A L2-Norm Regularized Pseudo-Code for Change Analysis in Satellite Image Time Series

Anamaria Radoi[1,2] and Mihai Datcu[3]

[1] Department of Applied Electronics, University Politehnica of Bucharest, Romania
[2] Research Center for Spatial Information, CEOSpaceTech, Romania
rdi.anamaria@gmail.com,
[3] Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany
mihai.datcu@dlr.de

Abstract. *The continuous progress in the acquisition of high-dimensional information (e.g., satellite image time series, or medical screening) implies an efficient characterization of changes that occur in a temporal series of data. A pseudo-encoding technique can be designed to represent the changes between two consecutive moments of time, based on the minimization of a convex error function which has an analytical solution. The domain transformed feature vectors are grouped into clusters using K-Means. The proposed approach results in a better separation between classes and, thus, in an enhanced characterization of temporal changes. The experiments are done on 5 Landsat multispectral images at 30 meters spatial resolution, covering an area of approximately 59 × 51 km^2 around Bucharest, Romania.*

Keywords: Domain adaptation, data mining, information retrieval, multitemporal images, satellite image time series

1 Introduction

Lately, with the increasing demand in Earth surveillance, the interest developed in the automatic analysis of change has constantly accrued. This work is centered around satellite image time series, for which most of the methods that analyze temporal series of data are mainly looking at the change detection aspect, neglecting the information that can be obtained from a more complex analysis.

Intuitively, features belonging to different classes are distributed differently across datasets. For these reasons, classifiers behave in a different manner for each of these distributions. Information-theoretical learning methods [3] and kernels designed to have specific properties [2] were used to correct the distribution differences. Similarly, change analysis has to deal with problems of revealing the changes in all types of classes, not only the dominant ones.

Moreover, one class of unsupervised domain adaptation methods, related to the proposed approach, is to change the feature representations such that the shared characteristics between the source and target domains are kept. In this

case, we aim to keep the information regarding the class provenience in the analysis applied to changes that occur between two temporal satellite images. One way to do this is to derive new feature representations that are able to model the particular characteristics of the classes.

This paper addresses the above mentioned issues from the perspective of minimizing a convex cost function, which results in a pseudo-encoder that quantifies the dissimilarities between the feature maps of two consecutive images, whereas the multispectral information is included in the descriptor of each pixel, or patch.

The rest of the paper is organized as follows. Section 2 presents some of the currently used methods for change detection, whereas Section 3 introduces the basic idea of the proposed pseudo-encoding based on a functional minimization. Section 4 comes with a set of experimental results, whilst Section 5 concludes the paper.

2 Related Work

The traditional change detection techniques in satellite image time series (SITS) are divided into several groups [1]: algebra based approach (i.e., image differencing, image rationing), linear transformations (i.e., Principal Component Analysis, Tasseled Cap), classification based methods (i.e., unsupervised change detection, artificial neural networks). These techniques can be also combined in order to yield better results in terms of change detection.

An automatic change detection method is depicted in [7] by finding the best threshold between "change" and "no change" through an Expectation-Maximization algorithm applied to the difference between two images. To smooth the detection and to get benefit from the interpixel dependencies, a Markov Random Field is used. Another approach to change detection is [6], which projects each difference pixels on the first principal components of the difference images.

Change map time series are tackled by [9] using a Latent Dirichlet Allocation (LDA) model to describe the dynamic evolution of the Earth's surface. The change detection process comprises four similarity measures, namely: correlation coefficient, Kullback-Leibler divergence, conditional information, and normalized compression distance.

Algebraic techniques are widely used in the change detection chain, even in specialized tools, due to their simplicity and low complexity. They are based on the differences (i.e., subtractions, ratios, or log-ratios depending on the type of satellite image) of the pixel values situated at the same location, or differences of the linear transformation results. Post-classification methods aim at providing an overview of the types of changes that occur in the temporal series, by classifying the resulted feature vectors into classes of change.

