to perform a large variety of monitoring studies, since the geographical area of interest can be covered with high frequency. This led to a variety of techniques for multitemporal [3], [4] and multiangular [5], [6] data processing for VHR data.

The increased access to information also induces a series of processing problems for the analyst performing the monitoring task: the images undergo a series of spectral distortions related to variations in the acquisition geometry, the atmospheric conditions, and the acquisition angle. To these acquisition-related differences, we must add a series of temporal specificities, such as the variations in the phenological cycle, or the appearance of small objects contaminating the class signature (flower pots on roofs, for example). All these changes in acquisition conditions and geometry, as well as differences in the properties of the sensors, produce local changes in the probability distribution function (PDF) of the images from one acquisition to another one, which in turn affect the performances of the classifier, when predicting data from another domain. As a consequence, the tempting direct application of a classifier optimal for one scene to another scene can lead to catastrophic results. The application of a classifier to a series of newly acquired scenes remains a priority for image analysts, who would like to reuse the available labeled samples, minimize the time/effort devoted to photointerpretation (or terrestrial campaigns) and eventually use images acquired by other sensors. Therefore, a classifier should

PREPRINT

be applicable to new scenes regardless of the sensor and acquisition specificities. This property of *portability* is of the greatest importance, especially in the VHR context or when considering archives of images, for which labels are often unavailable.

To meet this objective, research has considered strategies of *adaptation*. By adaptation, we mean the ability of a method to modify its characteristics to new scenes acquired under different acquisition conditions, but representing a similar problem (typically sharing the same classes). Adaptation has been carried out in three main ways in remote sensing data processing: 1) at the level of the classifier, 2) to encode invariances of interest, and 3) at the level of the data representation. Let us briefly review the three families of approximations.

Regarding the adaptation of the classifier, most strategies are issued from semi-supervised learning: the adaptation is usually performed either by modifying the weights of the classifier using unlabeled data coming from the PDF of the image to be classified [7]–[9], by spatial regularization [10], or by adding few informative labeled examples carefully chosen from the new image [11], [12]. This family of approaches often comes with several free parameters, requires expertise in machine learning and statistics, and involves high computational costs.

When considering the incorporation of invariances in the classifier, one aims at making the classifier robust to variations in the data representation. In the specific example of remote sensing images, such variations can be rotations (objects can be arbitrarily oriented), presence of shadows (attenuations of the signal), and scale (objects of the same class can have different sizes, or the new image may have different spatial resolution), just to name a few. Invariance to these changes can be achieved in several ways. For some of these effects, one may develop physical models [13], for example to remove atmospheric of illumination effects. This, however, requires the accurate modeling of the physical processes involved and a detailed knowledge of the atmospheric conditions at the time of acquisition. Alternatively, the classifier can become invariant if one includes in the training set synthetic examples representing the phenomena that the classifier should be invariant to. For example, if using patches of the images as inputs, adding rotated versions of them will force invariance to rotation of the objects represented in the patches [14]. While good results are obtained in general, it becomes complex to encode several invariances simultaneously that cover a reasonable subspace of the image manifold.

The third family is the one considered in this paper. The rationale is to modify the data representation to perform the adaptation. Such a modification can be driven by physical properties

PREPRINT

of the atmosphere [13], [15], by a data-driven feature extraction [16], [17], or both [18]. In all cases, however, the aim is to find a common data representation, in which the data are more similar (or *aligned*) to each other [19], independently from the processing steps that will follow. Since the data representation itself is modified, it is then possible to use any classifier on the aligned data.

This type of adaptation can be seen in a global or local perspective: in the first case, the PDFs of all images are compared and aligned, either by matching the histograms [20], [21] or by projecting the data distributions onto a common space using, for instance, Principal Component Analysis (PCA, [22]), Canonical Correlation Analysis (CCA [16]) or other more advanced nonlinear kernel multivariate methods [23].

In the case of local matching, the point clouds are matched regionally, in order to account for deformations that affect only certain parts of the image manifold. Local methods generally either minimize a cost function bringing the nodes of two graphs closer while maintaining the local neighborhoods in the projected space unchanged [24]–[26] or consider mixtures of local models [27], which are then merged in a global representation [28]–[31]. In the remote sensing literature, an application of the first of these two approaches is found in [17], where an iterative method is used to match graphs of the same sensor under angular and temporal variations.

A desirable alignment method must be capable of aligning images 1) under strong distortions, 2) of different sensors and 3) not co-registered. The methodologies so far proposed in remote sensing literature do not address all these issues simultaneously: PCA does not require co-registration, but fails in multisensor scenarios and under strong distortions. On the contrary, (K)CCA is naturally multisensor and works under strong distortions, but requires strict co-registration. Table I summarizes these properties for a series of alignment methodologies found in remote sensing.

In this paper, we tackle these problems simultaneously by proposing a linear, efficient method for feature extraction and dataset alignment issued from the manifold alignment literature [19], [32]. The main idea of the method is to use labeled samples from both domains to bring the manifolds closer, while keeping their respective inherent structure unchanged using two proximity graphs built with unlabeled samples [33]. Therefore, the method includes constraints on the local manifold geometry in the aligned space. The alignment transformation is defined by a linear projection function that depends both on labeled and unlabeled samples, making the method

PREPRINT

semi-supervised. In the aligned space, pixels from both domains can be used simultaneously and a classifier effective in all domains can be learned. We will refer to this method as *semi-supervised manifold alignment* (SS–MA).

SS–MA is effective for large deformations, since it is not based on inter-graphs distances, as the methods in [17], [34]. It does not require co-registration of the image sources as [16], [35], and is naturally multisource, as it can align images of different dimensionality, unlike standard domain adaptation algorithms [8], [9]. Since the eigenvectors are defined with a discriminative term and are sorted by their eigenvalues, the classifier can then be learned using only the first dimensions: this makes SS–MA an interesting solution also for dimensionality reduction. The price to pay is the availability of some (typically few) labeled pixels in each domain.

The remainder of the paper is organized as follows. Section II presents the proposed methodology and the manifold alignment algorithm. Section III presents the VHR data used in the experiments, that are detailed and discussed in Section IV. We will illustrate the method in several problems: 2D toy datasets under different deformations, as well as multiangular adaptation for the same sensor, multitemporal and multisource image adaptation. Section V concludes the paper with some discussion and further work.

II. Semisupervised manifold alignment

The main idea of the proposed method is to align the data manifolds by projecting them into a common representation, or *joint latent space*, $\mathcal{F}$. Such space has two desirable properties: 1) it preserves the local geometry of each dataset and 2) it brings regions belonging to the same class closer together, while pushing those belonging to different classes apart. The method searches for a set of projection functions, one *per* image, that achieve this double objective [33]. Figure 1 illustrates the two properties and the expected alignment.

A. Notation

Consider a series of M images (or *domains*) and their corresponding data matrices $\mathbf{X}^m$, $m = 1, \ldots, M$. Each matrix $\mathbf{X}^m$ contains a large set of unlabeled samples $\{\mathbf{x}_i^m\}_{i=1}^{u_m}$, and some input-output labeled sample pairs $\{\mathbf{x}_j^m, y_j^m\}_{j=1}^{l_m}$, typically with $l_m \ll u_m$. To fix notation, examples are column vectors and the data matrices are $\mathbf{X}^m \in \mathbb{R}^{d_m \times n_m}$ where $n_m = l_m + u_m$. These matrices contain both labeled and unlabeled points. Let us additionally denote the block diagonal

PREPRINT

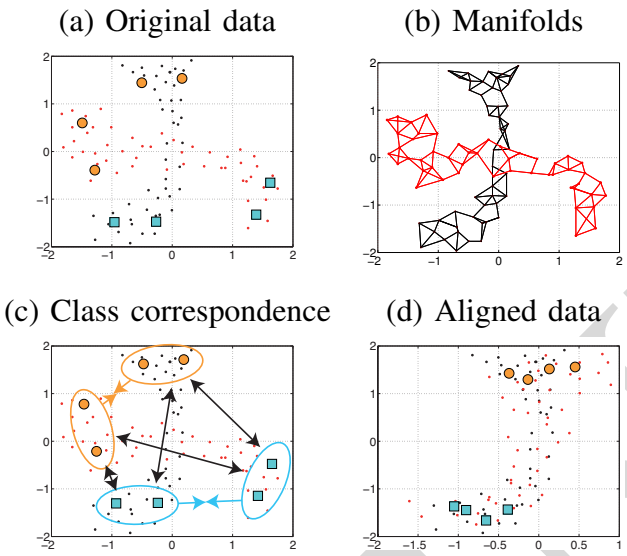


Fig. 1. Ingredients of the proposed algorithm: (a) two datasets (● and ●) distorted by a 90° rotation and labeled pixels of two classes (● and ■); (b) Geometrical structures of each manifold to be preserved (G term, Eq. (2)); (c) Label similarity induced attraction (orange and cyan arrows, S term, Eq. (3)) and repulsion (black arrows, D term, Eq. (4)); (d) aligned dataset.

data matrix containing all samples $\mathbf{X}^m$, $m = 1, \ldots, M$, as $\mathbf{X} = diag(\mathbf{X}_1, \ldots, \mathbf{X}_M) \in \mathbb{R}^{d \times N}$, $N = \sum_m n_m$, $d = \sum_m d_m$. It is important to stress here that the images do not necessarily represent the same location and can be acquired by different sensors with different spatial and spectral resolutions, so in principle we may have different numbers of samples per data matrix, $n_m \neq n_{m'}$, and different dimensions in each domain, i.e. $\mathbf{x}^m \in \mathbb{R}^{d_m}$ possibly with $d_m \neq d_{m'}$, $\forall m, m' = 1, \ldots, M$.

