CLOUD-SCREENING FROM MULTISPECTRAL SATELLITE IMAGE TIME SERIES

Luis Gómez-Chova, Julia Amorós-López, Emma Izquierdo-Verdiguier, Juan Carlos Jiménez-Muñoz, and Gustavo Camps-Valls

University of Valencia, Image Processing Laboratory, Valencia, Spain; chovago@uv.es

ABSTRACT

This paper faces the challenging problem of cloud screening in multispectral image time series acquired by space-borne sensors working in the visible and near-infrared range of the electromagnetic spectrum. The main objective of this paper is to provide new operational tools for masking clouds in time series from Earth observation satellites. In particular, we propose a novel multitemporal method for change detection that maximizes specific changes between two dates while minimizes sources of errors, such as land-cover changes and co-registration. The feature extraction method is based on kernels and can deal with non-linear relations between samples at different dates. The effectiveness of the proposed method is successfully illustrated in a cloud screening application using a Landsat time series acquired over Albacete, Spain, in 2009. Results show that the proposed method provides the most discriminative features in terms of cloud detection when confronted with standard PCA, IRMAD, and advanced nonlinear KPCA. In particular, extracted features with the proposed method enable automatic cloud detection in multispectral time series.

INTRODUCTION

Time series from space-borne sensors working in the visible and near-infrared range of the electromagnetic spectrum are an extremely valuable data source. However, processing and analysis of satellite image time series poses some challenges that can seriously hamper their operational use. The first basic but important requirement is that all the images that form the series should be perfectly co-registered, as any spatial mismatch between images would produce erroneous results. The other factor that has a critical impact on the results is the presence of clouds in the acquired scenes. For instance, Landsat scenes are globally estimated to be 35% cloud covered [1]. Therefore, an automatic and accurate cloud screening is essential in order to improve the temporal consistency and the usability of the time series by considering partially cloudy images.

In this context, cloud screening of image time series can be takled as a specific change detection problem. Most unsupervised change-detection approaches are based on the difference image between two single dates, which is further processed to obtain the changes [2]. Basically, information from the diference image is transformed to extract new features where changes can be easily obtained by simple thresholding or clustering techniques [3, 4]. Linear feature extractors have been extensively used for remote sensing change detection. Among these methods, the most common is principal component analysis (PCA), which is applied to maximize the variance of the differences. In order to mitigate problems due to radiometric differences and noise, canonical correlation analysis (CCA/MAD) plus minimum noise fraction (MNF/MAF) transforms are becoming increasingly popular in unsupervised change detection [5]. These methods, however, can only deal with affine transforms

*This work has been supported by the Universitat de València under project UV-INV-AE11-41223.

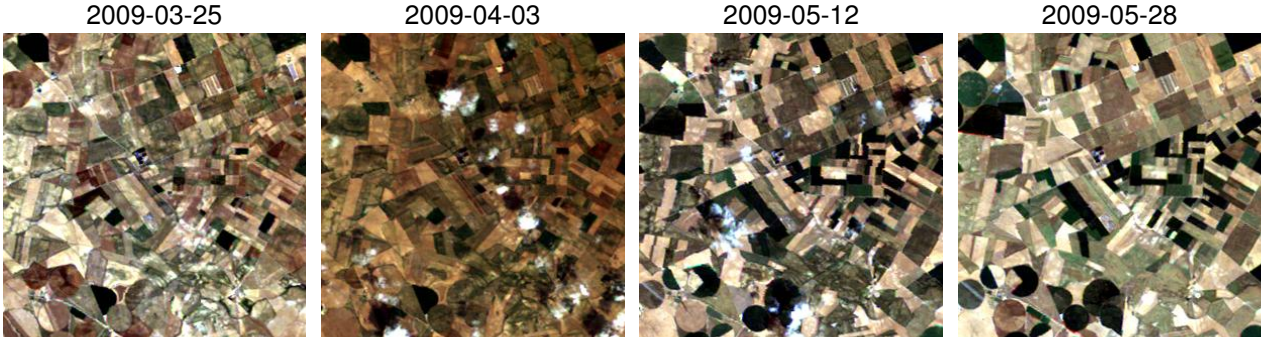

Figure 1: Example of Landsat TM images at four dates. Information about the presence of clouds in the images is available and it is used to select the anterior (X_1) and posterior (X_3) cloud-free images to each cloudy image (X_2) to be analyzed.

and additive uncorrelated Gaussian noise, and cannot be easily adapted to find a particular type of change, such as clouds.

