Multi-temporal and Multi-source Remote Sensing Image Classification by Nonlinear Relative Normalization

Devis Tuia[a], Diego Marcos[a], Gustau Camps-Valls[b]

[a] *MultiModal Remote Sensing, University of Zurich, Switzerland.*
http://www.geo.uzh.ch/en/units/multimodal-remote-sensing
[b] *Image Processing Laboratory, Universitat de València, Spain.*
http://isp.uv.es

Abstract

Remote sensing image classification exploiting multiple sensors is a very challenging problem: data from different modalities are affected by spectral distortions and mis-alignments of all kinds, and this hampers re-using models built for one image to be used successfully in other scenes. In order to adapt and transfer models across image acquisitions, one must be able to cope with datasets that are not co-registered, acquired under different illumination and atmospheric conditions, by different sensors, and with scarce ground references. Traditionally, methods based on histogram matching have been used. However, they fail when densities have very different shapes or when there is no corresponding band to be matched between the images. An alternative builds upon *manifold alignment*. Manifold alignment performs a multidimensional relative normalization of the data prior to product generation that can cope with data of different dimensionality (e.g. different number of bands) and possibly unpaired examples. Aligning data distributions is an appealing strategy, since it allows to provide data spaces that are more similar to each other, regardless of the subsequent use of the transformed data. In this paper, we study a methodology that aligns data from different domains in a nonlinear way through *kernelization*. We introduce the Kernel Manifold

*Corresponding Author: Devis Tuia, devis.tuia@geo.uzh.ch, Tel: +4144 635 52 60, Fax: +4144 635 68 46.
This is the pre-acceptance version, to read the final, published version, please go to the DOI: 10.1016/j.isprsjprs.2016.07.004

Alignment (KEMA) method, which provides a flexible and discriminative projection map, exploits only a few labeled samples (or semantic ties) in each domain, and reduces to solving a generalized eigenvalue problem. We successfully test KEMA in multi-temporal and multi-source very high resolution classification tasks, as well as on the task of making a model invariant to shadowing for hyperspectral imaging.

Keywords: Feature extraction, Manifold learning, Domain adaptation, Graph-based methods, Hyperspectral imaging, Very high resolution, Classification, Kernel methods.

---

1. Introduction

Many real-life problems currently exploit heterogeneous sources of remote sensing data: forest ecosystems studies (Asner et al., 2005, 2006; Roth et al., 2015), post-catastrophe assessment (Brunner et al., 2010; Taubenböck et al., 2011) or land-use updating (Bruzzone and Fernandez-Prieto, 2001; Nielsen, 2002; Amorós-López et al., in press) take advantage of the wide coverage and short revisit time of remote sensing sensors. They typically design specific image processing pipelines to produce maps of a product of interest. Despite the promises of remote sensing to tackle such ambitious problems, two main obstacles prevent this technology from reaching a broader range of applications: on the one hand, there is generally a lack of labeled data present at each acquisition and, on the other hand, the models need to be capable of dealing with images obtained under different conditions and thus potentially with different sensors.

Working under label scarcity has been extensively considered in recent remote sensing image processing literature by means of optimizing the use of the few available labels (Camps-Valls et al., 2014). In our view, the problem of adapting remote sensing classifiers boils down to compensating for a variety of distortions and mis-alignments: for example, data resolution may differ or seasonal conditions might offer remarkable differences in the spectral signatures observed. When the images cover the same area, registration can be approximate. Moreover, each scene depends on its particular illumination and viewing geometry, which causes spectral signatures to shift among acquisitions (Matasci et al., 2015). As a consequence, it becomes difficult, often impossible, to re-use field data acquired on a given campaign to process newly acquired images. Transferring models from one remote sensing image

acquisition to the other can be a very challenging task.

Adapting classifiers to (even slightly) shifted data distributions is an old problem in remote sensing, which started in the 1970s with the signature extension field (Fleming et al., 1975; Olthof et al., 2005), and then evolved, due to the technological advances in both sensor and processing routines, into what is generally referred to as the *transfer learning* problem (Pan and Qiang, 2010; Patel et al., 2015). By transfer learning, we mean all kind of methodologies aiming at making models *transferable* across image/data acquisitions. In recent remote sensing literature, works have mainly considered three research directions (Tuia et al., in press): 1) unifying the data representation, for example via atmospheric correction (Guanter et al., 2009), feature selection (Bruzzone and Persello, 2009), or feature extraction (Volpi et al., 2015; Sun et al., 2016, in press); 2) incorporating invariances in the classifier, for example via synthetic ('virtual') examples (Izquierdo-Verdiguier et al., 2013) or physically-inspired features (Pacifici et al., 2014; Verrelst et al., 2010); and 3) adapting the classifier to cope with the shift among acquisitions, for example via semi-supervised-inspired strategies (Rajan et al., 2006; Bruzzone and Marconcini, 2010) or active learning (Matasci et al., 2012).

Most of the methodologies above rely on the fact that all images are acquired by the same sensor (i.e. they share the same d-dimensional data space, as well as the nature -and physical meaning- of the features), or that all information and know-how necessary to convert to surface reflectance is available to the user performing the analysis, which is unfortunately often not the case. Moreover, at the application level there is generally no requirement of sticking to a specific sensor (taking the example of post-catastrophe intervention, the fact of waiting for the next cloud-free image of a specific sensor can mean the loss of human lives): since more and more images are currently available to the general public and organizations, new transfer learning approaches must be capable to unify data from different sensors, at different resolutions, without co-registration, and without being specific to a given end classifier (Gómez-Chova et al., 2015). The recently proposed *manifold alignment* methods gather all these properties.

Manifold alignment (Wang et al., 2011) is a machine learning framework aiming at matching, or *aligning*, a set of domains (the images) of potentially different dimensionality using feature extraction under pairwise proximity constraints (Ham et al., 2005). In some sense, manifold alignment performs registration in the feature space and matches corresponding samples, where the correspondence is defined by a series of proximity graphs encoding some

prior knowledge of interest (e.g. co-location, class consistency). An intuition of how manifold alignment functions is provided in Fig. 1. Its application to remote sensing data is relatively recent: in Tuia et al. (2014), authors presented the semi-supervised manifold alignment method (SSMA), which gathers all properties above, but at the price of requiring labeled pixels in all domains to perform the alignment. Yang and Crawford (2016b) study issued of spatial consistency and in Yang and Crawford (2016a) they propose a multi-scale alignment procedure not relying on labels in all domains. Finally, true colour visualization for hyperspectral data was tackled in Liao et al. (in press).

In this paper, we study the effectiveness of the nonlinear counterpart of SSMA, the Kernel Manifold Alignment (KEMA, Tuia and Camps-Valls (2016)), as well as its relevance for remote sensing problems. KEMA is a flexible, scalable, and intuitive method for aligning manifolds. KEMA provides a flexible and discriminative projection function, only exploits a few labeled samples (or semantic ties (Montoya-Zegarra et al., 2013), when images are roughly registered – see Section 3.3) in each domain, and reduces to solving a simple generalized eigenvalue problem.

KEMA is introduced in Section 2. In Section 3, we test it in several real-life scenarios, including multi-temporal and multi-source very high resolution image classification problems, as well as in the challenging task of making a model shadow-invariant in hyperspectral image classification. Section 4 concludes the paper.

[Figure 1 about here.]

2. Kernel Manifold Alignment (KEMA)

In this section, we detail the KEMA method. We first recall the linear counterpart, the SSMA method (Wang and Mahadevan, 2011). Noting the main problems of this method, we introduce KEMA as a solution to address them. The reader interested in more theoretical details of KEMA can find them in Tuia and Camps-Valls (2016). Code can be found at the URL: `https://github.com/dtuia/KEMA`.

2.1. Notation

To fix notation, we consider a series of M domains. For each one of them, we have a data set: $\mathcal{M} := \{\mathbf{x}_i^m \in \mathbb{R}^{d_m} | i = 1, \ldots, n_m\}$, where n_m

is the number of samples issued from domain m with data dimensionality d_m, and $m = 1, \ldots, M$. Some of the pixels in $\mathbf{x}_i$ are labeled $(l_1, ..., l_M)$, and most are unlabeled. From one domain to another, the data are not necessarily semantically paired, i.e. $n_1 \neq n_m \neq n_M$, nor it is mandatory that all domains have the same dimension, i.e. $d_1 \neq d_m \neq d_M$.