Image Differencing. This algebra-based method consists of the pixel-wise difference between the satellite images registered at different times, over the

same space location. More precisely, we compute, for each subband of frequency:

$$\mathbf{DIFF}^{(t)} = \mathbf{I}^{(t)} - \mathbf{I}^{(t-1)}, \tag{1}$$

where $\mathbf{I}^{(t)}$ is the matrix of pixel values at time t, i.e. the satellite image captured at time t, and $\mathbf{DIFF}^{(t)}$ represents the pixel-wise difference between two images registered at two consecutive moments of time. No temporal changes in a location is translated in a 0 entry in the corresponding position of the image differencing matrix. Due to residual differences (i.e., not caused by the temporal changes), the 0's are, in fact, gray-levels, for which a threshold has to be found in order to demarcate between the change and no change states.

Image Rationing. In a similar manner, image rationing is defined as:

$$\mathbf{R}^{(t)} = \frac{\mathbf{I}^{(t)}}{\mathbf{I}^{(t-1)}}, \tag{2}$$

where $\mathbf{R}^{(t)}$ represents the ratio between pixel values of both images, $\mathbf{I}^{(t-1)}$ and $\mathbf{I}^{(t)}$, and the division is done point-wise. As before, the images are registered at consecutive moments of time. In the ideal case, no temporal changes imply a ratio of 1, whereas changes are represented by ratios higher or lower than 1.

3 Proposed Method for Change Analysis

3.1 Overview of the Proposed Method

Let us denote by $\mathbf{I}^{(t-1)}$ and $\mathbf{I}^{(t)}$ two temporal images, rescaled between $[0, 1]$. The corresponding descriptors are $\mathbf{D}^{(t-1)}$ and $\mathbf{D}^{(t)}$, that can be taken at a pixel-level, or at a patch-level, as we will see in the next section. The change matrix, $\mathbf{C}_{\lambda}^{(t)}$, quantifies the dissimilarity between the two temporal images, using different algebraic measures such as: image differencing, image rationing, or the proposed pseudo-encoder described in this section. Post-classification is done using a simple and fast K-means, where K is the number of classes of change, and the classifier's inputs are exactly the resulted pseudo-codes. The idea behind the post-classification is to show that each change belongs to a particular class, which is strongly correlated to the classes perceived by a user.

Fig. 1 summarizes the change analysis chain that was described above. As mentioned before, throughout this paper, we bring into discussion change analysis as opposed to change detection, the former being less deliberated in most of the related papers.

3.2 Pseudo-Encoder with L_2-Norm Regularization Term

Let us consider a temporal series (e.g., SITS) with T observations (e.g., satellite images), $\{\mathbf{I}^{(t)} \in \mathbb{R}^{S_1 \times S_2}\}_{t \in \{1,\ldots,T\}}$, where S_1 and S_2 are the dimensions of the images. Each image is divided in N patches of size $p \times p$ pixels. At limit, the patch can be considered of 1 pixel (i.e., $p = 1$). Additionally, let $\mathbf{D}^{(t)} \in \mathbb{R}^{d \times N}$ contain d-dimensional local descriptors (i.e., patch features, pixel values) extracted from the image $\mathbf{I}^{(t)}$ captured at time t.

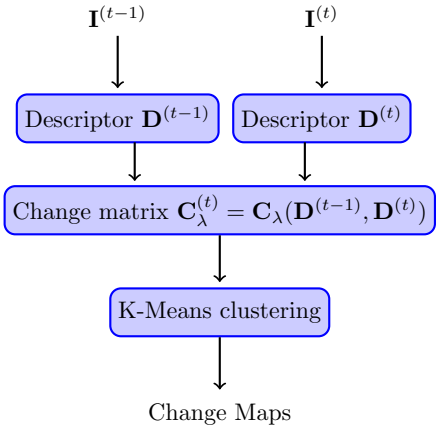


Fig. 1. Block diagram of the proposed change analysis.