B. Semi-supervised loss function

The problem of aligning the M-images dataset to a common representation boils down to constructing M mapping functions to $\mathcal{F}$ by means of M projection matrices, $\mathbf{f}^m \in \mathbb{R}^{d_m \times d}$, $m = 1, \ldots, M$. The common latent space $\mathcal{F}$ is of dimension $d = \sum_{m=1}^{M} d_m$. Mapping to $\mathcal{F}$ requires that samples belonging to the same class become closer, while those of different classes are pushed far apart. Moreover, the mapping should preserve the geometry of each data manifold. Since we have terms to be minimized and others to be maximized, the problem reduces to solving

PREPRINT

a standard Rayleigh quotient[1]:

$$\mathbf{F}^{opt} = \arg\min_{\mathbf{F}} \{\text{tr}((\mathbf{F}^\top \mathbf{B}\mathbf{F})^{-1}\mathbf{F}^\top \mathbf{A}\mathbf{F})\} \quad (1)$$

where $\mathbf{F}$ is a $d \times d$ projection matrix, whose row-blocks correspond to the $d_m \times d$ domain-specific projection functions, $\mathbf{f}^m$. $\mathbf{A}$ is a matrix that contains the quantity of data relations to be decreased by the projection, and $\mathbf{B}$ is a matrix that contains the quantity of data relations to be increased. In our case, we want find the set of projectors $\mathbf{F}^{opt}$ that maximizes the distances between samples of different classes, represented by an affinity matrix $\mathbf{D}$ (thus $\mathbf{B} = \mathbf{D}$) and at the same time minimizes the distances between samples that are either close in each manifold (affinity matrix $\mathbf{G}$) or of the same class (affinity matrix $\mathbf{S}$). These two last matrices are combined as $\mathbf{A} = \mu\mathbf{G} + \mathbf{S}$, where μ is a tradeoff parameter. $\mathbf{A}$ favors discriminant projections (induced by $\mathbf{S}$) and at the same time preserves the original geometry of each manifold (summarized in $\mathbf{G}$). The effect of the three terms $\mathbf{G}, \mathbf{S}$ and $\mathbf{D}$ is illustrated in Fig. 1. The three matrices $\mathbf{D}$, $\mathbf{S}$ and $\mathbf{G}$ will be approached by graphs Laplacians, as detailed below.

Let us now describe the computation of the corresponding scalar terms G, S, and D, which can be interpreted as the global quantities to be minimized (for $\mathbf{S}$ and $\mathbf{G}$), respectively maximized (for $\mathbf{D}$), by the optimization.

The term G ensures that the geometry of each manifold is preserved in the transform. For this reason it does not act among the sources, but only within each source. To this end, local similarity matrices $\mathbf{W}_g^m \in \mathbb{R}^{n_m \times n_m}$ need to be defined. In this paper, similarity matrices are built using standard graphs, such as the kNN or ϵ-ball [36], computed for each domain m separately. Entries of the matrices are $W_g^m(i,j) = 1$ if samples $\mathbf{x}_i$ and $\mathbf{x}_j$ from domain m are neighbors in the kNN graph of that domain and 0 otherwise. The term ensures that neighboring samples in the original domains are mapped close in the projected space[2]. No geometric relation between domains are considered, the geometry is preserved solely within each graph. The geometry term reduces to:

[1]Also known as the Rayleigh–Ritz ratio, and widely used in the field of Fisher's discriminant analysis.

[2]The term acts as a form of graph regularization enforcing the smoothness (or manifold) assumption. Intuitively, this is equivalent to penalize "rapid changes" of the projection function $\mathbf{f}^m$ evaluated between samples that are close in the graphs $\mathbf{W}_g^m$.

PREPRINT

$$G = \sum_{m=1}^{M} \sum_{i,j=1}^{n_m} W_g^m(i,j) \|\mathbf{f}^{m\top}\mathbf{x}_i^m - \mathbf{f}^{m\top}\mathbf{x}_j^m\|^2$$
$$= \text{tr}(\mathbf{F}^\top \mathbf{X}\mathbf{L}_g\mathbf{X}^\top\mathbf{F}) \quad (2)$$

where $\mathbf{f}^m$ are the functions projecting the domain m in the latent space $\mathcal{F}$ and $\mathbf{L}_g = diag(\mathbf{L}_g^1, \ldots, \mathbf{L}_g^M)$ is a block-diagonal matrix containing the domain-specific graph Laplacians, reflecting the geometric similarity in each domain. For domain m, the corresponding Laplacian is $\mathbf{L}_g^m = \mathbf{U}_g^m - \mathbf{W}_g^m$, with degree matrix $\mathbf{U}_g^m(i,i) = \sum_j \mathbf{W}_g^m(i,j)$.

The term S enhances class similarity between the labeled instances of all domains. Its role is to pull labeled samples of the same class close together. To this end, we use a matrix of class-similarities, $\mathbf{W}_s = [\mathbf{W}_s^{(m,m')}]$, $m, m' = 1, \ldots, M$, where the components of each $\mathbf{W}_s^{m,m'} \in \mathbb{R}^{n_m \times n_m}$ are $W_s^{m,m'}(i,j) = 1$ for samples with the same label and 0 otherwise (including unlabeled data). Minimizing S for all the image sources simultaneously corresponds to minimize the distance among all samples of the same class in the latent space:

$$S = \sum_{m,m'=1}^{M} \sum_{i,j=1}^{l_m, l_{m'}} W_s^{m,m'}(i,j) \|\mathbf{f}^{m\top}\mathbf{x}_i^m - \mathbf{f}^{m'\top}\mathbf{x}_j^{m'}\|^2$$
$$= \text{tr}(\mathbf{F}^\top \mathbf{X}\mathbf{L}_s\mathbf{X}^\top\mathbf{F}), \quad (3)$$

where $\mathbf{L}_s$ is the corresponding joint graph Laplacian to $\mathbf{W}_s$. What we want to achieve is to map all samples of the same class close to each other, independently from the source they come from. Note that this term considers relations of labeled samples in different domains.

The dissimilarity term D encodes the opposite behavior. By maximizing D, one tries to pull samples belonging to different classes apart from each other. To this end, we employ a matrix of class-dissimilarities, $\mathbf{W}_d = [\mathbf{W}_d^{(m,m')}]$ where the components of each $\mathbf{W}_s^{m,m'} \in \mathbb{R}^{n_m \times n_m}$ are $W_s^{m,m'}(i,j) = 1$ for samples of different classes and 0 otherwise (including unlabeled data). The dissimilarity term is:

$$D = \sum_{m,m'=1}^{M} \sum_{i,j=1}^{l_m, l_{m'}} W_d^{m,m'}(i,j) \|\mathbf{f}^{m\top}\mathbf{x}_i^m - \mathbf{f}^{m'\top}\mathbf{x}_j^{m'}\|^2$$
$$= \text{tr}(\mathbf{F}^\top \mathbf{X}\mathbf{L}_d\mathbf{X}^\top\mathbf{F}), \quad (4)$$

PREPRINT

where $\mathbf{L}_d$ is the corresponding joint graph Laplacian for $\mathbf{W}_d$. The term is maximized when labeled samples of different classes are mapped far from each other in the latent space.

C. Projection functions

All the $\mathbf{W}$ matrices ($\mathbf{W}_g$, $\mathbf{W}_s$ and $\mathbf{W}_d$) and corresponding graph Laplacians are of size $N \times N$. Now, by plugging Eqs. (2), (3) and (4) into (1), we obtain:

$$\min_{\mathbf{F}} \{\text{tr}((\mathbf{F}^\top \mathbf{X}\mathbf{L}_d\mathbf{X}^\top\mathbf{F})^{-1}\mathbf{F}^\top\mathbf{X}(\mu\mathbf{L}_g + \mathbf{L}_s)\mathbf{X}^\top\mathbf{F})\}. \tag{5}$$

The solution of the minimization problem in Eq. (1) is given by the eigenvectors $\boldsymbol{\varphi}_i$ corresponding to the smallest eigenvalues of the following generalized eigenvalue decomposition [33]:

$$\mathbf{X}(\mu\mathbf{L}_g + \mathbf{L}_s)\mathbf{X}^\top\boldsymbol{\varphi} = \lambda\mathbf{X}\mathbf{L}_d\mathbf{X}^\top\boldsymbol{\varphi}. \tag{6}$$

The optimal matrix $\mathbf{F}^{opt}$ contains the projectors from the original spaces to the joint latent space in row blocks:

$$\mathbf{F}^{opt} = \left[\sqrt{\lambda_1}\boldsymbol{\varphi}_1 | \ldots | \sqrt{\lambda_d}\boldsymbol{\varphi}_d\right] = \begin{bmatrix} \mathbf{f}_1^1 & \ldots & \mathbf{f}_d^1 \\ \mathbf{f}_1^2 & \ldots & \mathbf{f}_d^2 \\ \vdots & \vdots & \vdots \\ \mathbf{f}_1^M & \ldots & \mathbf{f}_d^M \end{bmatrix} \tag{7}$$

By looking at the structure of the $\mathbf{F}^{opt}$ matrix, we observe that each row block contains the projection function from domain m to the latent space. Each domain can thus be projected to the joint space by simple matrix multiplication (as in any multivariate analysis method like PCA). Projecting a data matrix in domain m, $\mathbf{X}_*^m$ to the d-dimensional latent space $\mathcal{F}$ just involves:

$$\mathcal{P}_f(\mathbf{X}_*^m) = \mathbf{f}^{m\top}\mathbf{X}_*^m.$$

D. Properties

The method has the following properties:

a) *Linearity*: the method defines explicitly the projection functions, which can in turn be used to project large datasets into the latent space explicitly optimized for joint classification of all the sources.

b) *Multisensor*: since it only exploits the geometry of each manifold separately, there is no restriction on the number of bands to be aligned nor on their properties.

PREPRINT

TABLE I

PROPERTIES OF THE DATA REPRESENTATION METHODS.