2.2. Semi-supervised manifold alignment (SSMA)

The linear SSMA method was originally proposed in Wang and Mahadevan (2011) and successfully adapted to remote sensing problems in Tuia et al. (2014). The SSMA method aligns data from all M domains by projecting them into a common *latent space* using a set of domain-specific projection functions, $\mathbf{f}^m$, collectively grouped into the projection matrix $\mathbf{F} := [\mathbf{f}^1, \ldots, \mathbf{f}^M]^\top$. The latent space has two properties: it is discriminant for classification and respects the original geometry of each manifold. To do so, SSMA tries to find a data projection matrix $\mathbf{F}$ that maximizes the following cost function

$$\mathcal{L} = \frac{\mu \texttt{GEO} + \texttt{SIM}}{\texttt{DIS}},$$

where we aim to maximize a topology/geometry (`GEO`) and a class similarity (`SIM`) terms while minimizing a class dissimilarity term (`DIS`) between all samples, and $\mu > 0$ is a parameter controlling the contribution of the similarity and the topology terms. The three terms correspond to:

1. a geometry-preservation term, `GEO`, forcing the local geometry of each manifold to remain unchanged, i.e. penalizing projections mapping neighbors in the input space far from each other,

$$\begin{aligned} \texttt{GEO} &= \sum_{m=1}^{M} \sum_{i,j=1}^{n_m} W_g^m(i,j) \|\mathbf{f}^{m\top}\mathbf{x}_i^m - \mathbf{f}^{m\top}\mathbf{x}_j^m\|^2 \\ &= \operatorname{tr}(\mathbf{F}^\top \mathbf{X} \mathbf{L}_g \mathbf{X}^\top \mathbf{F}), \end{aligned} \tag{1}$$

   where W_g^m is a similarity matrix returning the value 1 if two pixels of domain m are neighbours in the original feature space and 0 otherwise. W_g^m is typically a k-NN graph. $\mathbf{L}_g$ is the $(\sum_m n_m \times \sum_m n_m)$ graph Laplacian issued from the similarity matrices $\mathbf{W}_g^m$, stacked in a block-diagonal matrix. All the out-of-diagonal blocks of $\mathbf{W}_g$ are empty, since we do not want to preserve neighbourhood relationships between the images.

2. a class similarity term, SIM, penalizing projections mapping samples of the same class far from each other,

$$\begin{aligned}\texttt{SIM} &= \sum_{m,m'=1}^{M} \sum_{i,j=1}^{l_m,l_{m'}} W_s^{m,m'}(i,j)\|\mathbf{f}^{m\top}\mathbf{x}_i^m - \mathbf{f}^{m'\top}\mathbf{x}_j^{m'}\|^2 \\ &= \mathrm{tr}(\mathbf{F}^\top\mathbf{X}\mathbf{L}_s\mathbf{X}^\top\mathbf{F}),\end{aligned} \tag{2}$$

where $W_s^{m,m'}$ is a similarity matrix returning the value 1 if two pixels from domains m and m' belong to the same class. These are the tie points performing registration in the spectral space, and are used to match the images to each other.

3. a class dissimilarity term, DIS, penalizing projections mapping pixels of different classes close to each other.

$$\begin{aligned}\texttt{DIS} &= \sum_{m,m'=1}^{M} \sum_{i,j=1}^{l_m,l_{m'}} W_d^{m,m'}(i,j)\|\mathbf{f}^{m\top}\mathbf{x}_i^m - \mathbf{f}^{m'\top}\mathbf{x}_j^{m'}\|^2 \\ &= \mathrm{tr}(\mathbf{F}^\top\mathbf{X}\mathbf{L}_d\mathbf{X}^\top\mathbf{F}),\end{aligned} \tag{3}$$

where $W_d^{m,m'}$ is a dissimilarity matrix returning the value 1 if two pixels from domains m and m' belong to different classes. These tie points prevent the solution to collapse in a single point and, together with the SIM term, foster the latent space to be discriminative.

Now, by combining Eqs. (1)-(3), it is straightforward to show that the solution boils down to finding the last eigenvalues of the following generalized eigenproblem (Wang and Mahadevan, 2011), which is directly derived:

$$\mathbf{X}(\mu\mathbf{L}_g + \mathbf{L}_s)\mathbf{X}^\top\boldsymbol{\varphi} = \lambda\mathbf{X}\mathbf{L}_d\mathbf{X}^\top\boldsymbol{\varphi}, \tag{4}$$

where $\mathbf{X}$ is a $(d \times \sum_m n_m)$ block-diagonal matrix containing the data from the different domains to be aligned. $\boldsymbol{\varphi}$ is the researched common projection matrix of size $d \times d$, with $d = \sum_{m=1}^{M} d_m$. The rows of $\boldsymbol{\varphi}$ contain a block of projectors for each domain, scaled by $\lambda^{1/2}$, in a particular block structure:

$$\mathbf{F} = \lambda^{\frac{1}{2}}\boldsymbol{\varphi} = \begin{bmatrix}\mathbf{f}^1 \\ \vdots \\ \mathbf{f}^M\end{bmatrix} = \begin{bmatrix} f_{1,1} & \cdots & f_{1,d} \\ \vdots & \ddots & \vdots \\ f_{d_1,1} & \cdots & f_{d_1,d} \\ f_{d_1+1,1} & \cdots & f_{d_1+1,d} \\ \vdots & \ddots & \vdots \\ f_{d,1} & \cdots & f_{d,d} \end{bmatrix}, \tag{5}$$

where the eigenvectors for the first domain are highlighted in green.

Once the projection matrix $\boldsymbol{\varphi}$ is obtained, any sample $\mathbf{x}_i^m \in \mathbb{R}^{d_m \times 1}$ from domain m (one of the domains considered) can be projected in the latent space by using the corresponding $(d_m \times d)$ block of eigenvectors $\mathbf{f}^m$:

$$\mathcal{P}(\mathbf{x}_i^m) = \mathbf{f}^{m\top}\mathbf{x}_i^m. \tag{6}$$

As for (k)PCA and other methods based on eigen-decomposition, the data can be projected onto a subspace of dimension p lower than d by simply using only the first $p \ll d$ columns of $\mathbf{f}^m$. In this sense, SSMA leaves some control on the dimensionality of the latent space for class separation.

2.3. Kernel Manifold Alignment (KEMA)

The idea behind *kernelization* is to map the data into a high dimensional Hilbert space $\mathcal{H}$ with the mapping function $\boldsymbol{\phi}(\cdot) : \mathbf{x} \mapsto \boldsymbol{\phi}(\mathbf{x}) \in \mathcal{H}$ such that the mapped data is better suited for solving our problem. This technique has found wide adoption in many remote sensing data analysis problems (Camps-Valls and Bruzzone, 2009). In practice, computing this mapping explicitly can be prohibitive due to its high dimensionality. This can be avoided by expressing the problem in terms of dot products within $\mathcal{H}$. We can then define an easy-to-compute kernel function $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \boldsymbol{\phi}(\mathbf{x}_i), \boldsymbol{\phi}(\mathbf{x}_j) \rangle_{\mathcal{H}}$ returning similarities between mapped samples without having to compute $\boldsymbol{\phi}(\cdot)$ explicitly.

In the multi-modal setting considered here, we would have to map the M datasets to M Hilbert spaces $\mathcal{H}_m$ of dimension H_m, $\boldsymbol{\phi}_m(\cdot) : \mathbf{x} \mapsto \boldsymbol{\phi}_m(\mathbf{x}) \in \mathcal{H}_m$, $m = 1, \ldots, M$. Then, we replace all the samples with their mapped feature vectors. The `GEO`, `SIM` and `DIS` terms become:

$$\texttt{GEO} = \sum_{m=1}^{M} \sum_{i,j=1}^{n_m} W_g^m(i,j) \|\mathbf{u}^{m\top} \boldsymbol{\phi}(\mathbf{x}_i)^m - \mathbf{u}^{m\top} \boldsymbol{\phi}(\mathbf{x}_j)^m\|^2$$
$$= \text{tr}(\mathbf{U}^\top \boldsymbol{\Phi} \mathbf{L}_g \boldsymbol{\Phi}^\top \mathbf{U}) \tag{7}$$
$$\texttt{SIM} = \sum_{m,m'=1}^{M} \sum_{i,j=1}^{l_m, l_{m'}} W_s^{m,m'}(i,j) \|\mathbf{u}^{m\top} \boldsymbol{\phi}(\mathbf{x}_i)^m - \mathbf{u}^{m'\top} \boldsymbol{\phi}(\mathbf{x}_j)^{m'}\|^2$$
$$= \text{tr}(\mathbf{U}^\top \boldsymbol{\Phi} \mathbf{L}_s \boldsymbol{\Phi}^\top \mathbf{U}), \tag{8}$$
$$\texttt{DIS} = \sum_{m,m'=1}^{M} \sum_{i,j=1}^{l_m, l_{m'}} W_d^{m,m'}(i,j) \|\mathbf{u}^{m\top} \boldsymbol{\phi}(\mathbf{x}_i)^m - \mathbf{u}^{m'\top} \boldsymbol{\phi}(\mathbf{x}_j)^{m'}\|^2$$
$$= \text{tr}(\mathbf{U}^\top \boldsymbol{\Phi} \mathbf{L}_d \boldsymbol{\Phi}^\top \mathbf{U}), \tag{9}$$