Problem Formulation. The main rationale behind the proposed method is to represent the changes in a self-contained manner so that the types of changes are apart one from another, whilst the algorithm can be generalized to all kinds of feature maps. In order to attain this aim, the changes that occur between two consecutive registrations, $\mathbf{I}^{(t-1)}$ and $\mathbf{I}^{(t)}$, are "encoded" by minimizing the following convex cost function:

$$J(\mathbf{C}_\lambda^{(t)}) = \sum_{i=1}^{N} \left(\|\mathbf{D}_i^{(t)} - \mathbf{C}_{\lambda,i}^{(t)} \odot \mathbf{D}_i^{(t-1)}\|_2^2 + \lambda \cdot \|\mathbf{d}_i \odot \mathbf{C}_{\lambda,i}^{(t)}\|_2^2 \right), \tag{3}$$

where $\mathbf{C}_\lambda^{(t)} = \left[\mathbf{C}_{\lambda,1}^{(t)}, \mathbf{C}_{\lambda,2}^{(t)}, \ldots, \mathbf{C}_{\lambda,N}^{(t)}\right] \in \mathbb{R}^{d \times N}$ represents the set of learned codes for each of the N patches inside the image, $\lambda \in \mathbb{R}^+$ is a sparsity coefficient, $\|\cdot\|_2$ is the L_2-norm, and $\odot$ denotes the point-wise multiplication $[A \odot B]_i = A_i B_i$. The term $\mathbf{d}_i \in \mathbb{R}^{D \times 1}$ acts like a weight for the sparsity coefficient λ on each dimension of patch $i \in \{1, \ldots, N\}$. Specifically,

$$\mathbf{d}_i = \left(1 + \exp\left(-\text{dist}(\mathbf{D}_i^{(t-1)}, \mathbf{D}_i^{(t)})\right)\right)^{-1}, \tag{4}$$

where $\text{dist}(\mathbf{a}, \mathbf{b}) = [|a_1 - b_1|, \ldots, |a_D - b_D|]^T$ measures the distance between each element inside the feature map of patch i captured at time $t-1$ and its correspondent captured at time t.

The choice of the cost function is argued by multiple facts. Firstly, $J(\mathbf{C}_\lambda^{(t)})$ is able to induce a measure of similarity that can be further used to cluster the temporal evolution between pairs of patches. Secondly, this optimization task is convex and removes the risk of falling into local minimums as it is the case of non-convex functions. Thirdly, a linear transform cannot adapt to differences between classes. Hence, non-linear transforms are required to reduce this information loss.

Furthermore, imposing the sparsity constraint in equation (3) leads to a better separation between codes, and, thus, to a better separation between clusters, as we will see in Section 4. The sparseness is not induced using a common L_0-norm, or L_1-norm, but by using the L_2-norm regularization term similar to the one proposed in [11]. This imposes the presence of few significant values in the code, and also an analytical global solution for the minimization problem, avoiding the computational burden carried by other sparse coding methods.

The weights $\{\mathbf{d}_i\}_{i=1,\dots,N}$ give different freedom on each dimension of the code $\mathbf{C}_{\lambda,i}^{(t)}$ depending on the distance between the elements of the feature maps. In addition, the function $f(x,y) = \frac{1}{1+\exp(-|x-y|)}$ in expression (4) has the following useful properties:

1. $|x-y|$ is not sensitive to the direction of change (i.e., the distance is the same between feature vector at time $t-1$ and those at time t, and vice-versa);
2. If $|x-y|$ is large enough (i.e., a major change took place in the respective area), $f(x,y) \nearrow 1$;
3. If $\liminf_{|x-y|\to 0} f(x,y) = 0.5$ (i.e., no major changes).

Solution to the minimization problem. The minimization of the convex function $J(\mathbf{C}_{\lambda}^{(t)})$ is equivalent to solving a linear system with $d \times N$ equations and $d \times N$ unknowns, namely $\mathbf{C}_{\lambda}^{(t)} \in \mathbb{R}^{d\times N}$:

$$-\left(\mathbf{D}^{(t)} - \mathbf{C}_{\lambda}^{(t)} \odot \mathbf{D}^{(t-1)}\right) \odot \mathbf{D}^{(t-1)} + \lambda \cdot \mathbf{d} \odot \left(\mathbf{d} \odot \mathbf{C}_{\lambda}^{(t)}\right) = \mathbf{0}.$$

The solution of this system of equations can be derived simply as:

$$\mathbf{C}_{\lambda}^{(t)} = \frac{\mathbf{D}^{(t-1)} \odot \mathbf{D}^{(t)}}{\mathbf{D}^{(t-1)} \odot \mathbf{D}^{(t-1)} + \lambda \cdot \mathbf{d} \odot \mathbf{d}}, \tag{5}$$

where $\mathbf{d} = [\mathbf{d}_1, \dots, \mathbf{d}_N] \in \mathbb{R}^{d\times N}$ and the division is taken element by element.