In this paper, we propose a nonlinear feature extraction method that can be easily adapted to find a specific change of interest in satellite image time series. In recent years, kernel methods have emerged as an excellent tool to develop nonlinear feature extraction methods [6] and have also been applied to unsupervised change detection [7]. The proposed method generates nonlinear data projections that enhance changes between acquisitions due to cloud covers compared with surface changes between cloud-free images (see Fig. 1).

KERNEL CHANGE DISCRIMINANT ANALYSIS

This section presents a novel kernel method for the nonlinear extraction of multitemporal features that maximizes specific changes between two dates while minimizes sources of errors such as land-cover changes and co-registration errors. We propose a formulation in which the difference between samples from different dates is computed and then a transform that maximizes the signal covariance of the changes of interest while minimizes the covariance of the changes of no-interest is found. First, a brief introduction of standard linear and kernel feature extraction methods is given. Then the formulation for the proposed method, which we denominate Kernel Change Discriminant Analysis (KCDA), is presented.

Notationally, we are given a set of training feature vectors $\{\mathbf{x}_i^t\}_{i=1}^n \in \mathbb{R}^d$ at different time instants, t, in an input space of dimension d, i.e. n samples with d spectral channels or bands. This can be also expressed using matrix notation, $\mathbf{X}_t = [\mathbf{x}_1^t, \ldots, \mathbf{x}_n^t]^\top$, where $^\top$ denotes matrix transposition. In a change detection setup, the difference image between two time instants $t = \{1, 2\}$ is computed as $\tilde{\mathbf{X}} = \tilde{\mathbf{X}}_2 - \tilde{\mathbf{X}}_1$, where $\tilde{\mathbf{X}}$ indicates the centered version of $\mathbf{X}$, and $\mathbf{C} = \frac{1}{n}\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}$ represents the empirical covariance matrix of the difference data. In this context, linear feature extraction can be carried out by projecting the data into the subspace characterized by the projection matrix $\mathbf{U}$, of size $d \times n_p$, so that the $n_p \leq d$ extracted features of the original data are given by $\tilde{\mathbf{X}}' = \tilde{\mathbf{X}}\mathbf{U}$.

Principal component analysis (PCA) projects linearly the data onto the directions of largest variance [8]. Therefore, to perform PCA, one has to solve:

$$\text{PCA:} \quad \mathbf{U} = \arg\max_{\mathbf{U}} \{\text{Tr}(\mathbf{U}^\top\mathbf{C}\mathbf{U})\} \quad \text{subject to} \quad \mathbf{U}^\top\mathbf{U} = \mathbf{I}, \tag{1}$$

where $\mathbf{I}$ is the identity matrix of size $n_p \times n_p$. This can also be expressed, using Lagrange multipliers, as the eigenvalue problem $\mathbf{C}\mathbf{u}_i = \lambda_i\mathbf{u}_i$ (or singular value decomposition of $\mathbf{C}$).

The main limitations of PCA for change detection are 1) that the same transform is applied to both $\mathbf{X}_1$ and $\mathbf{X}_2$, which suggested the use of canonical correlation in MAD and IRMAD methods [5]; and 2) it does not consider the noise present in the input vectors, which suggested a postprocessing minimizing the noise (MAF) [5].