As for the SSMA case, combining Eqs. (7)-(9) leads to a generalized eigendecomposition problem:

$$\boldsymbol{\Phi}(\mathbf{L}_g + \mu \mathbf{L}_s) \boldsymbol{\Phi}^\top \mathbf{U} = \lambda \boldsymbol{\Phi} \mathbf{L}_d \boldsymbol{\Phi}^\top \mathbf{U},$$

where $\boldsymbol{\Phi}$ is a block diagonal matrix containing the data matrices $\boldsymbol{\Phi}^m = [\boldsymbol{\phi}_m(\mathbf{x}_1), \ldots, \boldsymbol{\phi}_m(\mathbf{x}_{n_m})]^\top$ and $\mathbf{U}$ contains the eigenvectors organized in rows for the particular domain defined in Hilbert space $\mathcal{H}_m$, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_H]^\top$ where $H = \sum_{m=1}^{M} H_m$. As stressed above, $\boldsymbol{\Phi}$ and $\mathbf{U}$ live in a high dimensional space that might be very costly or even impossible to compute. Therefore, we express the eigenvectors as a linear combination of mapped samples using the Representer's theorem (Yan et al., 2007) , $\mathbf{u}^m = \boldsymbol{\Phi}^m \boldsymbol{\alpha}^m$ (or $\mathbf{U} = \boldsymbol{\Phi} \boldsymbol{\Lambda}$ in matrix notation):

$$\mathbf{K}(\mathbf{L}_g + \mu \mathbf{L}_s) \mathbf{K} \boldsymbol{\Lambda} = \lambda \mathbf{K} \mathbf{L}_d \mathbf{K} \boldsymbol{\Lambda}, \tag{10}$$

where $\mathbf{K}$ is a block diagonal matrix containing the kernel matrices $\mathbf{K}^m$. Now the eigenproblem becomes of size $n \times n$ instead of $d \times d$, and we can extract a maximum of n components.

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\alpha}^1 \\ \vdots \\ \boldsymbol{\alpha}^M \end{bmatrix} = \begin{bmatrix} \alpha_{1,1} & \ldots & \alpha_{1,n} \\ \vdots & \ddots & \vdots \\ \alpha_{n_1,1} & \ldots & \alpha_{n_1,n} \\ \alpha_{n_1+1,1} & \ldots & f_{\alpha_1+1,n} \\ \vdots & \ddots & \vdots \\ \alpha_{n,1} & \ldots & \alpha_{n,n} \end{bmatrix}. \tag{11}$$

This dual formulation is advantageous when dealing with very high dimensional datasets, $d \gg n$ for which the SSMA problem is not well-conditioned. Operating in Q-mode endorses the method with numerical stability and computational efficiency in current high-dimensional problems, e.g. when using Fisher vectors or deep features for data representation. As shown in Tuia and Camps-Valls (2016), KEMA performs very well when aligning input spaces of such high dimension. This type of problems with much more dimensions than points are becoming more and more prominent in remote sensing (Lagrange et al., 2015; Marmanis et al., in press). In this sense, even KEMA with a linear kernel (which corresponds to the SSMA solution) becomes a valid solution for these problems, as it has all the advantages of methods related to (kernel) Canonical Correlation Analysis ((k)CCA (Lai and Fyfe, 2000)), but can also deal with unpaired data.

Projection of a new test vector $\mathbf{x}_i^m$ to the latent space requires first mapping it to its corresponding kernel form $\mathbf{K}_i^m$ and then applying the corresponding projection vector $\boldsymbol{\alpha}^m$ defined therein:

$$\mathcal{P}(\mathbf{x}_i^m) = \mathbf{u}^{m\top}\boldsymbol{\Phi}_i^m = \boldsymbol{\alpha}^{m\top}\boldsymbol{\Phi}^{m\top}\boldsymbol{\Phi}_i^m = \boldsymbol{\alpha}^{m\top}\mathbf{K}_i^m, \tag{12}$$

where $\mathbf{K}_i^m$ is a vector of kernel evaluations between sample $\mathbf{x}_i$ and all samples from domain m used to define the projections $\boldsymbol{\alpha}^m$. Therefore, projection to the kernel latent space is possible through the use of dedicated reproducing kernel functions.

2.4. Computational complexity of KEMA

A shortcoming of KEMA may be its computational cost. As for the SSMA method (and CCA-based approaches), KEMA is based on an eigendecomposition, for which many efficient solvers are available. Their cost is comparable to kernel and linear canonical correlation analysis techniques, respectively. Compared to SSMA, the KEMA problem to be solved might be of larger side, since SSMA involves a $d \times d$ decomposition (with $d = \sum_m d_m$ being the sum of the dimensionality of all domains), while KEMA involves a $n \times n$ decomposition (with $n = \sum_m n_m$ being the total number of samples involved in the kernel matrices). Hence, for small d (for instance, when aligning VHR images in $\mathbb{R}^4$) SSMA is computationally more interesting, while for large d (for instance, when considering DeCAF deep features in $\mathbb{R}^{4096}$ for each domain), KEMA will involve a smaller cost and higher stability (see also Tuia and Camps-Valls (2016) for a more detailed discussion). Another

issue is related with storage costs, since KEMA stores a set of $n_m \times n_m$ kernel matrices and requires the evaluation of n_m kernel functions for each sample at test time: to alleviate such costs, we proposed two approaches based on reduced rank approximation and random features. The interested reader can find all the details in Tuia and Camps-Valls (2016).

3. Experimental Results

In this section, we present experimental results in three challenging remote sensing problems: multi-temporal / multi-source VHR classification, shadow removal in hyperspectral images, and multi-source image alignment without labels.

3.1. Multi-temporal and multi-sensor VHR classification

The first experiment is a direct comparison to the multi-source experiment reported in Tuia et al. (2014). We consider three VHR images (Fig. 2) depicting peri-urban settlements:

- *Prilly*: the first image is acquired by the WorldView-2 VHR satellite (8 visible and near-infrared bands) over Prilly, a residential neighborhood of Lausanne, Switzerland. The image is acquired on August 2, 2011 and has been pansharpened using the Gram-Schmid transform to a resolution of approximatively 0.7m.

- *Malley*: the second image is also acquired by WorldView-2 over another residential neighborhood of Lausanne, Montelly. The image is acquired on September 29, 2010 and has also been pansharpened using the Gram-Schmid transform to 0.7m.

- *Zurich*: the third image is acquired by the QuickBird satellite (4 bands, RGB- NIR) over a residential neighborhood of Zurich, Switzerland. The image has been acquired on October 6, 2006 and pansharpened.

[Figure 2 about here.]

[Figure 3 about here.]

For each image, a ground truth consisting of 9 classes is available (see bottom row of Fig. 2). We follow the experimental protocol of Tuia et al. (2014): we use the original DN values of each image as input features. From all the available labeled pixels in each image, 50% are kept apart as the testing set. The remaining 50% are used to extract the labeled and unlabeled pixels composing the $\mathbf{x}_m$ sets. We then extract $l_1 = 100$ labeled pixels per class from what we call the leading domain image, which is the image carrying most labeled samples (we take each image in turn as the leading domain image). Experiments run on smaller labeled sets led to the same conclusions, only with lower performance for all models. In our setting, we also need labeled pixels from the two other acquisitions: we tested an increasing additional of labeled samples, $l_2 = l_3 = [10, 30, 50, 90]$ pixels per class. As in Tuia et al. (2014), the unlabeled examples are selected using an iterative clustering algorithm, the bisecting k-means (Kashef and Kamel, 2009), which runs k-means with 2 clusters iteratively, by splitting the current largest cluster in the dataset. This way, we sample 500 unlabeled examples per each image source. We use the labeled and unlabeled examples to extract both the SSMA and KEMA projections and then project all images in the latent space. Finally, we use all the projected labeled examples to train a single classifier (a linear SVM) in the latent space. This classifier is used to predict all the test pixels of all three images at once (i.e. no specific training is performed for the specific images separately).

In KEMA, we use RBF kernels with the bandwidth σ_m fixed as half the median distance between the samples of the specific image (labeled or unlabeled). By doing so, we allow different kernels in each domain, thus tailoring the similarity function to the data structure observed (Tuia and Camps-Valls, 2016). To build the graph Laplacians, we used a series of graphs built using k-NN graphs with $k = 9$ as in Tuia et al. (2014). We validated the optimal number of dimensions, as well as the optimal C parameter in the SVM classifier using the labeled samples in a cross-validation setting. Finally, as in Tuia et al. (2014) we add a baseline, which is the classifier learned with the original features. Since the Zurich image has a different input space than the two others, only the common bands between QuickBird and WorldView-2 are considered.

The results are reported in Fig. 3. Two distinct behaviours are observed:

- Diagonal blocks of Fig. 3 (when predicting the leading domain image, which carried most labels): in this cases, the predictions of KEMA are

better than those of SSMA by $\approx 2-5\%$ and remain consistent when adding samples from the other domains. This means that the images are aligned correctly and the inclusion of labels from other images does not disturb the classifier (as in the 'no adaptation' case). On the contrary, adding labeled samples from the other images is beneficial, as one can observe by comparing the KEMA results with the optimal case obtained when using only the 100 labeled pixels per class from the leading image (green bars): the final prediction is 5-10% more accurate than in the case, where the leading image is used alone (i.e. without extra labeled samples coming from the other acquisitions). This means that the extra labeled are aligned correctly, since the classifier trained with $100 + l_2 + l_3$ aligned examples per class outperforms the one obtained with 100 pixels per class.