Advantages. Firstly, as already mentioned, having an analytical solution (5) to the minimization of the error function (3) represents a major advantage.

Secondly, the sparsity (i.e., of course, in the sense of fewer important values) in the codes is beneficial: It helps the clustering algorithm to distinguish better between the classes, as we will see in Section 4, reducing the impact of the noise on the data analysis, a frequent source of errors in satellite imagery (e.g., different atmospheric conditions, different seasons between the registrations).

Furthermore, besides the direct advantages of the encoding procedure (i.e., analytical solution, better separation), we mention also the generality of the algorithm: the method can be applied to any type of feature maps, at a pixel-level, or patch-level, as we will show shortly.

Relation with other algebraic measures for change detection. As a first remark, if no transform is applied to the images and $\lambda = 0$, then the pseudo-encoder is equivalent to an image rationing operation:

$$\mathbf{C}_0^{(t)} = \mathbf{R}^{(t)} \text{ if } \mathbf{D}^{(j)} = \mathbf{I}^{(j)}, \forall j \in \{t-1, t\}, \tag{6}$$

where the pixel level case is considered for $\mathbf{C}_0^{(t)}$. However, image rationing is not well-defined if the values of the pixels are close to 0, which is often the case. For this reason, the term $\lambda \cdot \|\mathbf{d}_i \odot \mathbf{C}_{\lambda,i}^{(t)}\|_2^2$ reduces this risk, being an added advantage of the proposed encoding model.

Moreover, the values that result by image differencing are transformed into weights for the cost function $J(\mathbf{C}_\lambda^{(t)})$, which tries to equalize the importance between the levels of change so that all changes are taken into account.

4 Experimental results

In this section, we report the results obtained by using the pseudo-encoder described in Section 3 on a challenging type of temporal series, namely satellite image time series (SITS). This paper uses 5 Landsat multispectral SITS captured between 2001 and 2003, at 30 meters spatial resolution, covering an area of approximately 59×51 km^2. The satellite images are captured using six subbands of frequency, namely: near-infrared (NIR – the wavelengths are between 0.77 – 0.90 μm) and shortwave infrared (SWIR1 and SWIR2 – the wavelenghts are between 1.55 – 1.75 μm and 2.09 – 2.35 μm).

4.1 Feature maps

The pseudo-encoding method presented in the previous section is a general algorithm that can be used with any type of feature maps, depending on the desired level of detail. Two cases can be depicted, namely: pixel-level, and patch-level descriptors.

Pixel–Level Descriptors At a given time t, for a single spectral subband, the pixel–level descriptor is built considering each pixel value, namely $\mathbf{D}^{(t)} \in \mathbb{R}^{1 \times S_1 S_2}$ is the vectorized form of the image, taken column by column. The multispectral information is included by a simple concatenation of the values of all the spectral subbands (i.e., $\mathbf{D}^{(t)} \in \mathbb{R}^{6 \times S_1 S_2}$ for 6 bands).

Patch–Level Descriptors There are many types of patch based descriptors that describe locally the shapes and textures in an image, starting from wavelet coefficients, edge descriptors, Fourier coefficients, and so on. In this case, we will opt for local image descriptors that result from the projections onto a learned basis, as shown below.

Table 1. The reconstruction error of the SITS decreases with the level of detail.