Method	No labeled pixels in targets	Multisensor	No Co-registration	More than 2 images simult.	Invertible
PCA	✓	×	✓	×	✓
k-PCA	✓	×	✓	×	×
TCA	✓	×	✓	×	×
CCA	✓	✓	×	✓	✓
k-CCA	✓	✓	×	✓	×
Graph matching [17]	✓	×	✓	×	✓
SS–MA	×	✓	✓	✓	✓

c) *Multidomain*: as for CCA, the method can align an arbitrary number of domains in a common latent space. It does not necessarily require (as in the experiments below) a leading source domain to which all the others are aligned to (contrarily to PCA, TCA and graph matching [17]).

d) *PDF-based*: the method aligns directly the PDFs of the sources, without requiring co-registered samples. Unlike CCA, it can thus align images, which are unregistered, of different areas or of different spatial resolutions.

e) *Invertibility*: using the projection functions defined, it is possible to synthesize the bands of one sensor from images acquired with another one. This opens many possibilities for the design and evaluation of new instruments.

These properties are also summarized and compared to other feature extractors used to align domains in Table I.

E. Computational complexity

The proposed method reduces to solving a generalized eigenvalue problem of relatively small size, involving $d \times d$ matrices. The problem is linear with the number of data sources (in our case, images). Most of the computational effort is associated to the construction of the graph Laplacian

PREPRINT

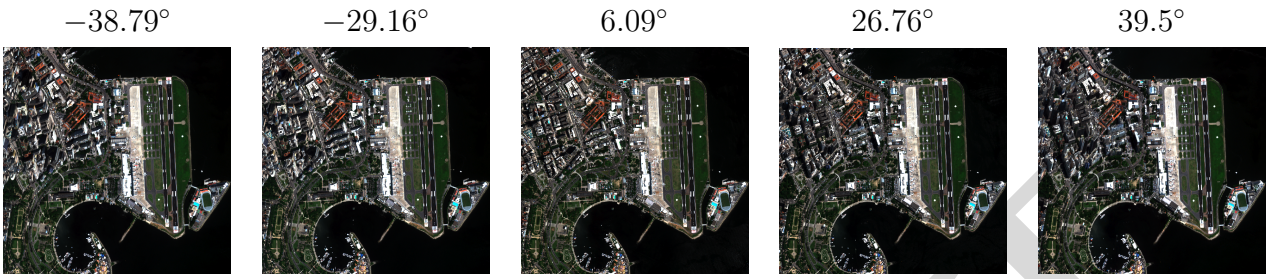


Fig. 2. The five images used in the multiangular experiments.

$\mathbf{L}_g$ involved in the geometry-preserving term G, because the two other graph Laplacians $\mathbf{L}_s$ and $\mathbf{L}_d$ are constructed with a simple convolution between the vectors of labels, and hence the cost of their construction is negligible. Dealing with graph Laplacians with many vertices and edges can be computationally demanding for two main reasons: the calculation/storage of the matrix itself and the cost of operations such as eigendecomposition and inversion when dealing with these large matrices.

The memory requirements cannot be solved easily, but strategies to decrease the size of the matrices exist, such as intelligent node selection via landmark points [37], active learning [38], inclusion of additional constraints to regularize the graph spatially [39] or graph partitioning and sparsification [40]. Efficient solvers to construct the graph Laplacian have been recently proposed: the solver proposed in [41] scales linearly with the number of edges in both run time and storage. Also note that in SS–MA the graph Laplacian is computed only once and its calculation is performed off-line.

The second problem (the cost of the eigendecomposition) is not critical for SS–MA: since Eq. (5) decomposes an eigenproblem of relatively small size ($d \times d$), there is no need to design specific strategies to reduce the size of the problem (by sample/nodes selection) or to improve the efficiency of the algebraic process. Note however that iterative methods can be used to reduce the computational load of the eigendecomposition to a logarithmic complexity (which is otherwise $\mathcal{O}(d^3)$ for our $d \times d$ matrix): examples can be found in [42], [43].

III. Data and setup

This section describes the data and the setups used in the experiments presented in Section IV.

PREPRINT

A. Datasets and evaluated scenarios

The proposed SS–MA is tested in four scenarios:

- *2D Toy dataset under different deformations*: the first example is designed to test the robustness of the proposed SS-MA transformation to different deformations of different levels of complexity. To do so, we generated toy datasets composed by two spirals embedded in a 2D space (red and blue spirals on the first column of Fig. 4). The red spiral remains unchanged, while the blue one is either 1) scaled (S), 2) scaled and rotated (S + R), and 3) scaled, rotated and translated (S + R + T). The two domains are composed of 2000 labeled examples each, whose classes are illustrated on different colors on the second column of the figure.
- *Multiangular adaptation*: the first real experiment considers portability of a model along a series of five angular views of the city center of Rio de Janeiro, Brazil, acquired by WorldView-2 [44]. The interest of this dataset is that the images are acquired at a very short time interval and capture the same area, so that the spectral deformations should be only due to angular effects, since the atmosphere is unchanged along the acquisitions. Figure 2 shows an RGB composite of the five images. The number of labeled pixels available for the twelve classes available in each domain is detailed in Tab. II.
- *Multitemporal adaptation for the same sensor*: this is the classical scenario for domain adaptation in which, given one image with sufficient ground truth, we want to apply the best possible classifier for that image to other images taken at different times and places and with minimal effort in photointerpretation (i.e. the least number of new labeled pixels). The images in this experiment are three WorldView-2 scenes extracted from two images of the city of Lausanne, Switzerland: the Montelly and Malley images are subsets of images acquired the 29th of September 2010, while the Prilly subset is part of a scene acquired the 2nd of August 2011. All scenes have been pansharpened using the Gram-Schmidt transform. Figure 3 illustrates the RGB composite and the exhaustive ground truth of these datasets. The number of labeled pixels available is detailed in Tab. III.
- *Multisource adaptation*: in the last experiment, we test the capability of the proposed SS-MA to align manifolds of different input dimensionality. To do so, we consider the 8-bands available in the Montelly and Prilly datasets, but use as a third image a 4-bands

PREPRINT

TABLE II

NUMBER OF LABELED PIXELS AVAILABLE FOR EACH DATASET IN THE MULTIANGULAR EXPERIMENTS (θ = OFF-NADIR ANGLE).

Class \ θ	$-38.79°$	$-29.16°$	$6.09°$	$26.76°$	$39.5°$
Water	83260	79937	66084	63492	54769
Grass	8127	8127	8127	8127	8127
Pools	244	244	223	195	195
Trees	4231	4074	3066	3046	3046
Concrete	707	719	719	719	696
Bare soil	790	790	790	790	811
Asphalt	2949	2949	2949	2827	2827
Grey buildings	6291	6061	5936	4375	4527
Red buildings	1147	1080	1070	1046	1042
White buildings	1683	1683	1571	1571	1571
Shadows	1829	1056	705	512	525
Tarmac	5179	5179	5179	2166	2758

QuickBird image of Zurich (Switzerland) acquired the 6th of October 2006. The image and the corresponding ground truth are illustrated in the last row of Fig. 3. The number of labeled pixels available is detailed in Tab. III. All the images show periurban areas and involve a common set of classes.

B. Experimental setup

As proposed in [33], all the similarity matrices $\mathbf{W}$ are scaled to have the same Frobenius norm in order to make them comparable in terms of energy. Consequently they all have the same contribution if $\mu = 1$. From all the available labeled pixels in each image, 50% are kept apart as the testing set. The remaining 50% samples are used to extract the labeled and unlabeled pixels used in the feature extractor.

The projection matrices $\mathbf{F}$ are learned in a traditional domain adaptation setting, where one domain provides many labeled examples l_1, and the other(s) have limited examples $l_{M\setminus 1}$. Since the projectors and models are trained with labeled samples from each domain, all the classes of all the images are represented simultaneously. This way, problems related to the appearance of

PREPRINT

TABLE III

NUMBER OF LABELED PIXELS AVAILABLE FOR EACH DATASET IN THE MULTITEMPORAL AND MULTISOURCE EXPERIMENTS

Class	Color in Fig. 3	Dataset			
		Prilly	Montelly	Malley	Zurich
Residential		151271	179695	24898	78018
Meadows		148604	47865	143674	12347
Trees		116343	177203	63956	52812
Roads		141353	104582	294687	43005
Shadows		39404	218189	194321	14071
Commercial		13692	22506	437633	25389
Rail		–	8570	76802	–
Bare		–	2812	33305	–
Highway		–	–	–	28827

TABLE IV

NUMBER OF LABELED AND UNLABELED PIXELS USED IN THE EXPERIMENTS.

Experiment	l_1	$l_{M\setminus 1}$	u_m
Toy 2D	20 per class	[5, ..., 20] per class	300 per source
Multiangular	100 per class	[10, 50] per class	500 per source
Multitemporal	100 per class	[10, ..., 90] per class	500 per source
Multisource	100 per class	[10, ..., 90] per class	500 per source

new classes are obviated. All the domains provide the same number of unlabeled samples u_m. The dataset sizes reported in Table IV have been considered.

In the multiangular experiment, only the near-nadir image ($6.07°$) is considered as the heavily labeled one (l_1). In the multitemporal and multisource experiments, all the domains are considered in turn as the heavily labeled.