- Off-diagonal blocks of Fig. 3 (when predicting the two other, scarcely labeled images): in the off-diagonal blocks we can observe a constant improvement of the results obtained by SSMA, which corresponds already to a strong improvement over the 'no adaptation' case. The improvement of KEMA with respect to the latter is more striking ($\approx 5-15\%$) when using little labels from the test images. In comparison to SSMA we observe a constant $3-5\%$ improvement.

3.2. Shadow compensation in hyperspectral image classification

In this experiment, we aim at compensating the reduction in reflectance due to a shadow casted by a large cloud. We consider a hyperspectral image acquired by the CASI sensor over Houston (see Fig. 4a)). The data were originally provided to the community for the data fusion contest 2013 (Debes et al., 2014)[1]. The contest was framed as a land use classification contest, where 15 land use classes were to be detected using two data sources: the hyperspectral image mentioned and a LiDAR DSM. The specificity of the contest is that the test pixels are partly located under a shadow cast by clouds (see Fig. 5d), thus raising the need for compensation algorithms. In our analysis, we compare three strategies for handling the hyperspectral image: using it without further processing ('Raw'), applying a histogram matching (HM) on the shadowed area (the strategy also used before extracting features

[1]The data can be found at http://www.grss-ieee.org/community/technical-committees/data-fusion/

in Tuia et al. (2015)), SSMA and the proposed KEMA aligning the pixels under the shadow and those illuminated. For HM, SSMA and KEMA, we define the shadowed pixels by defining a cloud mask by thresholding band 130 and then applying morphological operators to remove salt and pepper noise within the bigger connected component representing the shadow (cf. the mask in Fig. 4d).

In this experiment, we align the dataset using 20 labeled pixels per class. We use only classes occurring in both domains (shadowed and illuminated). Additionally, we sample randomly 200 unlabeled pixels per class. We align the 144 reflectance band of the two domains to each other. As for the first example, the kernel used in KEMA is an RBF with σ_m bandwidth estimated as half of the median distance between the points of the domain. This is very important in this experiment, since it allows to have a much narrower bandwidth for the kernel acting on the shadowed domain than the one used in the illuminated domain. We classify using a support vector machine with RBF kernel, whose parameters are found by cross validation ($\sigma \in [0.01, 0.1]$, $C \in [1, 100]$). We train the classifier on 95% of the training set available and predict on two validation datasets: the entire test set and the test samples under the shadowed area. We consider three feature sets, as detailed in Table 1, and use them in three experiments: the first using only the HSI, the second adding LiDAR-derived features, and the third adding contextual features extracted from the optical bands (this type of filters is known to improve accuracy of classifiers considerably, as they break the assumption of spatial independence of pixel features (Khatami et al., 2016)). A last setting, called MV, uses all features, and also applies a majority voting on the solution. Note that the LiDAR features do not change in the different experiments, as only the HSI (and the contextual filters applied to those) are affected by the normalization with SSMA, KEMA or HM. The experiments are repeated 10 times by varying the labeled pixels in KEMA and those picked for classification: therefore we report the average and standard deviation.

[Figure 4 about here.]

[Table 1 about here.]

The projections extracted by KEMA are visualized in Fig. 4 (geographical space, for projections $[1-3]$ and $[4-6]$) and Fig. 6 (feature space for dimensions $[1-3]$). At a first glance, the aligned features seem to be less

dependent on the presence of the shadow than the original image (some artifacts remain at the border, due to the binary nature of the cloud mask). This is confirmed in the feature space, where the two domain seem correctly aligned both in terms of classes and domains.

[Figure 5 about here.]

[Figure 6 about here.]

The classification results reported in Table 2 confirm these intuitions: KEMA is able to provide higher classification performance by working in the aligned latent space. The use of the raw images ('Raw' column), even though satisfactory on the global test set (OA of 85.5% in the best case), completely fails under the shadowed area (best OA: 23.8%). This can be also appreciated in the classification maps (first row in Fig. 5): from the maps it is clear that the shadow drains most of the shadowed pixels in the class 'water' (in cyan). Even including LiDAR features (right column of Fig. 5) does not solve entirely the problem and basically shifts most of the shadowed pixels in the class 'highway' (in beige). Using HM improves drastically the solution under the shadow, since the accuracy goes from 23.8% to 75.1% on average. Histogram matching solves the problem globally and provides the scaling and centering of the histogram necessary to make the images more similar, but still fails at accounting for subtle local variations, thus still leading to heavy misclassifications in the final map, in particular the highway being classified as buildings (see second row of Fig. 5). Finally, KEMA solves the problem locally by the flexibility of the kernel mapping: the accuracies are the highest (also matching those of the winners of the contest, who created an entirely *ad-hoc* system for this specific image) and reach an average of 94.3%, but also show an almost identical performance in the shadowed area (91.5%). The alignment has made the two domains more similar and the mismatch between domains becomes almost invisible in the classification maps (third row of Fig. 5). Compared with the linear SSMA, KEMA provides indeed similar results overall, but provides a more desirable solution in the shadowed area of the test set (with improvements in accuracy between 12% when only the spectral bands are used to 1% when all the additional features are injected), thus showing again the advances of using a flexible mapping via the use of kernels.

[Table 2 about here.]

3.3. Multi-source image classification without labels

In the last experiment, we break the requirement for labeled data in all domains. To do so, we need to reduce the flexibility of KEMA by adding a requirement on *partial spatial overlap between the scenes*. This can be understood as follows: KEMA is a spectral registration method that uses the labels as anchor points (or *ties*) to register the domains spectrally. If one of the domains is unlabeled, it is not possible to register them, since the $\mathbf{L}_s$ and $\mathbf{L}_d$ matrices in Eq. (10) cannot be computed. As a consequence, we can only preserve the inner domain geometry using $\mathbf{L}_g$, but there is no way to find the matching between domains.

[Figure 7 about here.]

[Figure 8 about here.]

When using geographical data (as remote sensing data), a special case can break this requirement: whenever the domains are (at least partially) co-located in space. In this case, represented in Fig. 7, the two images share a spatial region, where we can co-locate objects, for instance by feature keypoint matching or by manual registration. Once these matches are found, they can be used to build the matrix $\mathbf{L}_s$, since, even if we ignore their class, we know that the pixels of the objects matched belong to the same class (they are known as *semantic ties* (Montoya-Zegarra et al., 2013)). This type of weakly supervised alignment has been recently proposed in Marcos et al. (2016) and we use it here prior to aligning the data spaces with KEMA. The experiment is set as follows:

- We use an RGB image (0.6m resolution) over the area of Prilly, a neighbourhood of Lausanne, Switzerland as source domain. The area is labeled into five classes (roads, buildings, trees, grass and shadows) by manual photo-interpretation, see Fig. 8a.

- An FCIR (false colour infrared with NIR-G-B bands) ortho-photo of the area of Renens (another neighbourhood of Lausanne), at 0.25 cm resolution, is used as target domain and the labels are this time kept hidden (they are only used for validation), see Fig. 8c.

- To find the projections with KEMA, we use an overlapping area between the two images. The overlapping areas are not registered nor

they are at the same spatial resolution: to match them, we provide 40 tie object by manual drawing in both images (the operation takes less than 5 minutes), see Fig. 8b.

We use the labels in the source and the semantic ties to construct the $\mathbf{L}_s$ matrix. For the $\mathbf{L}_d$ matrix, we extracted the graph Laplacian from a dissimilarity matrix with values 1 for pixels from different classes in the source and 0.5 when issued from different objects in the semantic ties. We give a smaller penalization in the latter case, since two pixels coming from different objects can still belong to the same class. Once the domains are aligned, we train a linear SVM with 100 labeled pixels per class from the source domain (the RGB image) and test 400 pixels per class in the target domain (the FCIR image).

The projections retrieved are illustrated in Fig. 9: as for the previous examples, KEMA shows aligned data spaces, but also discriminative in terms of objects aligned: the bottom line in Fig. 9 illustrates six objects among the 40 semantic ties used to find the alignment. Figure 10 reports the classification performance in the FCIR domain: starting with six dimensions, KEMA outperforms the case where the RGB image is used to predict the FCIR one without any adaptation[2]: when using 13 dimensions, KEMA performs comparably to a model trained on labeled pixels form the target domain itself (green line in the figure). We compare these results to those obtained by applying kCCA (Lai and Fyfe, 2000). We can use kCCA as a fair competitor in this case, since the images share a common geographical extent: therefore, in the common area each location is roughly seen by both views, which is a condition for kCCA to function correctly, i.e. each sample aligned is viewed by every domain[3]. In order to compute the kCCA projection, we considered each object (each semantic tie in Fig. 8b) as a sample. We used the spectrum of the most representative pixel (i.e. the pixel closest to the object average) to describe it. We then extract the kCCA projections between the 40 pairs of corresponding objects across image acquisitions. Numerically (Fig. 10), the performance of KEMA is consistently better than that of kCCA. This is probably due to two reasons: 1) the fact that KEMA does not need a

[2]To maximize the performance of the 'no alignment' case, we use the bands that share comparable wavelengths across domains: $X^s = [R, G]$, $X^t = [R, G]$.