Patch size	MSE
11	$46.44 \cdot 10^{-3}$
9	$44.07 \cdot 10^{-3}$
7	$42.65 \cdot 10^{-3}$
5	$42.20 \cdot 10^{-3}$

Let us denote by $\mathbf{Y}_i \in \mathbb{R}^{p^2 \times 1}$ the column-wise form of a patch $\mathbf{X}_i \in \mathbb{R}^{p \times p}$ in an image from the database. As described in [10], sparse dictionaries can be learned starting from n randomly selected patches from the dataset, by minimizing the following convex function:

$$J\left(\mathbf{B}, \{\mathbf{t}_i\}_{i=1,\ldots,n}\right) = \sum_{i=1}^{n} \left(\|\mathbf{Y}_i - \mathbf{B} \cdot \mathbf{t}_i\|_2^2 + \mu \cdot \|\mathbf{t}_i\|_1 \right) \tag{7}$$

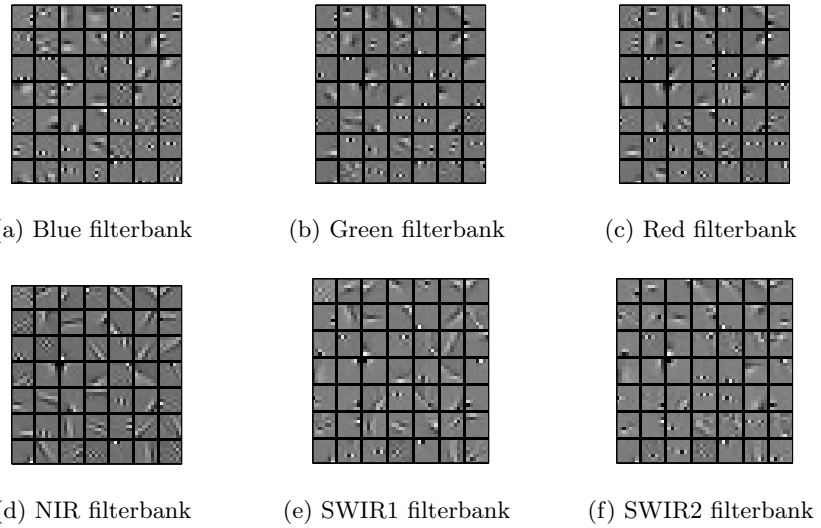
where $\mathbf{B} = [\mathbf{B}_j]_{j=1,\ldots,d}$ has the filters that compose the filterbank on each column, $\mathbf{t}_i$ are d - dimensional vectors that represent the projection of vector $\mathbf{Y}_i$ onto the learned dictionary $\mathbf{B}$, whereas $\|\cdot\|_2$ and $\|\cdot\|_1$ represent the L_2 - norm and L_1 - norm, respectively. μ models the degree of sparsity considered for the representation.

Table 1 shows that the mean squared reconstruction error (MSE) decreases with the size of the patch:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{p^2} \|\mathbf{Y}_i - \mathbf{B} \cdot \mathbf{t}_i\|_2^2. \tag{8}$$

In order to learn specific filterbanks for SITS, we considered $n = 100$ patches of 7×7 pixels (i.e., 210×210 m^2 covered area), $D = p^2 = 49$, and $\mu = 0.5$. The learned filterbanks for each spectral band of the SITS are shown in Fig. 2. The choice $p = 7$ is made in accordance with the analysis of change in SITS, developed in this paper. More precisely, a more detailed level of analysis (i.e., a smaller size of the patch) would lead to inconsistent detection of change, whereas a coarser one would have included different classes in the same patch, which is not desirable.

Furthermore, the corresponding feature vector $\mathbf{t}$ of an arbitrary patch $\mathbf{X}$, with the corresponding vector form $\mathbf{Y}$, can be computed using the approximation $\mathbf{t} = \mathbf{B}^T \cdot \mathbf{Y}$ that incorporates all the projections on the columns of the filterbank $\mathbf{B}$. The feature vectors corresponding to the six subbands are concatenated, leading to a new feature vector, i.e. $[\mathbf{t}_1^T, \ldots, \mathbf{t}_6^T]^T$, where $\mathbf{t}_l$ is the corresponding descriptor that represents the l^{th} subband of the patch in the satellite image. Under these circumstances, the length of each patch's descriptor is $6p^2$.