All these linear methods simply perform a coordinate rotation that aligns the transformed axes with the directions maximizing a given criterion: in this case, the variance of the difference data distribution. Therefore, they assume that the best extracted features, $\tilde{\mathbf{X}}'$, detecting the changes have a linear relation with the original data matrix, $\tilde{\mathbf{X}}$. However, in many situations this linearity assumption is not satisfied, and nonlinear feature extraction is needed to obtain acceptable performance. In this context, *kernel methods* are a promising approach, as they constitute an excellent framework to formulate nonlinear versions from linear algorithms [6]. In addition, previous methods are completely unsupervised and look for the largest changes, so there is no guarantee that the directions of maximum variance will find a specific change of interest. In the rest of the section, we describe the proposed Kernel Change Discriminant Analysis (KCDA) formulation.

Notationally, consider a nonlinear function $\phi(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathcal{H}$ that maps the input data into some *kernel space* of very high or even infinite dimension. The data matrix for performing the linear feature extraction in $\mathcal{H}$ is now given by $\mathbf{\Phi} = [\phi(\mathbf{x}_1), ..., \phi(\mathbf{x}_n)]^\top$. As before, the centered versions of this matrix is denoted by $\tilde{\mathbf{\Phi}}$. The projection of the input data will be given by $\tilde{\mathbf{\Phi}}' = \tilde{\mathbf{\Phi}}\mathbf{U}$, where the projection matrix $\mathbf{U}$ is now of size $\dim(\mathcal{H}) \times n_p$. Note, that the covariance matrix in $\mathcal{H}$, which is usually needed by the different methods, becomes of size $\dim(\mathcal{H}) \times \dim(\mathcal{H})$ and cannot be directly computed. However, making use of the *representer's theorem*, we can express the projection matrix as a linear combination of the training samples $\mathbf{U} = \tilde{\mathbf{\Phi}}^\top \mathbf{A}$, where $\mathbf{A} = [\alpha_1, ..., \alpha_{n_p}]$ and α_i is an n-length column vector containing the coefficients for the ith projection vector. Now the maximization problem can be reformulated solely in terms of the dot product of mapped samples, which is defined by the *kernel function* $\tilde{K}(\mathbf{x}_i, \mathbf{x}_j) = \tilde{\phi}(\mathbf{x}_i)^\top \tilde{\phi}(\mathbf{x}_j)$. Note that, in these kernel feature extraction methods, the projection matrix $\mathbf{U}$ in $\mathcal{H}$ might not be explicitly calculated, but the projections for an input pattern $\mathbf{x}_*$ are given by:

$$\tilde{\phi}'(\mathbf{x}_*)^\top = \tilde{\phi}(\mathbf{x}_*)^\top \mathbf{U} = \tilde{\phi}(\mathbf{x}_*)^\top \tilde{\mathbf{\Phi}}^\top \mathbf{A} = [\tilde{K}(\mathbf{x}_*, \mathbf{x}_1), ..., \tilde{K}(\mathbf{x}_*, \mathbf{x}_n)]\mathbf{A}, \tag{2}$$

which is expressed in terms of the inner products in the centered feature space that, as in all kernel methods, can be computed via a positive semidefinite Mercer kernel function.

As mentioned before, previous approaches consider a bi-temporal change detection problem between $\mathbf{X}_1$ and $\mathbf{X}_2$, and they cannot focus on a particular type of change such as cloud detection. The aim of the proposed KCDA is to find projections that maximize the variance of differences due to clouds, which will be observed between the image $\mathbf{X}_1$ (*cloud-free*) and $\mathbf{X}_2$ (*cloudy*), and also minimize any type of land-cover change or co-registration errors, which might be observed between the *cloud-free* images $\mathbf{X}_1$ and $\mathbf{X}_3$. Therefore, the proposed KCDA want to find projections maximizing the ratio between the variance of the mapped differences from 1 to 2, $\tilde{\mathbf{\Phi}}_{21} = [\phi(\mathbf{x}_1^2 - \mathbf{x}_1^1), ..., \phi(\mathbf{x}_n^2 - \mathbf{x}_n^1)]^\top$, and the variance of the mapped differences from 1 to 3, $\tilde{\mathbf{\Phi}}_{31} = [\phi(\mathbf{x}_1^3 - \mathbf{x}_1^1), ..., \phi(\mathbf{x}_n^3 - \mathbf{x}_n^1)]^\top$, in $\mathcal{H}$.