[3]Also note that this is the reason why we could not use kCCA as a competitor in the previous examples, as there is no spatial overlap between the domains.

one-to-one correspondence and thus all the pixels in an object are taken into account for the projection and 2) that class separability is explicitly taken into account by using the labels in the source domain.

[Figure 9 about here.]

[Figure 10 about here.]

4. Conclusions

In this paper, we presented a manifold alignment method based on kernels. The presented KEMA method is a feature extractor that finds projections from all the available source domains into a joint *latent* space, where data is semantically aligned and class separability enhanced. Compared to recent manifold alignment methods, KEMA offers a more flexible framework, going beyond simple linear transformations (scalings and rotations) of the input data. KEMA exploits a few labeled samples (or semantic ties) in each domain along with the wealth of unlabeled samples. KEMA reduces to solving a simple generalized eigenvalue problem, and has very few (and interpretable) hyperparameters to tune. We successfully tested KEMA in multi-temporal and multi-source very high resolution classification tasks, as well as on the task of making a model invariant to shadows for hyperspectral imaging.

KEMA can be seen as a multivariate method for data pre-processing in general applications where multi-sensor, multi-modal, sensory data is acquired. The generality of the approach opens a wide field in remote sensing data processing applications. Even though the applications showcased in this paper are urban areas, the method is generic and can be applied to any classification problem that comes with (scarcely) labeled, multi source image data. Our next steps with KEMA involve 1) performing semi-automatic atmospheric compensation in multi-temporal settings, 2) reduce the impact of the few labeled examples needed to perform the alignment, and 3) extend KEMA for challenging regression problems.

Acknowledgements

This work has been partly supported by the Swiss National Science Foundation (grant PZ00P2-136827, http://p3.snf.ch/project-136827), and the European Reseach Council (ERC) funding under the ERC-CoG-2014 SEDAL

under grant agreement 647423. The authors would like to thank the Hyperspectral Image Analysis group and the NSF Funded Center for Airborne Laser Mapping (NCALM) at the University of Houston for providing the hyperspectral data used in the shadow correction experiment, and the IEEE GRSS Data Fusion Technical Committee for organizing the 2013 Data Fusion Contest. They also would like to thank Swisstopo (www.swisstopo.admin.ch) for making available the FCIR orthophotos for academic use.

5. References

References

Amorós-López, J., Gómez-Chova, L., Alonso, L., Guanter, L., Zurita-Milla, R., Moreno, J., Camps-Valls, G., in press. Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. Int. J. Appl. Earth Obs. Geoinf. .

Asner, G.P., Broadbent, E.N., Oliveira, P.J.C., Keller, M., Knapp, D.E., Silva, J.N.M., 2006. Condition and fate of logged forests in the Brazilian Amazon. Proc. Nat. Ac. Science (PNAS) 103, 12947–12950.

Asner, G.P., Knapp, D., Broadbent, E., Oliveira, P., Keller, M., Silva, J., 2005. Ecology: Selective logging in the Brazilian Amazon. Science 310, 480–482.

Brunner, D., Lemoine, G., Bruzzone, L., 2010. Earthquake damage assessment of buildings using VHR optical and SAR imagery. IEEE Trans. Geosci. Remote Sens. 48, 2403–2420.

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

Bruzzone, L., Marconcini, M., 2010. Domain adaptation problems: A DASVM classification technique and a circular validation strategy. IEEE Trans. Pattern Anal. Mach. Intell. 32, 770–787.

Bruzzone, L., Persello, C., 2009. A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability. IEEE Trans. Geosci. Remote Sens. 47, 3180–3191.

Camps-Valls, G., Bruzzone, L. (Eds.), 2009. Kernel methods for Remote Sensing Data Analysis. Wiley & Sons, UK.

Camps-Valls, G., Tuia, D., Bruzzone, L., Benediktsson, J.A., 2014. Advances in hyperspectral image classification. IEEE Signal Proc. Mag. 31, 45–54.

Debes, C., Merentitis, A., Heremans, R., Hahn, J., Frangiadakis, N., van Kasteren, T., Liao, W., Bellens, R., Pizurica, A., Gautama, S., Philips, W., Prasad, S., Du, Q., Pacifici, F., 2014. Hyperspectral and lidar data fusion: Outcome of the 2013 GRSS Data Fusion Contest. IEEE J. Sel. Topics Appl. Earth Observ. and Remote Sensing, 7, 2405–2418.

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

Gómez-Chova, L., Tuia, D., Moser, G., Camps-Valls, G., 2015. Multimodal classification of remote sensing images: A review and future directions. Proceedings of the IEEE 103, 1560–1584.

Guanter, L., Richter, R., Kaufmann, H., 2009. On the application of the MODTRAN4 atmospheric radiative transfer code to optical remote sensing. Int. J. Remote Sens. 30, 1407–1424.

Ham, J., Lee, D.D., Saul, L.K., 2005. Semisupervised alignment of manifolds, in: Proc. Int. Workshop Artificial Intelligence and Statistics.

Izquierdo-Verdiguier, E., Laparra, V., Gómez-Chova, L., Camps-Valls, G., 2013. Encoding invariances in remote sensing image classification with SVM. IEEE Geosci. Remote Sens. Lett. 10, 981–985.

Kashef, R., Kamel, M., 2009. Enhanced bisecting k-means clustering using intermediate cooperation. Pattern Recogn. 42, 2257–2569.

Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. Remote Sens. Enviro. 177, 89–100.

Lagrange, A., Le Saux, B., Beaupere, A., Boulch, A., Chan-Hon-Tong, A., Herbin, S., Randrianarivo, H., Ferecatu, M., 2015. Benchmarking classification of earth-observation data: From learning explicit features to convolutional networks, in: Proc. IGARSS, Milan, Italy. pp. 4173 – 4176.

Lai, P.L., Fyfe, C., 2000. Kernel and nonlinear canonical correlation analysis., in: Int. J. Neural Sys., pp. 365–377.

Liao, D., Qian, D., Zhou, J., Tang, Y., in press. A manifold alignment approach for hyperspectral image visualization with natural color. IEEE Trans. Geosci. Remote Sens. .

Marcos, D., Hamid, R., Tuia, D., 2016. Geospatial correspondence for multimodal registration, in: Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV.

Marmanis, D., Datcu, M., Esch, T., Stilla, U., in press. Deep-learning earth observation classification using imagenet pre-trained networks. IEEE Geosci. Remote Sensing Lett. .

Matasci, G., Longbotham, N., Pacifici, F., M., K., Tuia, D., 2015. Understanding angular effects in VHR imagery and their significance for urban land-cover model portability: a study of two multi-angle in-track image sequences. ISPRS J. Int. Soc. Photo. Remote Sens. 107, 99–111.

Matasci, G., Tuia, D., Kanevski, M., 2012. SVM-based boosting of active learning strategies for efficient domain adaptation. IEEE J. Sel. Topics Appl. Earth Observ. 5, 1335–1343.

Montoya-Zegarra, J., Leistner, C., Schindler, K., 2013. Semantic tie points, in: Proc. IEEE WACV, Clearwater Beach, FL.

Nielsen, A.A., 2002. Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data. IEEE Trans. Im. Proc. 11, 293–305.

Olthof, I., Butson, C., Fraser, R., 2005. Signature extension through space for northern landcover classification: A comparison of radiometric correction methods. Remote Sens. Environ. 95, 290–302.

Pacifici, F., Longbotham, N., Emery, W.J., 2014. The importance of physical quantities for the analysis of multitemporal and multiangular optical very high spatial resolution images. IEEE Trans. Geosci. Remote Sens. 52, 6241–6256.

Pan, S.J., Qiang, Y., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359.

Patel, V.M., Gopalan, R., Li, R., Chellappa, R., 2015. Visual domain adaptation: a survey of recent advances. IEEE Signal Proc. Mag. 32, 53–69.

Rajan, S., Ghosh, J., Crawford, M., 2006. Exploiting class hierarchy for knowledge transfer in hyperspectral data. IEEE Trans. Geosci. Remote Sens. 44, 3408–3417.

Roth, K.L., Roberts, D.A., Dennison, P.E., Peterson, S.H., Alonzo, M., 2015. The impact of spatial resolution on the classification of plant species and functional types within imaging spectrometer data. Remote Sens. Enviro. 171, 45–57.

Sun, H., Liu, S., Zhou, S., Zou, H., 2016. Unsupervised cross-view semantic transfer for remote sensing image classification. IEEE Geosci. Remote Sens. Lett. 13, 13–17.

Sun, H., Liu, S., Zhou, S., Zou, H., in press. Transfer sparse subspace analysis for unsupervised cross-view scene model adaptation. IEEE J. Sel. Topics Appl. Earth Observ. .

Taubenböck, H., Wurm, M., Netzband, M., Zwenzner, H., Roth, A., Rahman, A., Dech, S., 2011. Flood risks in urbanized areas - multi-sensoral approaches using remotely sensed data for risk assessment. Nat. Hazards Earth Sys. Science 11, 431–444.