The u_m unlabeled examples are selected using an iterative clustering algorithm, the bisecting k-means [45]. As in [46], we run binary partitions of the dataset until retrieving u_m clusters. The centroid of the cluster is then used as unlabeled example, thus ensuring an uniform sampling. The number of unlabeled samples used is reported for each experiment in Table IV. To build the geometrical Laplacian $\mathbf{L}_g$ in Eq. (2), we used a series of M graphs built using k-NN with $k = 9$ and all the n_m labeled and unlabeled samples. The parameter k defines the locality of the graph and, in our experimental tests, seemed not to influence the final results in a significant

PREPRINT

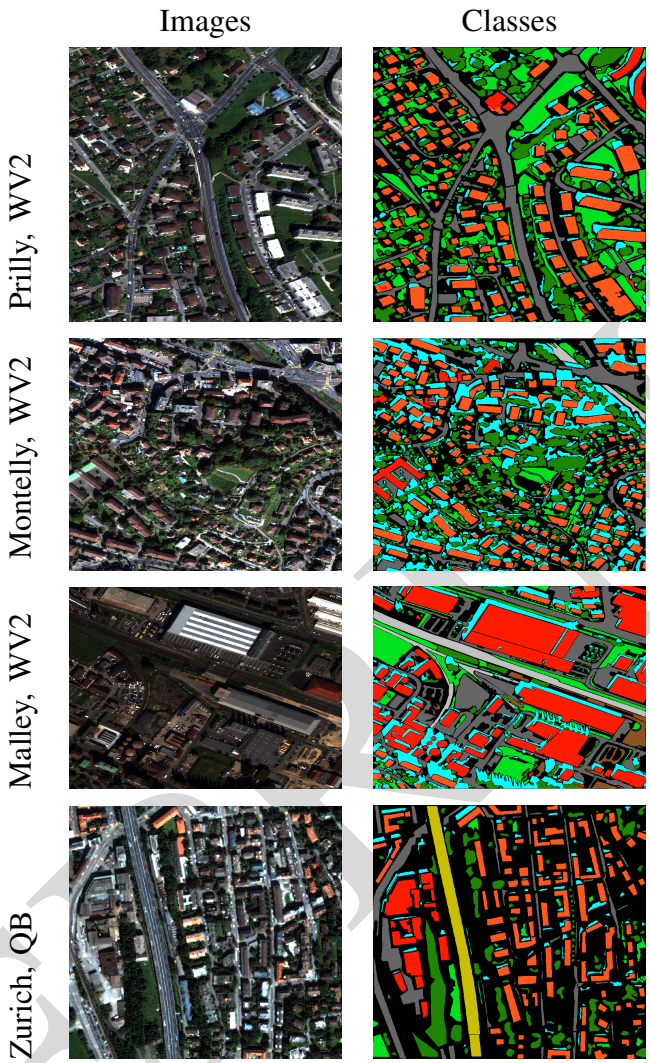


Fig. 3. The four images used in the multitemporal and multisource experiments.

way, as long as it is set in a reasonable range, i.e. a range where the manifold structure is not lost in an over-connected graph or oversimplified by an under-connected graph.

After extraction of the projection matrix $\mathbf{F}$, the best number of dimensions for classification is cross-validated exploiting the same labeled pixels already used to minimize Eq. (1). To assess the performance of the alignment, a linear SVM has been used to classify the original and projected data. Even though one could use more sophisticated and eventually nonlinear classifiers, our interest here is to assess the quality and discriminative power of the encountered projection features. The algorithm is actually nonlinear through the definition of the graph Laplacian so in

PREPRINT

principle there is no need to perform nonlinear classification. The libSVM library [47] has been used for such a classification and the regularization C parameter has been cross-validated using the entire labeled set in the range $[100, \ldots, 1000]$. In the multi-angular experiment, additional results obtained with other projectors (PCA, KPCA and graph matching [17]) and classifiers (LDA) are also reported. Finally, an in-depth analysis of the best dimensionality of the latent space is also conducted for LDA.

A final remark concerns the multisource experiment. Since the domains have different dimensionality, it is not possible to learn a classifier using simultaneously pixels from WorldView-2 and QuickBird. To do so, we downgraded the WorldView-2 images to the four QuickBird bands and provide the results obtained using only the four bands.

In all cases, the results reported are averages of the estimated Cohen's Kappa agreement score (κ hereafter [48]) over five realizations of the sequence *projection plus classification* in all domains with the projected data. This means that the projections are derived once, a single classifier is trained on the projected training samples, and then applied to each of the projected images, considered in turn as the test image. We set the random generator in a way that the samples of each run are automatically included in the corresponding run using more labeled pixels, i.e. the pixels included in the first run using 10 pixels *per class* are also used in the first run using $20, 30, \ldots$ pixels per class. Note that no specific training process for each test image is performed, as the aim is to classify in the latent space, regardless of the image which will be projected into it.

IV. Results and discussion

This section presents and discusses the results obtained in the four case studies described in Section III.

A. 2D toy datasets

The two first columns of Fig. 4 represent the three deformations considered: 1) scaling (S), 2) scaling plus rotation (S + R) and 3) scaling, rotation and translation (S + R + T). The third column illustrates the result of SS–MA in one of the five experiments performed in each case. In all cases, the proposed method compensates for the present deformations and aligns the two

PREPRINT

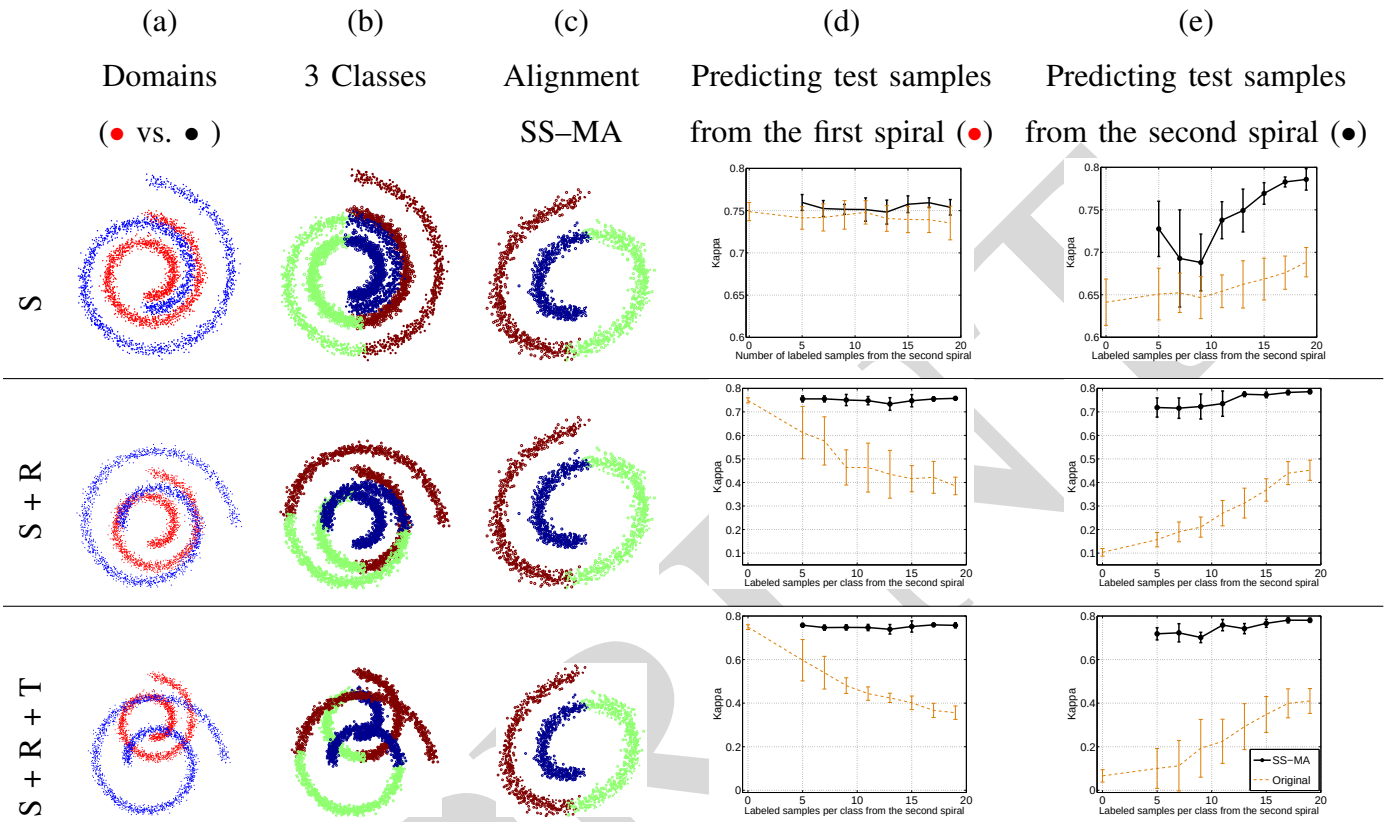


Fig. 4. Toy examples considered to test the invariance to scaling (S) rotation (R) and translation (T) of the SS-MA solution. (a): the two sources (● and ●) under different deformations (S = scaling; R = rotation; T = translation). (b) The distribution of the three classes. (c) Result of alignment with SS–MA. (d) Classification results with a linear SVM trained using 20 samples from the first source ($l_{\bullet}$= 20 *per* class) plus an increasing number of samples from the second ($l_{\bullet}$ = [0, 5, ..., 20] *per* class) when predicting the test samples from the first spiral (Test = ●). (e) Classification results in the same conditions, but predicting the second spiral (Test = ●).

spirals in the correct way. Subsequent classifications by a linear SVM (reported in the fourth and fifth columns of Fig. 4 show two trends:

- When predicting data from the first spiral (●, 20 labeled pixels *per* class), adding samples from the other one does not influence the classification results. On the contrary, when using unprojected data, a strong decrease in performance is observed in the three cases. SS–MA shows a stable behavior, since the distributions have been aligned correctly.
- When predicting data from the second spiral (●, [5, ..., 20] labeled pixels *per* class), SS–MA can exploit the common information of the two datasets efficiently and it always outperforms the unprojected data.