$$\text{KCDA:} \quad \mathbf{U} = \arg\max_{\mathbf{U}} \left\{ \text{Tr}\left(\frac{\mathbf{U}^\top \tilde{\mathbf{\Phi}}_{21}^\top \tilde{\mathbf{\Phi}}_{21} \mathbf{U}}{\mathbf{U}^\top \tilde{\mathbf{\Phi}}_{31}^\top \tilde{\mathbf{\Phi}}_{31} \mathbf{U}} \right) \right\} \quad \text{subject to } \mathbf{U}^\top \tilde{\mathbf{\Phi}}_{31}^\top \tilde{\mathbf{\Phi}}_{31} \mathbf{U} = \mathbf{I}. \tag{3}$$

This is equivalent to solve the generalized eigenproblem $\tilde{\mathbf{\Phi}}_{21}^\top \tilde{\mathbf{\Phi}}_{21} \mathbf{u}_i = \lambda_i \tilde{\mathbf{\Phi}}_{31}^\top \tilde{\mathbf{\Phi}}_{31} \mathbf{u}_i$ that can be also expressed using matrix notation $\tilde{\mathbf{\Phi}}_{21}^\top \tilde{\mathbf{\Phi}}_{21} \mathbf{U} = \tilde{\mathbf{\Phi}}_{31}^\top \tilde{\mathbf{\Phi}}_{31} \mathbf{U}\mathbf{\Lambda}$, which is not directly solvable since matrices $\tilde{\mathbf{\Phi}}^\top \tilde{\mathbf{\Phi}}$ and projection matrix $\mathbf{U}$ have possibly infinite dimension, $\dim(\mathcal{H}) \times \dim(\mathcal{H})$.