(a) Blue filterbank (b) Green filterbank (c) Red filterbank

(d) NIR filterbank (e) SWIR1 filterbank (f) SWIR2 filterbank

Fig. 2. Learned filterbanks from SITS

4.2 Clustering the changes in SITS

Further, we perform a post-classification over the resulting pseudo-codes using K-Means, noting that our aim is not a simple separation between "change" and "no change" as in [6] or [8], but a more complex one: The changes are grouped into "types of changes", that mark the transitions between the two temporal moments.

In addition, the analysis shows that the clusters are specific to individual user-defined classes. In order to attain this desideratum, the images are manually labeled, by direct observation, as: *Urban*, *Forest*, *Water*, and *Agriculture*. For example, a change occurring in an urban environment is more likely to be included in an urban prototype of changes, rather than grouped with changes related to water, forest, or agriculture. Moreover, these classes are time-invariant, meaning that a change doesn't imply any kind of modification over these classes. This assumption holds for relatively short periods of time (i.e., one year in our case). If the analysis spans a longer period of time, the pseudo-encoder acts more like a similarity measure, that, in conjunction with the K-Means algorithm, has the added capability of enhancing the separability between types of changes that a human observer cannot easily distinguish.

An example of 10-Means clustering is shown in Table 2, where each cluster is assigned to the class which is most frequent in the cluster. This kind of mapping, for this example in particular, can be observed in Fig. 3. The figure displays only the transition between June 24^{th}, 2001 (T_1) and October 14^{th}, 2001 (T_2), whilst Table 2 shows the results obtained over the 5 SITS.

Table 2. Repartition of each of the 10-Means clusters over the manually labeled classes in percentages [%]

Descriptor	10-Means clusters	**Urban**	**Forest**	**Water**	**Agriculture**
Pixel-level	C_1	0	0	**100**	0
	C_2	1.31	**86.56**	0.27	11.86
	C_3	2.10	0.23	0	**97.67**
	C_4	0.07	0.15	0	**99.78**
	C_5	0.39	2.29	0	**97.32**
	C_6	0.89	0.26	0	**98.85**
	C_7	0.23	7.63	0	**92.14**
	C_8	**98.41**	0.77	0.82	0
	C_9	**100**	0	0	0
	C_{10}	**99.94**	0	0.06	0
Patch-level	C_1	**99.22**	0.58	0	0.19
	C_2	0.77	**86.17**	12.43	0.63
	C_3	**95.78**	0.32	2.00	1.90
	C_4	1.75	2.43	2.99	**92.83**
	C_5	1.03	0	1.45	**97.52**
	C_6	0.34	0	**98.10**	1.56
	C_7	0.67	0.91	11.72	**86.70**
	C_8	0.30	0	0.30	**99.40**
	C_9	6.60	2.83	10.09	**80.48**
	C_{10}	0.80	7.40	5.39	**86.41**

In order to evaluate the results of our proposed method with respect to the performance of clustering in terms of domain separation, we consider *Purity* and *Adjusted Rand Index* as the criteria for analyzing the clustering quality with respect to the number of clusters and sparsity constraint λ. In what follows, let us denote by $\mathcal{S} = \{S_j\}_j$ the manually labeled partition (i.e., user-defined classes) and, by $\mathcal{C} = \{C_k\}_k$, the resulted K-means partition.

Formally, *purity* [5] is defined as:

$$Purity = \frac{1}{N} \sum_{k=1}^{K} \max_{j=1,\ldots,|\mathcal{S}|} |C_k \cap \mathcal{S}_j|, \tag{9}$$

where C_k represents the kth cluster, S_j is the jth user-defined class, and $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$. It measures the agreement between the two partitions in terms of class separation. A complete agreement translates into a 100% purity, whilst independent partitions give a 0% purity.