PREPRINT

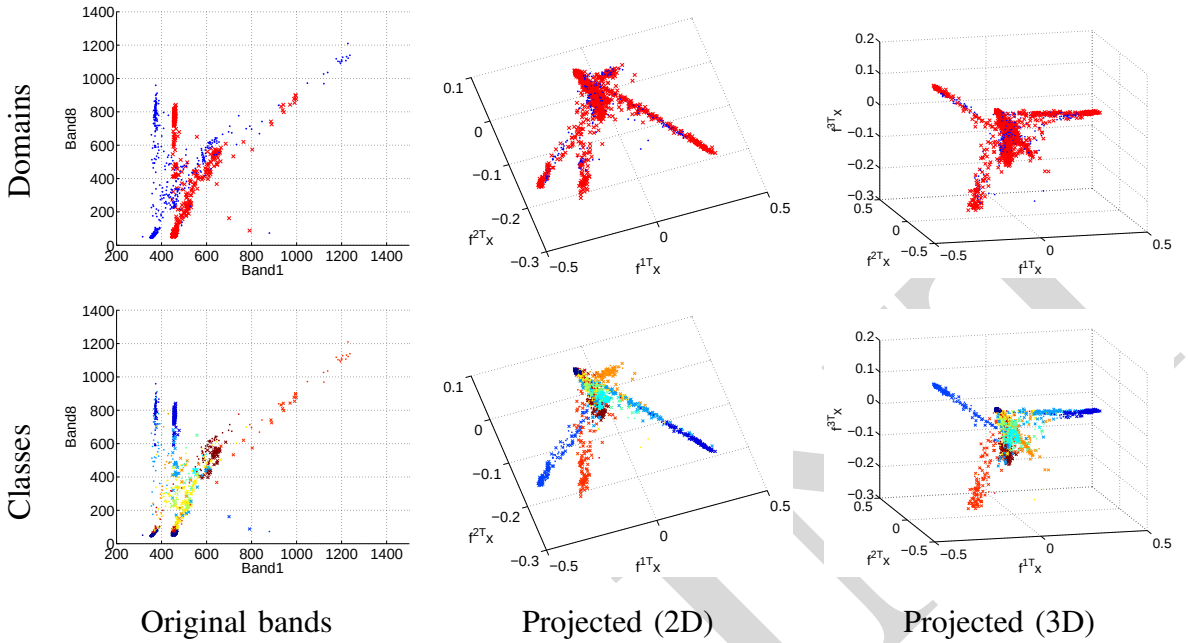


Fig. 5. SS–MA projections. Left: original acquisitions at $6.09°$ and $-38.79°$. Center: two first dimensions of the latent space. Right: three first dimensions of the latent space. In the top row, each color corresponds to a domain (an image), in the bottom row, each color corresponds to a class.

B. Multiangular adaptation

The first real experiment considers the multiangular sequence of Rio de Janeiro. As this dataset comprises a set of images from the same sensor and acquired at a very short time interval, the shifts observed are only related to angular effects of increasing strength. Besides analyzing the classification accuracy with respect to a set of competing methods, we also use this dataset to study the evolution of the performance with respect to the dimensionality of the latent space.

First, we visualize the latent space, to assess its discriminative power. Figure 5 illustrates the projections of the images at $6.09°$ and $-38.79°$ in the 2-dimensional (2 first eigenvectors) and 3-dimensional (3 first eigenvectors) latent space. From the plots, we can observe that the model defined a discriminative space, where joint linear classification is facilitated.

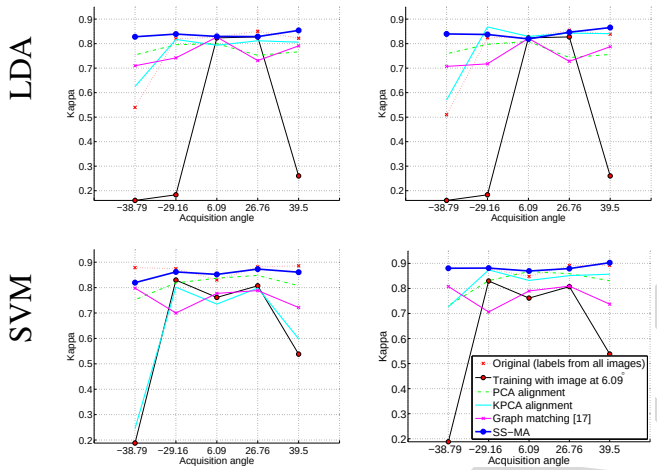
Figure 6 illustrates the results obtained by 1) the baseline (using only the labeled pixels from the $6.09°$ image as training set), 2) the joint use of all the acquisitions, but without projection, 3) the joint use of all the acquisitions after projection with the proposed SS–MA and 4) three aligning methods such as PCA, KPCA and graph matching [17]. To be fair in the evaluation,

the projections for these methods are obtained in an unsupervised way, but then the classifier is trained using the original training points from the nadir acquisition, stacked to the transformed labeled pixels of the domain to be tested.

Using only the nadir acquisition leads to poor results for increasing off-nadir angles, at the point that the κ statistic observed are close to 0.2 (black solid line with red markers). Adding labeled information from the specific viewing angle helps and improves notably the results (red dashed line). However, and especially when using a simple parametric model as LDA, the decrease in performance is still observed, especially for strong off-nadir acquisition. The proposed SS–MA (blue solid line) results in an almost flat κ surface along the acquisitions, thus showing that projection in the latent space permits 1) to align efficiently the spectra of the acquisitions, and 2) to reduce all the image-specific problems to the same (latent) classification problem. The results observed for SS–MA are always accurate and stable along the acquisitions: they join the stability of the unsupervised method designed specifically for domain adaptation (graph matching, pink line) and the higher numerical results of a generic supervised one (the one depicted by the red curve, where samples from all acquisitions are used to train the model). SS–MA always outperforms PCA alignment (green line) and is always at least as accurate as KPCA (which is also nonlinear – cyan line). Compared to KPCA, SS–MA shows more stable results along the acquisitions, regardless of the classifier and of the number of labeled pixels used.

The use of the LDA classifier also allows to assess the discriminative power of the latent space across its dimensions. Unlike SVM, which is robust to high dimensionality with noisy dimensions, LDA is strongly affected by non-discriminative dimensions and by non-Gaussian class-conditional distributions. Therefore, we studied the evolution of accuracy, when increasing the dimension of the latent space in which the classifier is trained. Fig. 7(a) shows the evolution of the κ statistic for one run of the algorithm in the nadir case ($\theta = 6.09°$). Fig. 7(b) shows the same evolution for the off-nadir angles. From these plots, we can conclude that when the shift between the domain providing most of the labeled samples and the destination domain is low (as for the $6.09°$ and the $26.76°$ images), a lower number of dimensions is sufficient to achieve the best performances. On the contrary, when the shift is larger, the latent space provides informative dimensions up to the last components. Another measure of the complexity of the alignment problem can be provided by studying the dimension of the latent space corresponding

PREPRINT



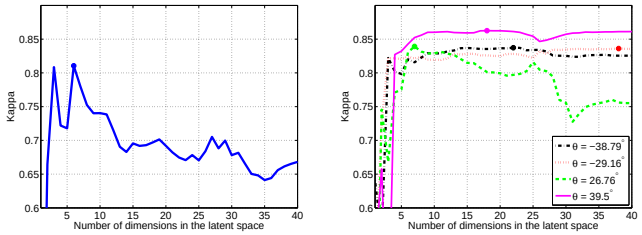
(a) (b)

10 pixels *per* class 50 pixels *per* class

Fig. 6. Numerical performances of the multiangular experiments. The nadir image provides 100 labeled pixels *per* class, while the other are aligned with an increasing number of labeled pixels.

to the highest classification performance in each experiment (and not on a global average, as the dimension represented by the filled dots in Fig. 7): Table V illustrates the average dimension for the five angles and varying number of training examples. Two tendencies are observed: the most nadir acquisition ($\theta = 6.09°$) requires the smallest number of dimensions of the latent space to achieve accurate classification. This is not surprising, as this domain is the one providing most of the labeled examples. For the other domains, a larger number of dimensions is necessary, ranging from 12 to 18 for the acquisition at $26.75°$ and higher dimensionality for the other acquisitions, related to larger shift (cf. the lower performance of the LDA baseline in Fig. 6). The only exception observed is the result for 10 pixels *per* class at $-39.79°$, which shows an unusually small average best dimensionality. This is due to the instability of the best dimensionality along the experiments, most likely due to the small number of pixels per class used to align the $-39.79°$ domain: for two experiments it required three dimensions and for the rest about 30. However, note that this does not impact the classification performance reported in Fig. 6, which is stable around 0.82 in κ statistic.

PREPRINT



(a) (b)

Fig. 7. Evolution of the average κ along the dimensions of the latent space defined by the five domains ($d = 5 \times 8 = 40$) for the experiment using 50 pixels *per* class. (a) Predicting the most nadiral domain ($\theta = 6.09°$); (b) predicting the four other domains. The filled dot corresponds to the dimension returning the higher average κ

TABLE V

AVERAGE NUMBER OF DIMENSIONS REQUIRED BY SS–MA TO OBTAIN THE BEST CLASSIFICATION PERFORMANCE (BLUE CURVE IN FIG. 6)

	Optimal dimensionality for θ				
	$-38.79°$	$-29.16°$	$6.09°$	$26.76°$	$39.5°$
10 *per* class	10.6	34.4	3.0	18.0	32.2
50 *per* class	25.8	24.2	6.2	11.6	23.8
90 *per* class	25.0	31.6	3.0	11.2	21.4

C. Multitemporal adaptation for the same sensor

Figures 8 and 9 illustrate the numerical results for the multitemporal experiments involving three World-View 2 scenes. The first figure shows the κ surfaces for increasing number of labeled samples in the auxiliary domains $\mathbf{X}^2$ and $\mathbf{X}^3$ for both the original data and the data aligned with the SS–MA method. The leading domain $\mathbf{X}^1$ is always sampled with $l_1 = 100$ labeled samples *per* class. Figure 9 summarizes the former by illustrating the diagonal of the surfaces, and compares it to the baseline case (in green), where only the tested domain is used for training the model. This baseline is what would happen in a traditional supervised classification study.