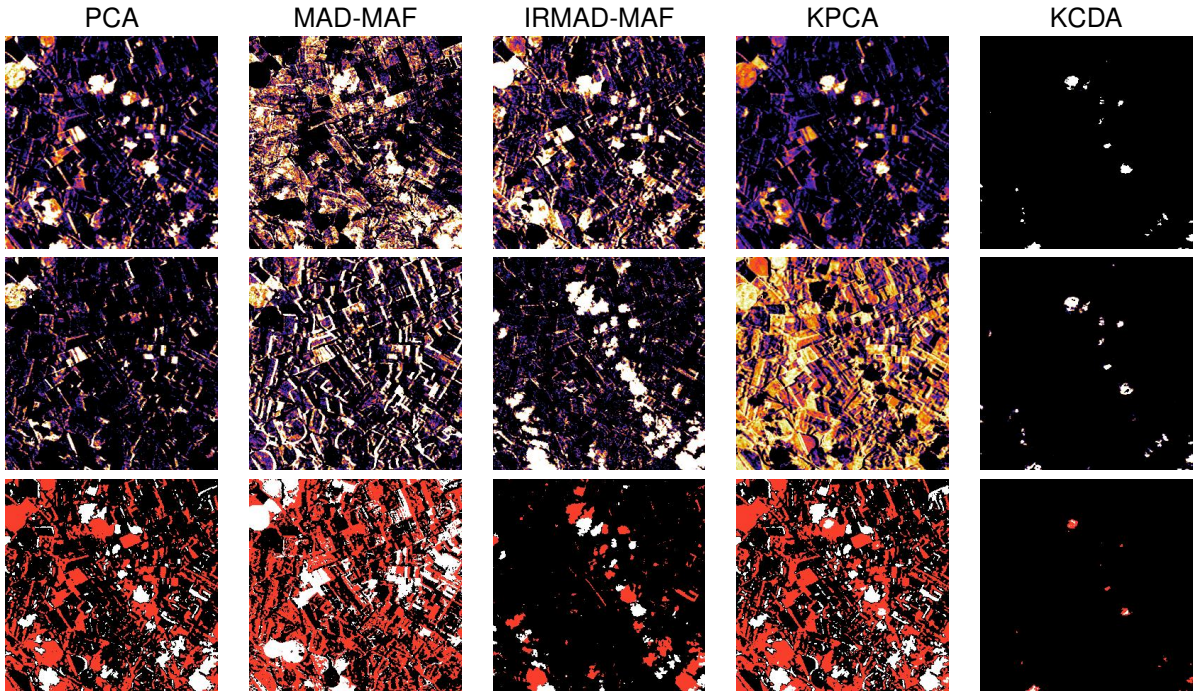

Figure 2: Extracted features (first and second rows) and clustering (third row) for the 2009-04-03 Landsat image (see Fig. 1) using (from left to right) PCA, MAD-MAF, IRMAD-MAF, KPCA, and the proposed kernel method.

Making use of the representer's theorem, $\mathbf{U} = \tilde{\mathbf{\Phi}}_{21}^{\top}\mathbf{A}$, and pre-multiplying both sides by $\tilde{\mathbf{\Phi}}_{21}$ it is possible to express the same problem in terms of dot products:

$$\tilde{\mathbf{\Phi}}_{21}\tilde{\mathbf{\Phi}}_{21}^{\top}\tilde{\mathbf{\Phi}}_{21}\tilde{\mathbf{\Phi}}_{21}^{\top}\mathbf{A} = \tilde{\mathbf{\Phi}}_{21}\tilde{\mathbf{\Phi}}_{31}^{\top}\tilde{\mathbf{\Phi}}_{31}\tilde{\mathbf{\Phi}}_{21}^{\top}\mathbf{A}\mathbf{\Lambda} \rightarrow \tilde{\mathbf{K}}_{22}\tilde{\mathbf{K}}_{22}\mathbf{A} = \tilde{\mathbf{K}}_{23}\tilde{\mathbf{K}}_{23}^{\top}\mathbf{A}\mathbf{\Lambda}, \tag{4}$$

where we have defined the symmetric centered kernel matrix $\tilde{\mathbf{K}}_{22} = \tilde{\mathbf{\Phi}}_{21}\tilde{\mathbf{\Phi}}_{21}^{\top}$ containing the inner products between any two difference vectors from 1 to 2 in the kernel space, and the non-symmetric kernel $\tilde{\mathbf{K}}_{23} = \tilde{\mathbf{\Phi}}_{21}\tilde{\mathbf{\Phi}}_{31}^{\top} = \tilde{\mathbf{K}}_{32}^{\top}$ containing the inner products between the differences from 1 to 2 and from 1 to 3 in the kernel space. Then, projections are obtained as $\tilde{\mathbf{\Phi}}' = \tilde{\mathbf{\Phi}}_{21}\mathbf{U} = \tilde{\mathbf{\Phi}}_{21}\tilde{\mathbf{\Phi}}_{21}^{\top}\mathbf{A} = \tilde{\mathbf{K}}_{22}\mathbf{A}$.

Finally, a simple but effective technique to reinforce the enhancement of the changes of interest and the rejection of the changes of no-interest in the extracted features consist in including in the training set: 1) samples whose changes between image 1 and 2 imply an increase in intensity (probably due to clouds), and also 2) samples that present any significant change between image 1 and 3 (any other changes) or belong to the edges of the land-cover regions (where changes are probably due to co-registration errors).

EXPERIMENTAL RESULTS

A time series of Landsat TM images[1] acquired over Albacete, Spain, in 2009 is used to illustrate the capabilities of the proposed multitemporal feature extraction algorithm. In this work, cloud-free images were manually selected from the time series by visual inspection and the Landsat TM images were geometrically and atmospherically corrected.