The second criteria, *Adjusted Rand Index* (or, shortly *ARI*), is one of the most popular performance measures for comparing two partitions of a set by

counting pairs of points that agree/disagree [4]. The index is computed as:

$$\mathrm{ARI}(\mathcal{C},\mathcal{S}) = \frac{\sum_{k,j}\binom{n_{k,j}}{2} - \frac{\sum_k \binom{n_k}{2}\sum_j \binom{n_j}{2}}{\binom{N}{2}}}{\frac{\sum_k \binom{n_k}{2}+\sum_j \binom{n_j}{2}}{2} - \frac{\sum_k \binom{n_k}{2}\sum_j \binom{n_j}{2}}{\binom{N}{2}}}, \tag{10}$$

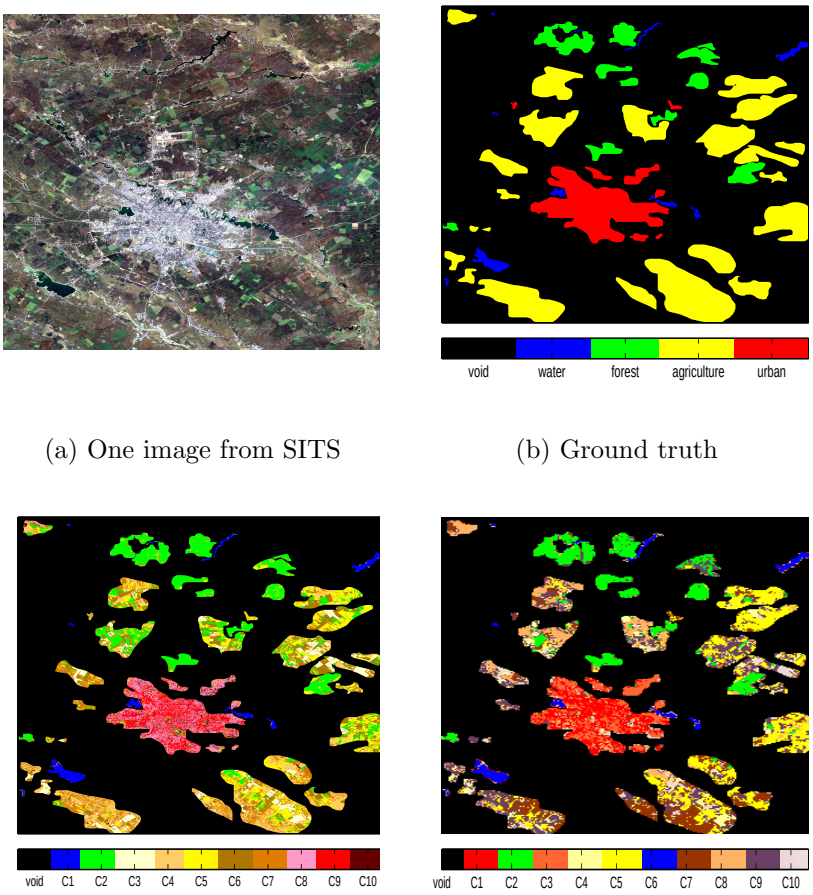
where $n_{k,j} = |C_k \cap \mathcal{S}_j|$, $n_k = \sum_j n_{k,j}$, $n_j = \sum_k n_{k,j}$, and $N = \sum_k \sum_j n_{k,j}$ is the total number of feature points. As proved in [4], ARI is equal to 1 if the partitions agree, whilst being 0 for no agreement.

The performance measures are plotted in Fig. 4 as functions of the number of clusters, for different sparsity constraints λ. We observe that the highest performance in terms of purity is obtained for $\lambda \in (0, 1]$, whereas for $\lambda > 1$, the purity of the K-Means clustering starts to decrease. Furthermore, as expected, we observe the tendency of ARI to decrease with the number of clusters, whereas the purity behaves exactly in the opposite way. Thus, a compromise between these measures has to be found. We note that the purity and ARI become approximately constant starting from 10 clusters. So, for our application, ten can be regarded as the optimal number of clusters. In addition, Fig. 4 shows that the performance of image rationing is limited, both at pixel-level and patch-level. In the latter case, ARI is even constantly equal to 0 at patch-level, showing that the two partitions (i.e., the K-means labels and the manual labels) are random and that this method is not suitable for change analysis as it is not able to reflect any information regarding the class from which the changed element comes.