In both cases, each row corresponds to the leading training domain $\mathbf{X}^1$ and each column correspond to a different testing image (the baseline is thus the same column-wise). We remind that in the 'SS–MA' case, one set of projections *per* combination of labeled samples in the three domains is extracted and all the images are then classified with the same SVM model, trained on the labeled pixels of the three domains. The three blocks of each row in the figures (one *per*

PREPRINT

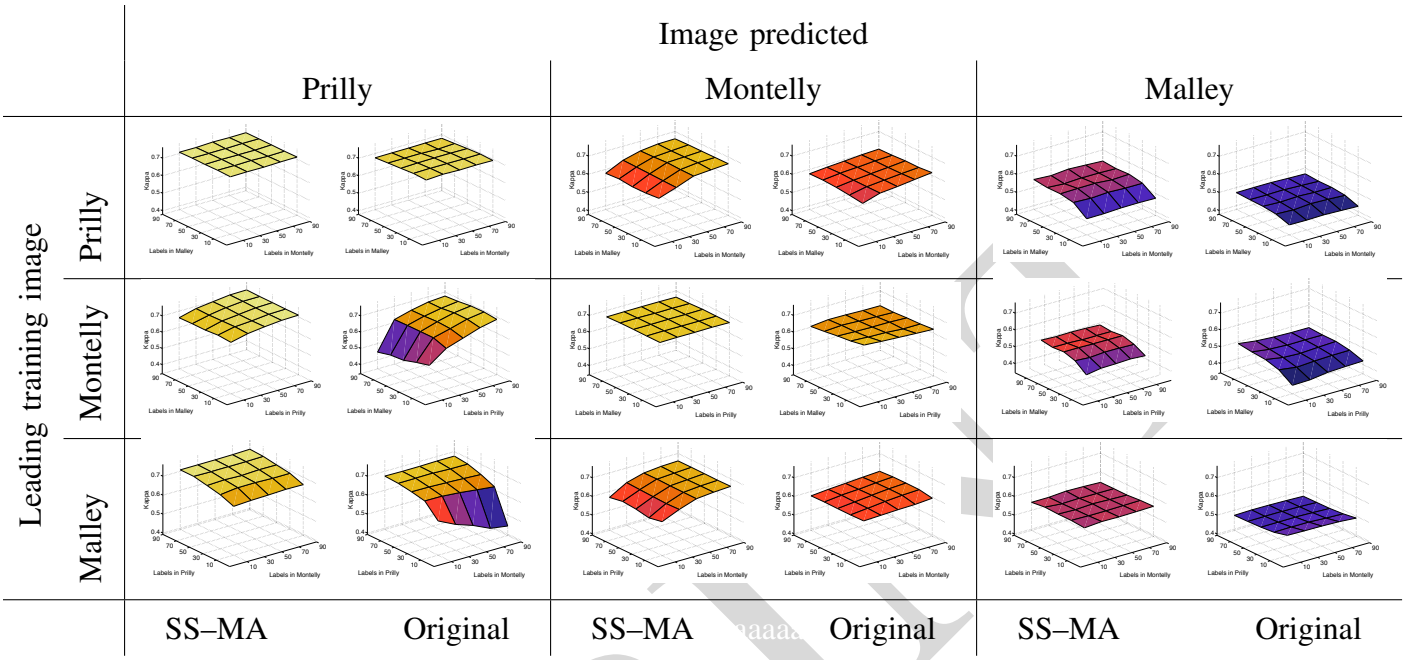


Fig. 8. Numerical results for the multitemporal experiment. Rows indicate the image from which 100 labeled pixels *per* class are used ($l_1 = 100$ *per* class). κ performance for increasing number of labeled pixels in the two other images ($l_{2,3} = [10, ..., 90]$ *per* class) are reported. Brighter tones correspond to higher κ. Columns correspond to the image that has been used for testing.

testing image) are issued from the same model. The same holds for the 'original' experiments: a common model is trained on the labeled pixels in the original space and then each image is predicted separately.

For all tested scenarios, the proposed SS–MA leads to higher κ scores than the method without alignment. In general, the inclusion of labeled pixels from the testing domain (e.g., from the Prilly scene, when testing on the Prilly scene) is beneficial, so that the ideal case is met when having labeled samples from all domains. Moreover, some interesting observations can be done:

- When predicting the same image as the leading training one (diagonal blocks of Figs. 8 and 9), adding pixels from the other images results in a decrease of the performance if no alignment is performed. This is particularly visible in the Montelly and Malley cases, where the performance of the 'original' model (the red bars in the diagonal panels of Fig. 9) decrease with the increase of labeled samples from the other domains. On the contrary, when aligning the datasets with the proposed SS–MA, the κ performance remains stable at a level generally higher than the baseline (blue bars in the same panels of Fig. 9). Since it aligns domains in a discriminative sense, SS–MA allows to use many domains without affecting

PREPRINT

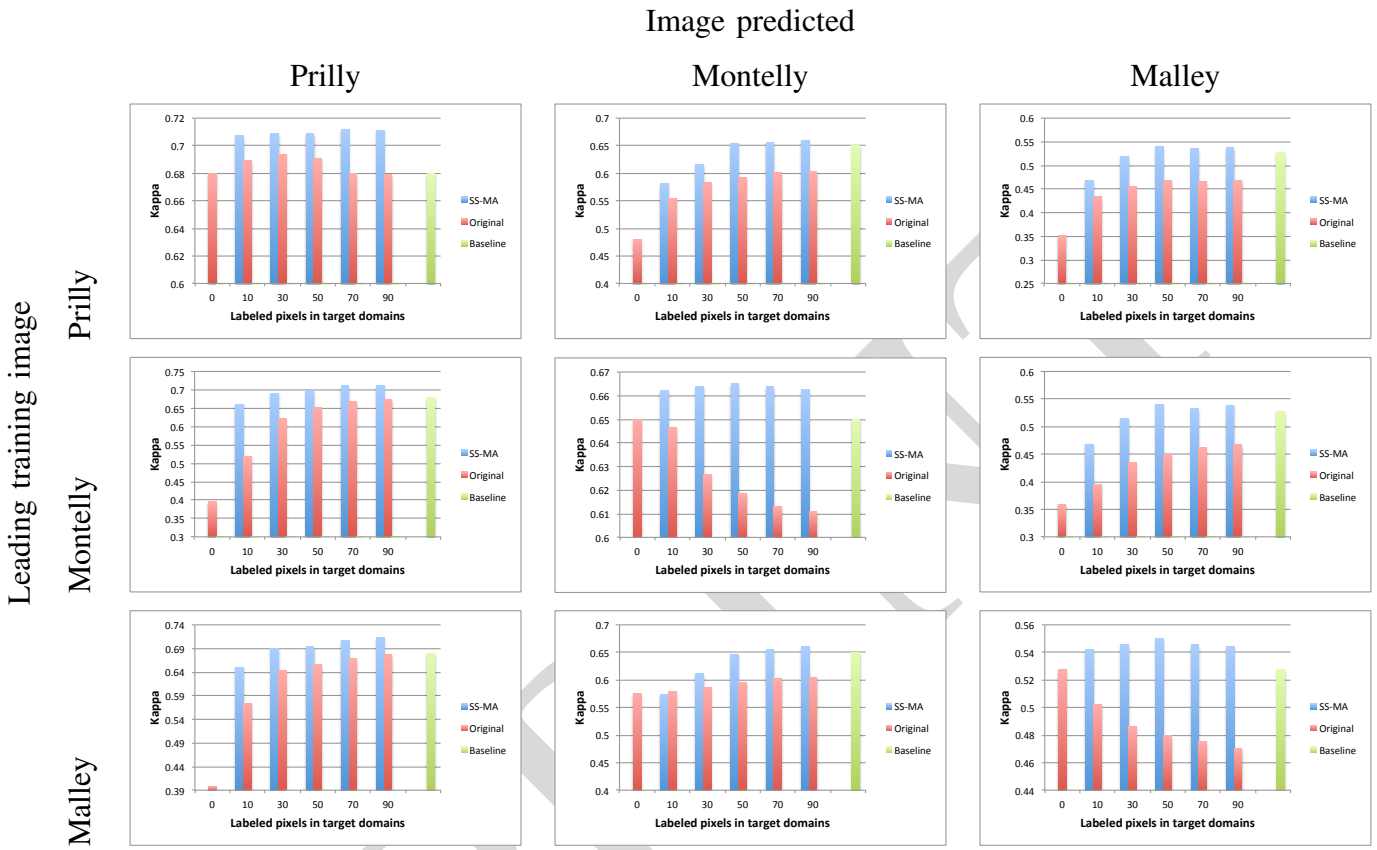


Fig. 9. Numerical results for the multitemporal experiment. Rows indicate the image from which 100 labeled pixels *per* class are used ($l_1 = 100$ *per* class). κ performances for increasing number of labeled pixels in the testing images ($l_2 = l_3 = [10, ..., 90]$ *per* class) are reported. The results correspond to the diagonal of the surfaces in Fig. 8. Columns correspond to the image that has been used for testing. The baseline is the model obtained using 100 pixels *per* class from that image only.

the quality of the classifier in the original domain.