The selected study area (10×10 km) is located in the overlap area between adjacent Landsat orbits in order to get an additional acquisition between two nominal Landsat acquisitions (revisit time of

[1]Authors want to thank the Instituto Geográfico Nacional (IGN) for granting access to the Landsat data.

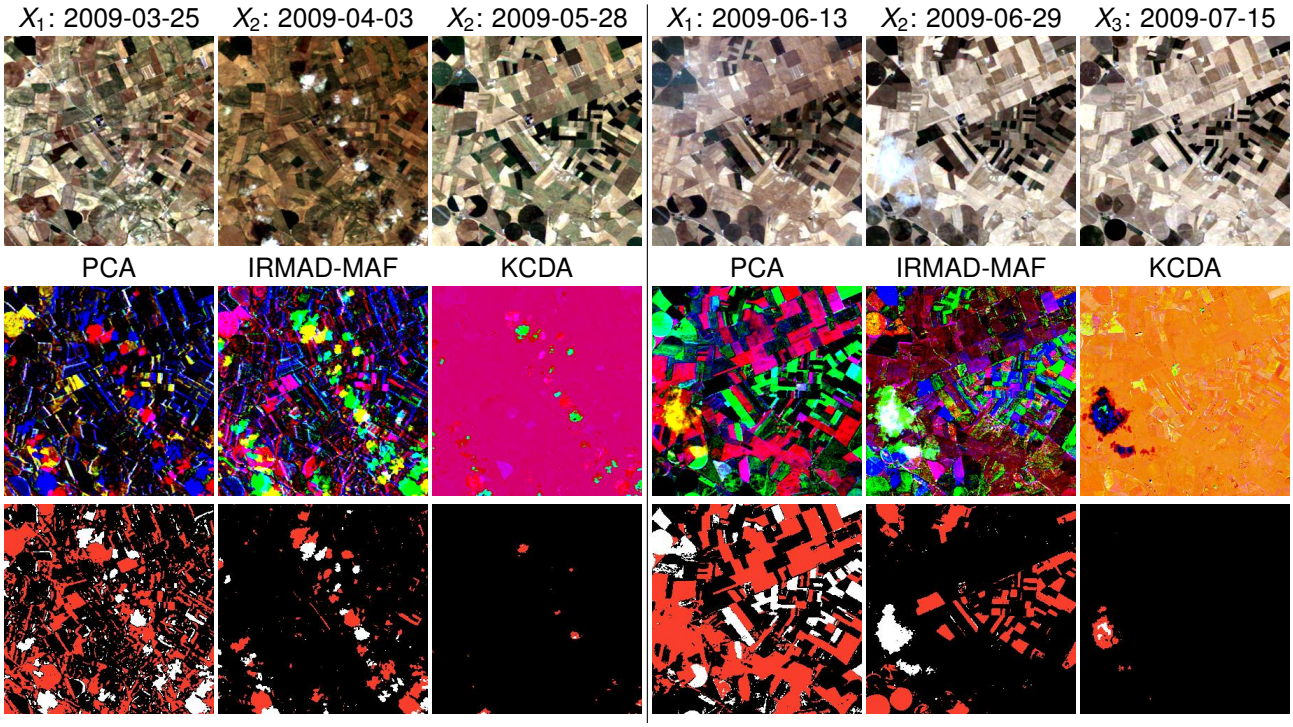

Figure 3: Results for two cloudy images in the Landsat time series using PCA, IRMAD-MAF, and KCDA. In each panel, first row shows the X_1, X_2, and X_3 Landsat images; second row shows a false color composite of the three most relevant extracted features with each method; and third row shows the clusters obtained using k-means with $k = 3$ and the Euclidean distance over the three most relevant extracted features.

16 days). These additional images are very useful to better assess changes. The region covers an irrigated agricultural area (Barrax site), which is characterized by large and uniform land-use units. The main cultivated crops in the region are: cereals (barley, wheat, corn, and oat), potatoes, sugar beet, sunflower, alfalfa, and vegetables (onion, garlic, melon, and tomatoes). Hence, a lot of land-cover changes can be observed during the growing season.