Tuia, D., Camps-Valls, G., 2016. Kernel manifold alignment for domain adaptation. PLoS ONE 11, e0148655.

Tuia, D., Courty, N., Flamary, R., 2015. Multiclass feature learning for hyperspectral image classification: sparse and hierarchical solutions. ISPRS J. Int. Soc. Photo. Remote Sens. 105, 272–285.

Tuia, D., Persello, C., Bruzzone, L., in press. Recent advances in domain adaptation for the classification of remote sensing data. IEEE Geosci. Remote Sens. Mag. .

Tuia, D., Volpi, M., Trolliet, M., Camps-Valls, G., 2014. Semisupervised manifold alignment of multimodal remote sensing images. IEEE Trans. Geosci. Remote Sens. 52, 7708–7720.

Verrelst, J., Clevers, J.G.P.W., Schaepman, M.E., 2010. Merging the Minnaert-k parameter with spectral unmixing to map forest heterogeneity with CHRIS/PROBA data. IEEE Trans. Geosci. Remote Sens. 48, 4014–4022.

Volpi, M., Camps-Valls, G., Tuia, D., 2015. Spectral alignment of cross-sensor images with automated kernel canonical correlation analysis. J. Int. Soc. Photo. Remote Sens. 107, 50–63.

Wang, C., Krafft, P., Mahadevan, S., 2011. Manifold alignment, in: Ma, Y., Fu, Y. (Eds.), Manifold Learning: Theory and Applications. CRC Press.

Wang, C., Mahadevan, S., 2011. Heterogeneous domain adaptation using manifold alignment, in: International Joint Conference on Artificial Intelligence (IJCAI).

Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q., Lin, S., 2007. Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Trans. Patt. Anal. Mach. Intell. 29, 40–51.

Yang, H., Crawford, M., 2016a. Domain adaptation with preservation of manifold geometry for hyperspectral image classification. IEEE J. Sel. Topics Appl. Earth Observ. 9, 543–555.

Yang, H., Crawford, M., 2016b. Spectral and spatial proximity-based manifold alignment for multitemporal hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 54, 51–64.

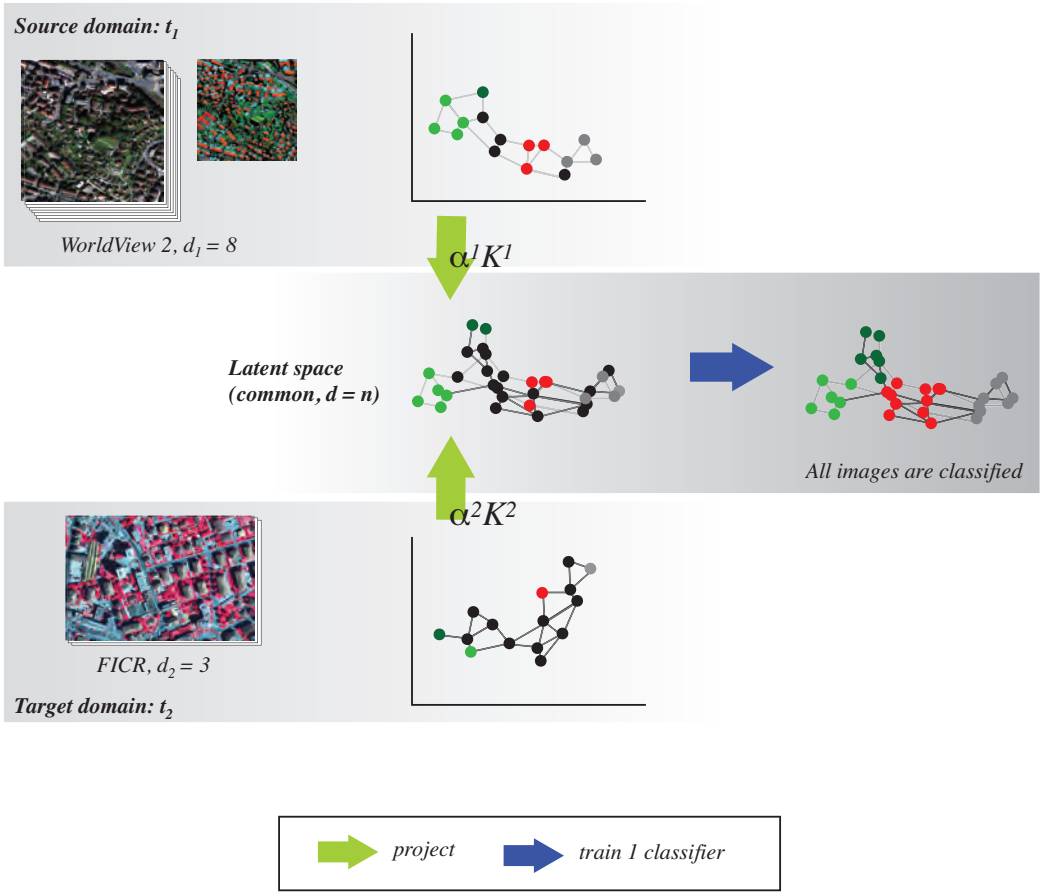


Figure 1: Illustration of KEMA aligning data distributions in a multi-sensor setting.

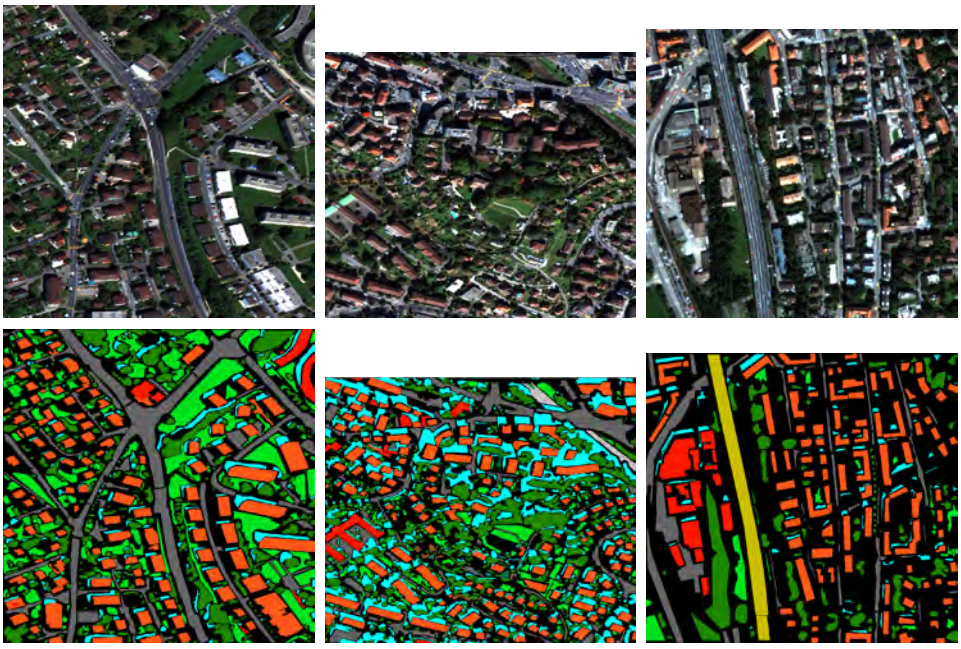


Figure 2: The WorldView-2 (WV2) and QuickBird (QB) images used in the remote sensing semantic classification experiments. Color legend: residential, meadows, trees, roads, shadows, commercial building, railway, bare soil, highway.