- When predicting other images than the one with a large amount of labeled samples (the off-diagonal blocks of Figs. 8 and 9) both the original and SS–MA classification lead to an improvement of the results of the case where only pixels from the leading domain are used (visible as the first bar of the histograms in Fig. 9. The alignment procedure boosts the performance of the classifier and leads to κ scores that are between 0.1 and 0.3 κ points higher with respect to the case, where only pixels from the leading domain are used for training.
- In all cases, SS–MA leads to more drastic improvements than the non-projected method and can meet the baseline performance in all the cases. This shows that, when using the

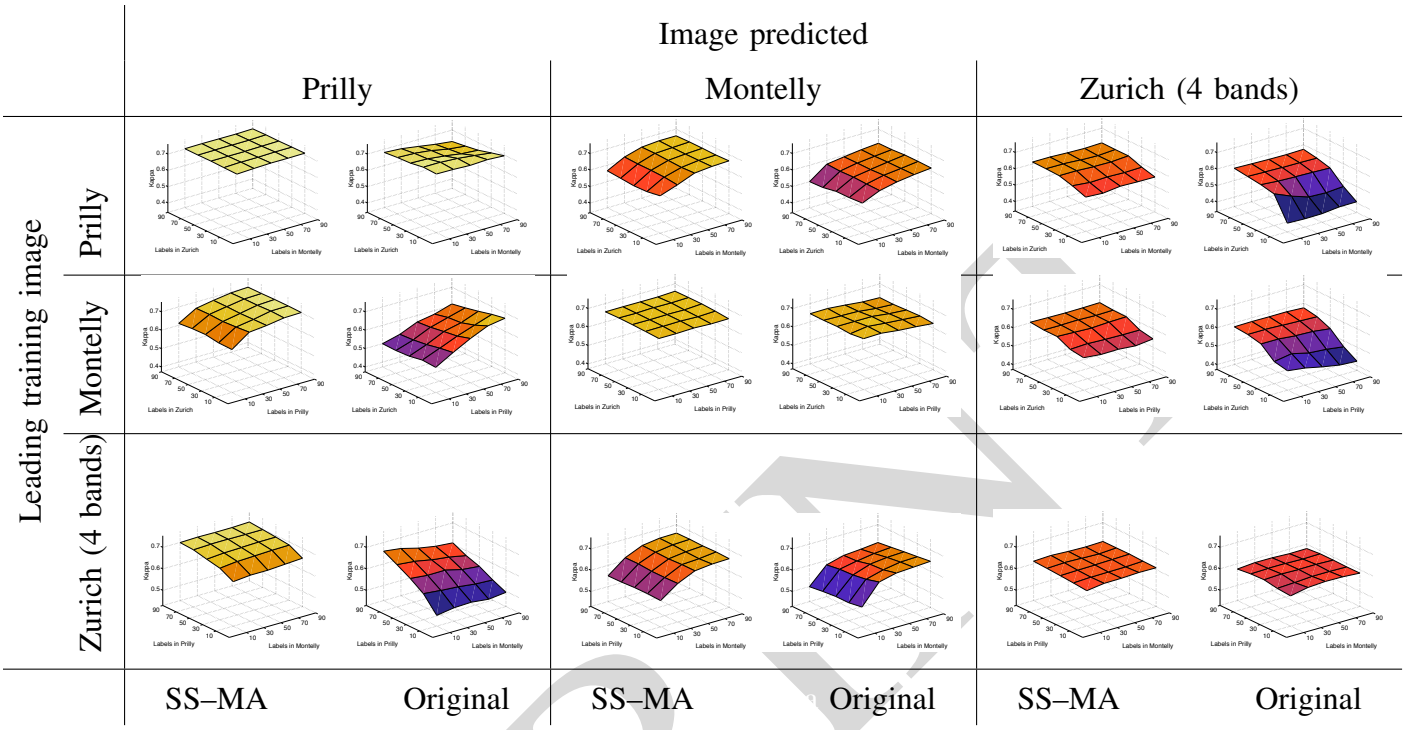


Fig. 10. Numerical results for the multisource experiment. Rows indicate the image from which 100 labeled pixels *per* class are used ($l_1 = 100$ *per* class). κ performances for increasing number of labeled pixels in the two other images ($l_{2,3} = [10, ..., 90]$ *per* class) are reported. Brighter tones correspond to higher κ. Columns correspond to the image that has been used for testing.

SS–MA projections, adding information from other domains improves the classification performances, provided the higher number of labeled samples and the discriminative nature of the projection functions.

D. Multisource adaptation

The last experiment deals with multisource data. Figures 10 and 11 illustrate the performance of the SS–MA projection against cases where no projection is performed and a baseline, where only pixels of the testing domain are used (green bar in Fig. 11). In all the results reported, the richness of the joint latent space provides better classification results, regardless of the dimensionality of the input space.

The trends observed in this experiment are similar to those observed in the multitemporal case, with the difference that the performance of the classifiers trained without adaptation is decreased: to allow multisource classification, the four bands missing in QuickBird were removed, and the performance of the SVMs exploiting a four dimensional input space are thus lower than in

PREPRINT

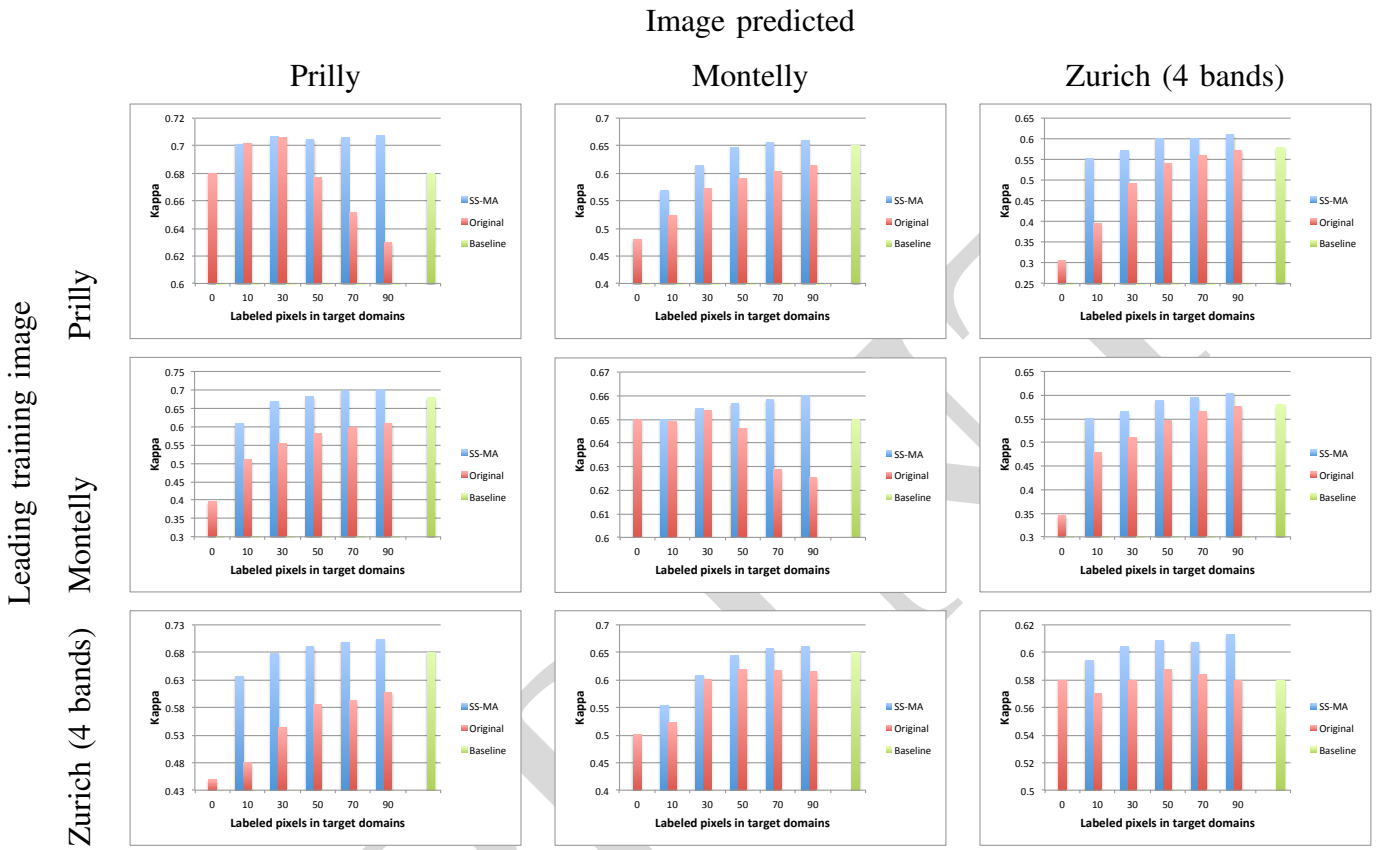


Fig. 11. Numerical results for the multisource experiment. Rows indicate the image from which 100 labeled pixels *per* class are used ($l_1 = 100$ *per* class). κ performances for increasing number of labeled pixels in the two other images ($l_2 = l_3 = [10, ..., 90]$ *per* class) are reported. The results correspond to the diagonal of the surfaces in Fig. 10. Columns correspond to the image that has been used for testing. The baseline is the model obtained using 100 pixels *per* class from that image only.

the previous case. On the contrary, the aligned data exploit the joint latent space of dimension $d = 8 + 8 + 4 = 20$ and return very high improvements in performances with respect to the unprojected counterpart.

V. Conclusions

In this paper, we presented a semisupervised method to align manifolds of remote sensing images. Alignment is performed in order to enhance model portability in multitemporal, multisource and multiangular classification. The method aligns the datasets by pulling close together pixels of the same class, while pushing those of different classes apart. At the same time, it preserves the geometry of each manifold along the transformation. The method performs well

even in case of strong deformations and provides models that can classify target domains at least as accurately as if a large set of labeled pixels for that domain were available.

We reported experiments in three problems involving VHR images acquired by different sensors and with different angular properties, that cannot be tackled in a satisfactory way with current relative normalization strategies, such as univariate histogram matching or PCA matching. None of the image datasets considered in this work were co-registered, thus showing the main interest of the proposed SS–MA with respect to methods such as canonical correlation analysis. Of course, the presence of labeled pixels in each image is a limitation, that we will address in future research.

ACKNOWLEDGEMENTS

The authors would like to thank the IADF Technical Committee of the IEEE and Digital Globe for the access to the multiangular data over Rio. They also would like to thank Adrian Moriette (trainee at LaSIG, EPFL) who, diligently and relentlessly, prepared the ground truths of the World-View 2 images considered in this work.

REFERENCES

[1] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jimenez, and J. Malo, *Remote Sensing Image Processing*, Synthesis Lectures on Image, Video, and Multimedia Processing. Morgan and Claypool, 2011.

[2] S. Prasad, L. Bruce, and J. Chanussot, *Optical Remote Sensing*, Advances in Signal Processing and Exploitation Techniques. Springer, 2011.