First, we compare the proposed algorithm against PCA, MAD-MAF, IRMAD-MAF, and KPCA. In all the experiments, projection matrices are computed using a training set of 1000 samples selected from the images, the RBF kernel is used for all kernel methods, and the kernel width is fixed to the average distance of the training samples, which is a standard and reasonable procedure in unsupervised kernel learning methods [6].

Figure 2 shows in its first two rows the two most relevant extracted features, in terms of their corresponding eigenvalues, for each method. Visual inspection of extracted features shows that the other methods focus on all changes between images. While the first component highlights clouds in all the methods, land-cover changes and linear features due to co-registration errors. On the other hand, the proposed features clearly highlight clouds from the background thus improving the discrimination of the changes of interest respect anything else.

To assess the discrimination power of the extracted features, an automatic detection of the changes is performed by an unsupervised image classification of the features using the k-means clustering. Usually, the number of clusters is fixed to $k = 2$ in order to automatically detect two classes: *'change'* and *'no-change'*. However, with only two clusters the other methods mix all changes in the same cluster and only the proposed method isolate clouds in the *'change'* cluster (results not shown). Therefore, we increase the number of clusters to $k = 3$ in order to separate clouds and other changes

in different clusters. In the third row of Fig. 2, one can see the classified changes for all the methods. Among the other methods included in the comparison, IRMAD-MAF provides the best results but mixes shadows and co-registration errors in the *'change'* clusters. The proposed KCDA method outperforms all the other methods in finding automatically the clouds present in the scene.

Finally, in order to validate the method in more images, Fig. 3 shows the three most relevant extracted features (false color) and the automatic classification of changes for two dates. Again, one can observe that PCA's components highlight clouds but they are still not well detected. IRMAD-MAF detects clouds but also shadows and borders due to co-registration errors in the first date; and mixes clouds with land-cover changes in the second one. The proposed KCDA approach provides more spatially homogeneous detection maps than the other methods and, although some cloud borders are missed, it detects all clouds present in the scene.

CONCLUSIONS

This paper presented a novel multitemporal method for change detection that maximizes specific changes between two dates while minimizes sources of errors, such as land-cover changes and co-registration. The feature extraction method is based on kernels and can deal with non-linear relations between samples at different dates. In addition, it presents good theoretical and practical properties since it represents a simple optimization problem solved through the generalized eigendecomposition of the kernel matrices, and the only parameter is the width of the employed RBF kernel.

Further work considers extending the application of the method to more than three images and also to apply it for bi-temporal change detection when only two images are available. We also plan to learn optimal kernel parameters in an automatic way. Finally, it would be interesting to test KCDA in more remote sensing applications, such as flooded or burned area detection and, in general, in any specific change event of interest.

REFERENCES

[1] J. Ju and D.P. Roy, "The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally," *Remote Sensing of Environment*, vol. 112, no. 3, pp. 1196–1211, 2008.

[2] L. Bruzzone and D.F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Trans. on Geos. and Remote Sensing*, vol. 38, no. 3, pp. 1171–1182, may 2000.

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

[4] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and k-means clustering," *IEEE Geos. and Remote Sensing Letters*, vol. 6, no. 4, pp. 772–776, oct 2009.

[5] A.A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data," *Image Processing, IEEE Transactions on*, vol. 16, no. 2, pp. 463–478, feb. 2007.

[6] G. Camps-Valls and L. Bruzzone, Eds., *Kernel methods for Remote Sensing Data Analysis*, Wiley & Sons, UK, dec 2009.

[7] M. Volpi, D. Tuia, G. Camps-Valls, and M. Kanevski, "Unsupervised change detection with kernels," *IEEE Geos. and Remote Sensing Letters*, p. In press, 2012.

[8] I. T. Jolliffe, *Principal Component Analysis*, Springer, 1986.