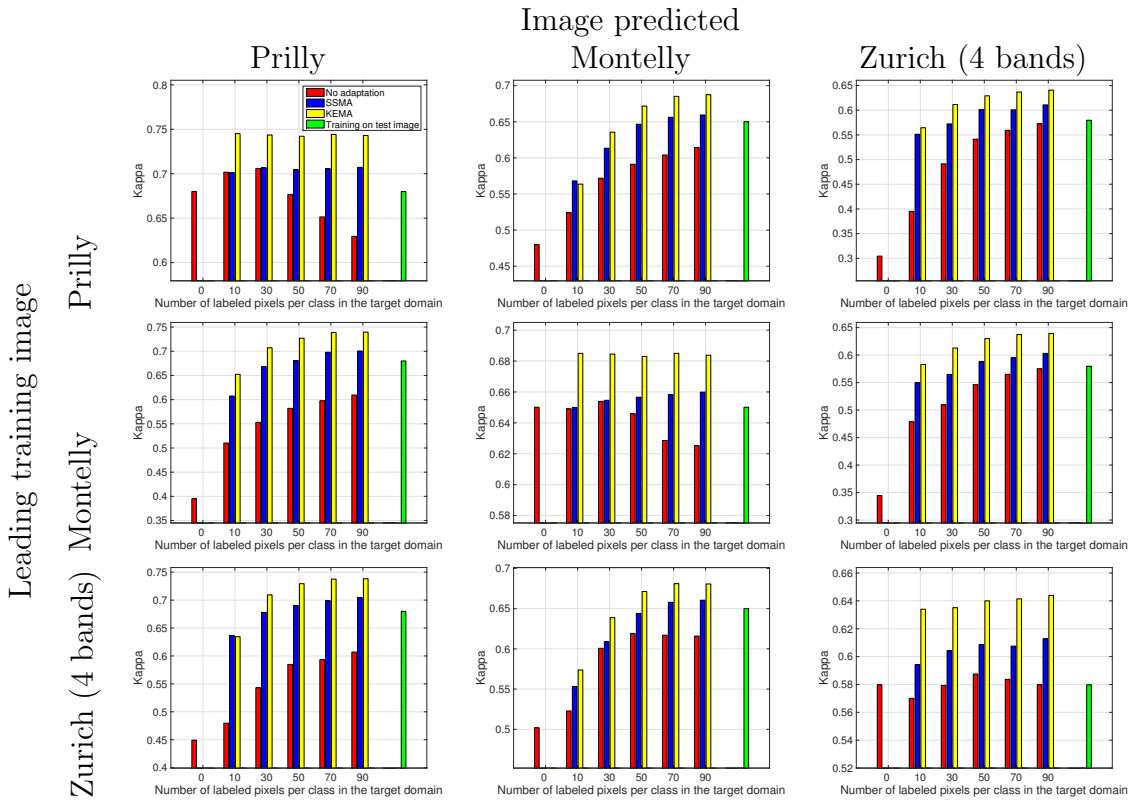


Figure 3: Numerical results for the multi-source experiment. Rows indicate the image from which 100 labeled pixels *per* class are used ($l_1 = 100$ *per* class). κ performances for increasing number of labeled pixels in the two other images ($l_2 = l_3 = [10, ..., 90]$ *per* class) are reported. Columns correspond to the image that has been used for testing. The baseline is the model obtained using 100 pixels *per* class from the test image only.

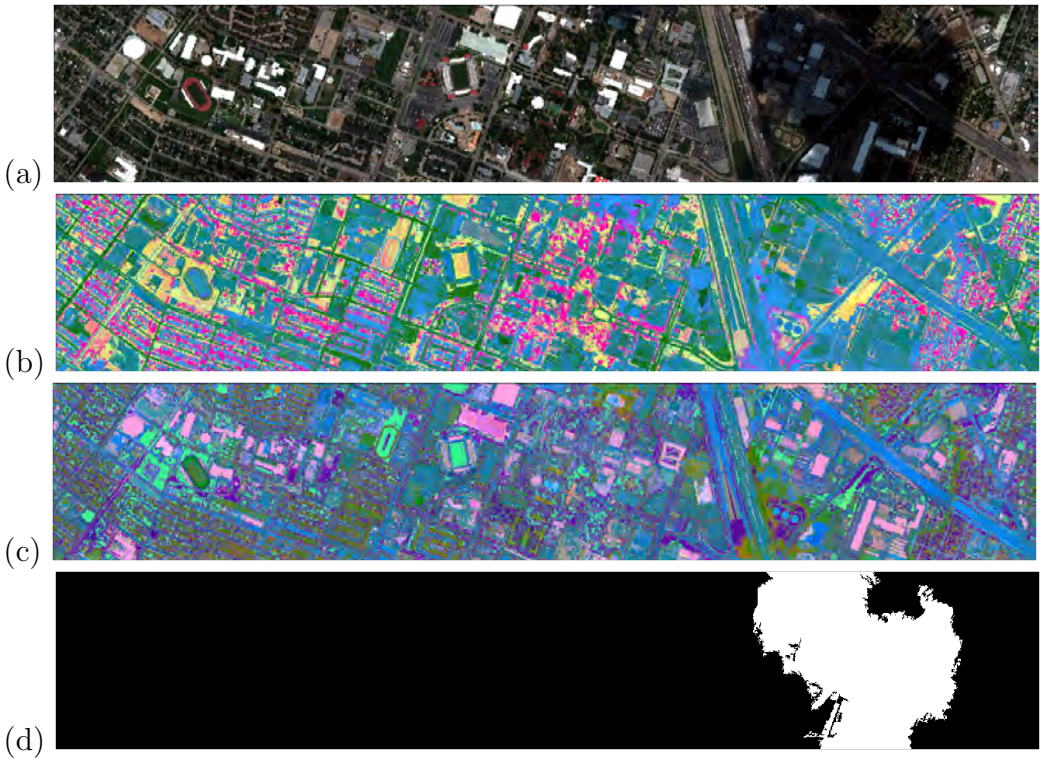


Figure 4: Domains reprojected by KEMA. (a): original CASI image. (b): first three dimensions of the latent space (R: 1, G: 2, B: 3). (c): dimensions 4-6. (d): cloud mask defining the two domains.

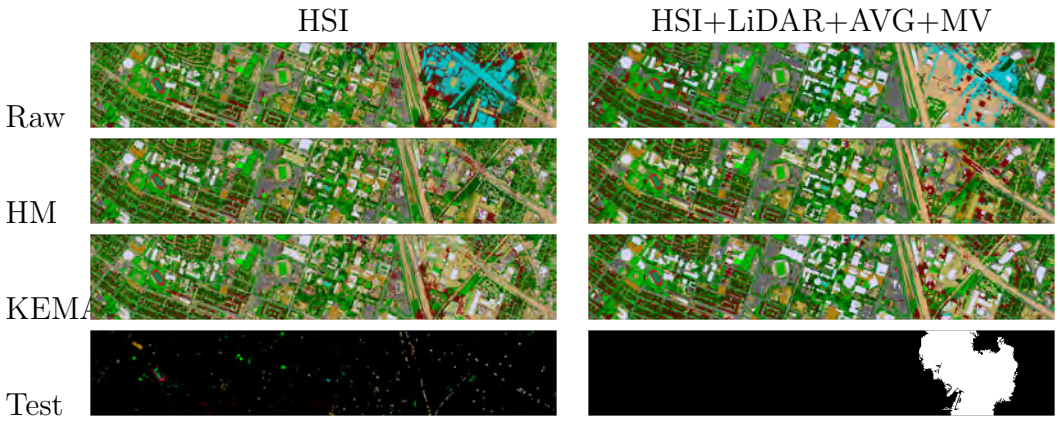
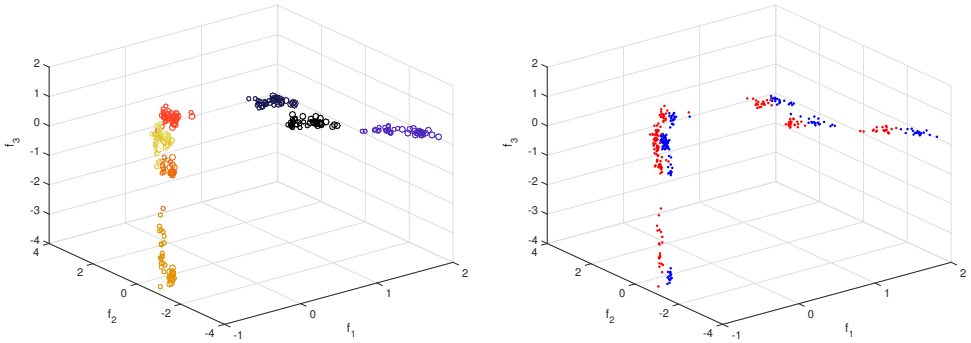


Figure 5: Classification maps for the three settings (Raw, HM and KEMA). (left) using the spectral bands; (right) performing a majority voting on the map obtained by staking HSI, LiDAR and AVG features (for averaged numerical results, see Tab. 2). Bottom line shows the test samples and the cloud mask.



(a) per class (b) per domain

Figure 6: Projection per class (a) and per domain (b, shadow is in blue and illuminated in red) for the Houston data.

