ISPRS Journal of Photogrammetry and Remote Sensing 120 (2016) 1–12

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

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[1]
[b] Image Processing Laboratory, Universitat de València, Spain[2]

ARTICLE INFO

Article history:
Received 12 April 2016
Received in revised form 5 July 2016
Accepted 7 July 2016

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

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 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.

1. Introduction

Many real-life problems currently exploit heterogeneous sources of remote sensing data: forest ecosystems studies (Asner et al., 2005; Asner et al., 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., 2016) take advantage of the wide coverage and short revisit time of remote sensing sensors. In these works, specific image processing pipelines are designed 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 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

* Corresponding author. Fax: +41 44 635 68 46.
E-mail address: devis.tuia@geo.uzh.ch (D. Tuia).
[1] http://www.geo.uzh.ch/en/units/multimodal-remote-sensing.
[2] http://isp.uv.es.

http://dx.doi.org/10.1016/j.isprsjprs.2016.07.004
0924-2716/

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., 2016): (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., 2016b; Sun et al., 2016a); (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 of unifying 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 issues 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 color visualization for hyperspectral data was tackled in Liao et al. (2016).

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.

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 $(l_1, \ldots, l_M)$, and most are unlabeled. From one domain to another, the data are not necessarily 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 finds a data projection matrix $\mathbf{F}$ that minimizes the following cost function

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

where we aim to minimize a topology/geometry (GEO) and a class similarity (SIM) terms while maximizing 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. minimizing 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,

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

where W_g^m is a similarity matrix returning the value 1 if two pixels of domain m are neighbors 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 neighborhood relationships between the images.