[3] 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.

[4] F. Petitjean, J. Inglada, and P. Gancarski, "Satellite image time series analysis under time warping," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3081–3095, 2012.

[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.*, vol. 50, no. 4, pp. 1155–1170, 2012.

[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] 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.

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

[9] 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.

PREPRINT

[10] 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.

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

[12] G. Matasci, D. Tuia, and M. Kanevski, "SVM-based boosting of active learning strategies for efficient domain adaptation," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 5, pp. 1335–1343, 2012.

[13] L. Guanter, A. Ruiz-Verdú, D. Odermatt, C. Giardino, S. Simis, V. Estelles, T. Heege, J. A. Domínguez-Gómez, and J. Moreno, "Atmospheric correction of ENVISAT/MERIS data over inland waters: Validation for European lakes," *Remote Sens. Enviro.*, vol. 114, pp. 467–480, 2010.

[14] E. Izquierdo-Verdiguier, V. Laparra, L. Gómez-Chova, and G. Camps-Valls, "Encoding invariances in remote sensing image classification with SVM," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 981–985, 2013.

[15] L. Guanter, R. Richter, and H. Kaufmann, "On the application of the MODTRAN4 atmospheric radiative transfer code to optical remote sensing," *Int. J. Remote Sens.*, vol. 30, pp. 1407–1424, 2009.

[16] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Trans. Im. Proc.*, vol. 11, no. 3, pp. 293–305, 2002.

[17] D. Tuia, J. Muñoz-Marí, L. Gómez-Chova, and J. Malo, "Graph matching for adaptation in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 329–341, 2013.

[18] G. Matasci, N. Longbotham, F. Pacifici, Kanevski M., and D. Tuia, "Statistical assessment of dataset shift in model portability of multi-angle image acquisitions," in *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, Melbourne, Australia, 2013.

[19] C. Wang, P. Krafft, and S. Mahadevan, "Manifold alignment," in *Manifold Learning: Theory and Applications*, Y. Ma and Y. Fu, Eds. CRC Press, 2011.

[20] G. Hong and Y. Zhang, "Radiometric normalization of IKONOS image using Quickbird image for urban area change detection," in *Proceedings of ISPRS 3rd Int. Symposium on Remote Sensing and Data Fusion Over Urban Areas*, Tempe, AZ, 2005.

[21] Z. Yang and R. Mueller, "Heterogeneously sensed imagery radiometric response normalization for citrus grove change detection," in *Proceedings of SPIE Optics East*, Boston, MA, 2007, vol. 6761.

[22] G. Matasci, L. Bruzzone, M. Volpi, D. Tuia, and Kanevski M., "Investigating the feature extraction framework for domain adaptation in remote sensing image classification," in *International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, Barcelona, Spain, 2013.

[23] J. Arenas-García, K. B. Petersen, G. Camps-Valls, and L. Kai Hansen, "Kernel multivariate analysis framework for supervised subspace learning," *Signal Proc. Mag.*, vol. 16, pp. 16–29, 2013.

[24] 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.

[25] 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.

[26] J. J. McAuley and T. S. Caetano, "Fast matching of large point sets under occlusion," *Pattern Recogn.*, vol. 45, pp. 563–569, 2012.

[27] N Kambathla and T. K. Leen, "Dimension reduction by local principal component analysis," *Neural Comp.*, vol. 9, no. 7, pp. 1493–1516, 1997.

PREPRINT

[28] S. T. Roweis, L. K. Saul, and G. E. Hinton, "Global coordination of local linear models," in *Advances in neural information processing systems (NIPS)*. 2002, vol. 14, pp. 889–896, MIT Press.

[29] J. J. Verbeek, N. Vlassis, and B. Krose, "Coordinating principal component analyzers," in *Proc. International Conference on Artificial Neural Networks*, 2002, pp. 914–919.

[30] Y. W. Teh and S. Roweis, "Automatic alignment of local representations," in *Advances in neural information processing systems NIPS*. 2003, vol. 15, pp. 841–848, MIT Press.

[31] M. Brand, "Charting a manifold," in *Advances in neural information processing systems NIPS*. 2003, vol. 15, pp. 961–968, MIT Press.

[32] J. Ham, D. D. Lee, and L. K. Saul, "Semisupervised alignment of manifolds," in *Proc. Int. Workshop Artificial Intelligence and Statistics*, 2005.

[33] C. Wang and S. Mahadevan, "Heterogeneous domain adaptation using manifold alignment," in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2011.

[34] H.L. Yang and M.M. Crawford, "Manifold alignment for multitemporal hyperspectral image classification," in *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, 2011, pp. 4332 –4335.

[35] R. Zurita-Milla, J. Clevers, and M. Schaepman, "Unmixing-based Landsat TM and MERIS FR data fusion," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 453–457, 2008.

[36] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-supervised learning*, MIT press, 2006.

[37] Y. Chen, M.M. Crawford, and J. Ghosh, "Improved nonlinear manifold learning for land cover classification via intelligent landmark selection," in *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, Denver, CO, 2006, pp. 545–548.

[38] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Muñoz-Mari, "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.

[39] J. Chi and M.M. Crawford, "Selection of landmark points on nonlinear manifolds for spectral unmixing using local homogeneity," *Geoscience and Remote Sensing Letters, IEEE*, vol. 10, no. 4, pp. 711–715, 2013.

[40] D. A. Spielman and S. Teng, "Spectral sparsification of graphs," *CoRR*, vol. abs/0808.4134, 2008.

[41] O. E. Livne and A. Brandt, "Lean algebraic multigrid (LAMG): Fast graph Laplacian linear solver.," *SIAM J. Scientific Computing*, vol. 34, no. 4, 2012.

[42] I. S. Dhillon, *A new $\mathcal{O}(n^2)$ algorithm for the symmetric tridiagonal eigenvalue/eigenvector problem*, Ph.D. thesis, Berkeley, CA, USA, 1998, UMI Order No. GAX98-03176.

[43] V. Y. Pan and Z. Q. Chen, "The complexity of the matrix eigenproblem," in *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, New York, NY, USA, 1999, STOC '99, pp. 507–516, ACM.

[44] F. Pacifici, J. Chanussot, and Q. Du, "2011 GRSS Data Fusion Contest: Exploiting WorldView-2 multi-angular acquisitions," in *IEEE International Geoscience and Remote Sensing Symposium, IGARSS*, 2011.

[45] R. Kashef and M.S. Kamel, "Enhanced bisecting k-means clustering using intermediate cooperation," *Pattern Recogn.*, vol. 42, no. 11, pp. 2257–2569, 2009.

[46] J. Muñoz-Marí, D. Tuia, and G. Camps-Valls, "Semisupervised classification of remote sensing images with active queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3751–3763, 2012.

[47] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

PREPRINT

[48] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Rem. S.*, vol. 50, no. 5, pp. 627–633, 2004.

Devis Tuia (S'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 València, Spain and the University 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. He serves as Associate Editor for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING and as a co-chair of the *Image Analysis and Data Fusion* technical committee of the GRSS. Visit http://devis.tuia.googlepages.com/ for more information.

Michele Volpi (S'08-M'13) was born in Lugano, Switzerland, in 1985. He received the M.Sc. and the Ph.D. degree in environmental sciences from the University of Lausanne, Lausanne, Switzerland, in 2009 and in 2013, respectively. He is currently a visiting postdoc at the CALVIN lab, Institute of Perception, Action and Behaviour at the School of Informatics of the University of Edinburgh under the Swiss National Science Foundation grant P2LAP2-148432. His research activities are in the area of computer vision and machine learning for remote sensing image processing and analysis. In particular, his interests include the development and application of semantic segmentation techniques for uni- and multi-temporal images, change detection and feature extraction from multi- and hyperspectral data using machine learning models and kernel methods. Visit https://sites.google.com/site/michelevolpiresearch/ for more information.

Maxime Trolliet was born in Lausanne, Switzerland, in 1986. He completed the M.S. degree at the Federal Institute of Technology of Lausanne (EPFL) in Environmental Engineering in 2012. During so he accomplished his Master Thesis at the LaSIG laboratory, EPFL, in the field of remote sensing and data fusion. He is currently working as environmental project manager for the city of Nyon, Switzerland.

PREPRINT

Gustau Camps-Valls (M'04, SM'07) received a Ph.D. degree in Physics (2002, *summa cum laude*) from the Universitat de València, Spain, where he is currently an Associate Professor in the Electrical Engineering Dep. His research is conducted as Group Leader of the Image and Signal Processing (ISP) group, http://isp.uv.es, of the same university. He has been Visiting Researcher at the Univ. Trento, (Italy) and the Max Planck Institute for Biological Cybernetics (Tübingen, Germany), and Invited Professor at the EPFL (Lausanne, Switzerland). His research interests are tied to the development of machine learning algorithms for signal and image processing with special focus on remote sensing data analysis.

He conducts and supervises research within the frameworks of several national and international projects, and he is Evaluator of project proposals and scientific organizations. He is the author (or co-author) of 115 international peer-reviewed journal papers, more than 150 international conference papers, 20 international book chapters, and editor of the books "Kernel methods in bioengineering, signal and image processing" (IGI, 2007), "Kernel methods for remote sensing data analysis" (Wiley & Sons, 2009), and "Remote Sensing Image Processing" (MC, 2011). He's a co-editor of the forthcoming book "Digital Signal Processing with Kernel Methods" (Wiley & sons, 2014). He holds a Hirsch's h index $h = 34$ and entered the ISI list of Highly Cited Researchers in 2011.

He is a referee and Program Committee member of many international journals and conferences. Since 2007 he is member of the Data Fusion technical committee of the IEEE GRSS, and since 2009 of the Machine Learning for Signal Processing Technical Committee of the IEEE SPS. He is member of the MTG-IRS Science Team (MIST) of EUMETSAT. He is Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE SIGNAL PROCESSING LETTERS, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. Visit http://www.uv.es/gcamps for more information.

PREPRINT