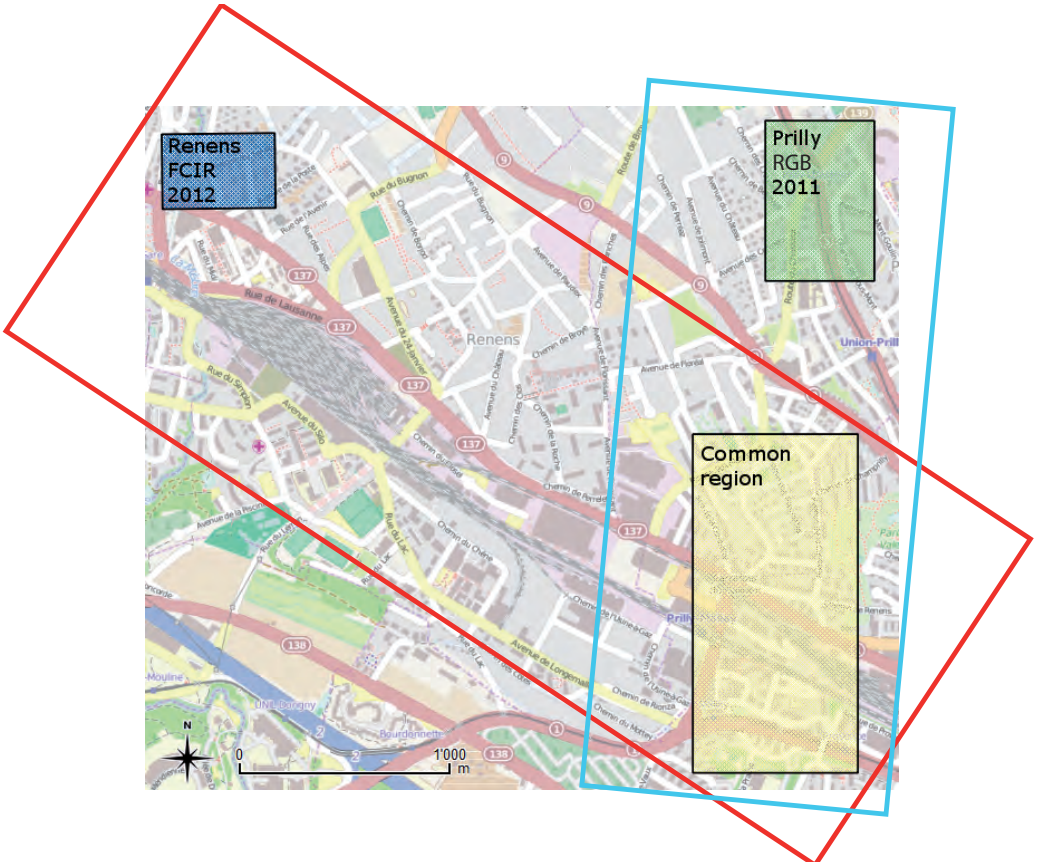
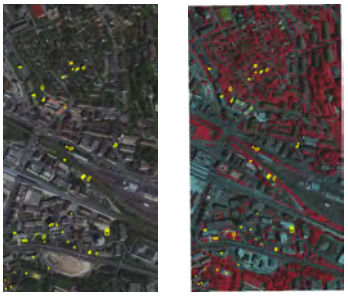


Figure 7: Setting of the multi-source experiment. The cyan square represents the source domain image (RGB) and the red square the target domain image (NIR-R-G). They share a spatial subset, where the semantic ties are used to align the domains. The dark blue, green and yellow square are the image detailed in Fig. 8, used for both the semantic ties definition and the numerical assessment.

Prilly: source domain
(RGB)
(a)



Spatially overlapping area
with semantic ties
(b)

Renens: target domain (unlabeled)
(NIR-R-G)
(c)

Figure 8: Images involved in the multi-source experiment (corresponding to the dark blue, green and yellow squares in Fig. 7).

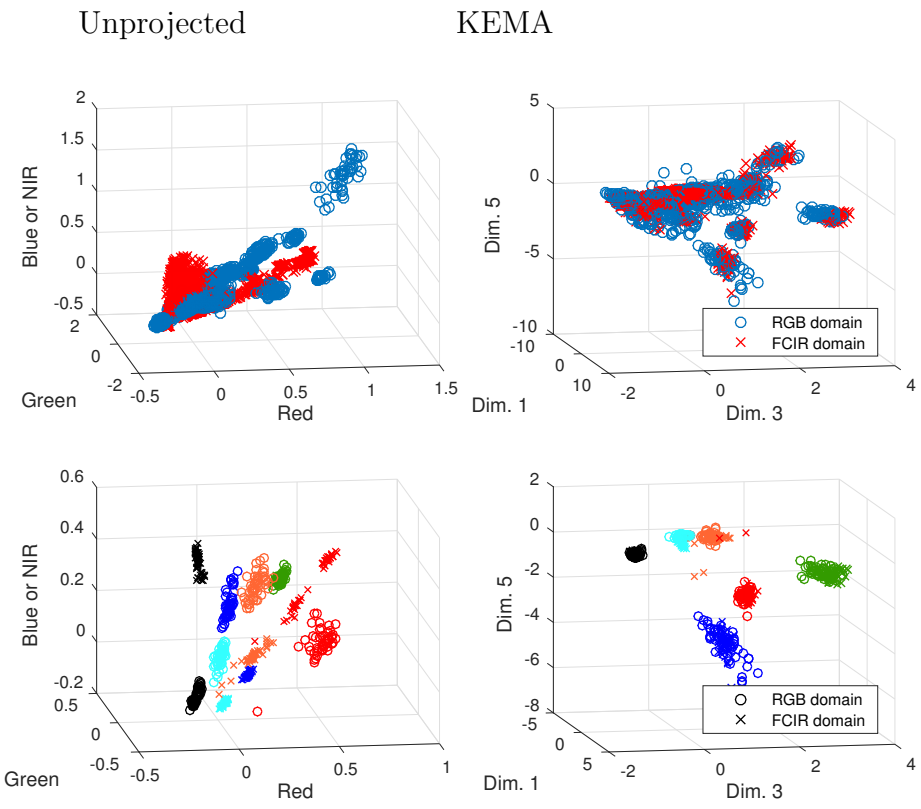


Figure 9: Projections found by KEMA, colored by domain (top) and by object in the semantic ties set (bottom, six objects shown). The left panel shows the unprojected data [x axis: R, y axis: G, z axis: NIR or B], the right panel shows the projections by KEMA [Projections 1, 3 and 5].

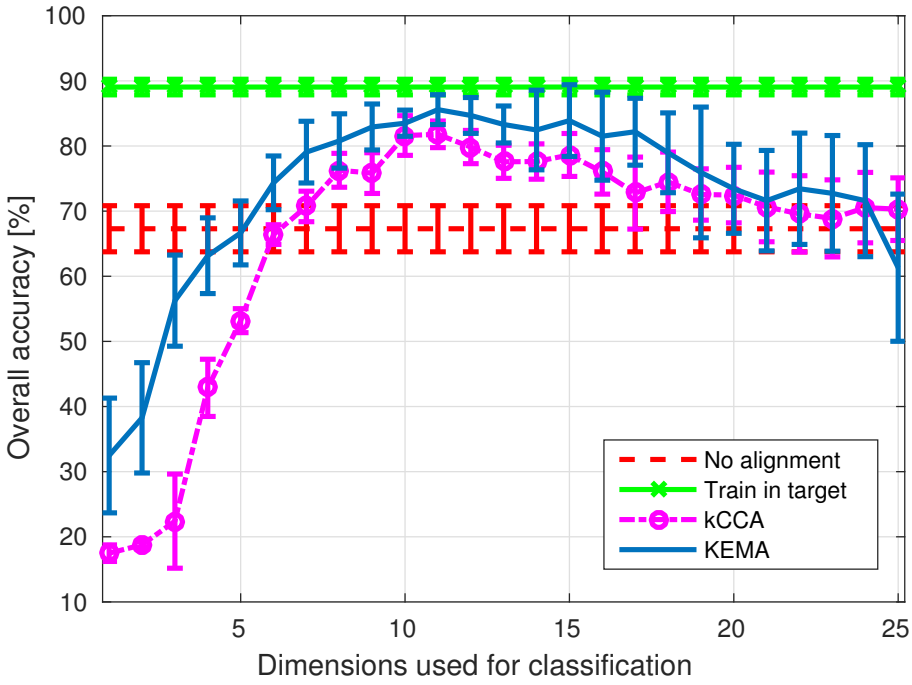


Figure 10: Classification performances by a linear SVM using the labeled samples from the source domain (RGB) as they are (red line) or projected by KEMA (blue line) or kCCA (magenta line). In green a baseline obtained by training with labeled pixels form the target domain (FCIR).

Table 1: Three feature types used in the experiment. Number in brackets is the number of features involved in each group.

	Raw / HM	KEMA
HSI	Hyperspectral bands (144)	KEMA aligned features (50)
LiDAR	LiDAR band + opening and closing by reconstruction features with convolution of size $[7, 19, 31]$ pixels (7)	
AVG	Average filters, window size 3, applied on the:	
	10 first principal component projections (10)	10 first KEMA projections (10)

Table 2: Classification results (Overall accuracy, in %) for the Houston data.

Entire test set

HSI processing:	Raw	HM	SSMA	KEMA (us)
HSI	71.0 ± 0.1	79.5 ± 0.4	81.4 ± 1.7	83.8 ± 1.9
↪ + LiDAR	83.4 ± 0.2	86.4 ± 0.7	89.3 ± 0.6	89.4 ± 1.4
↪ + AVG	85.1 ± 0.2	84.5 ± 0.4	92.3 ± 0.8	93.0 ± 0.8
↪ + MV	85.5 ± 0.2	86.0 ± 0.3	93.8 ± 0.8	94.3 ± 0.8

Shadowed areas in the test set

HSI	04.2 ± 0.1	67.4 ± 0.7	58.0 ± 9.4	70.0 ± 1.0
↪ + LiDAR	22.5 ± 0.3	77.1 ± 1.3	78.8 ± 2.9	82.6 ± 5.4
↪ + AVG	23.2 ± 1.2	73.6 ± 0.8	89.0 ± 4.1	90.4 ± 4.9
↪ + MV	23.8 ± 1.2	75.1 ± 0.9	90.6 ± 3.7	91.5 ± 4.5