Furthermore, the main objective of the proposed method is also attained. Inspecting Fig. 4 again, the proposed pseudo-encoder manages to distinguish better among the types of changes, which leads to a better separation of the change domains, and, thus, to a better analysis of the changes in the time series compared to the widely used differencing and rationing methods.

5 Conclusions

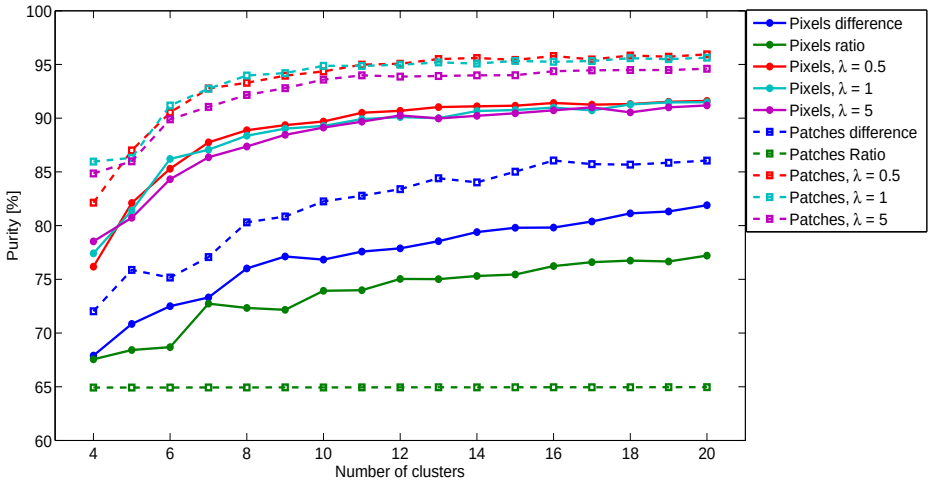
This paper presents a new image representation mainly designed to address analysis issues regarding time series. The proposed pseudo-encoding method aims to describe the changes that take place into a temporal series through a domain shift, which tries to give almost the same importance to all types of changes. Moreover, the method can be used at a patch level, or pixel level, depending on the desired level of precision. Experiments show good performance in terms of *purity* and *ARI* if compared to other commonly used methods (i.e., image differencing and image rationing). Furthermore, it is proven that change structures are endowed in each semantic class perceived by a user, meaning that a certain type of change is relevant for one class, while for another, not.



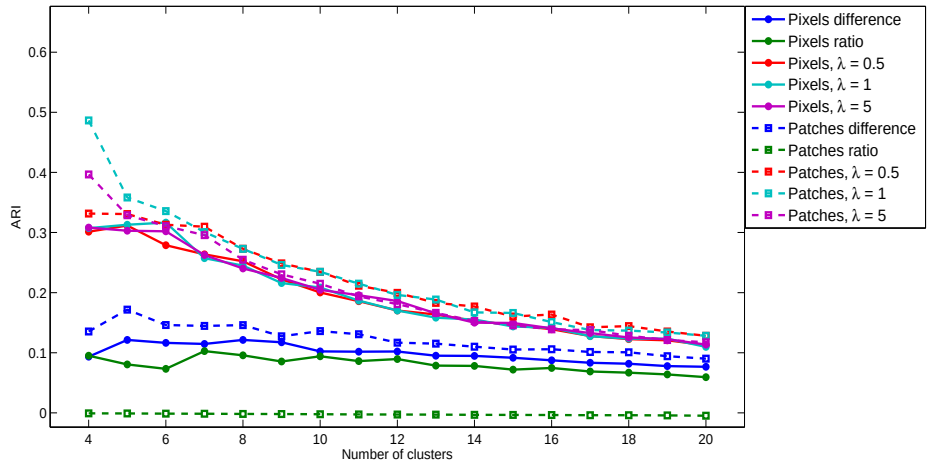
(a) One image from SITS (b) Ground truth

(c) Clustering map pixel-level (d) Clustering map patch-level

Fig. 3. The same space location is registered using six bands of frequency at two different times, namely June 24^{th}, 2001 (T_1) and October 14^{th}, 2001 (T_2). The ground truth is manually defined. Only the colored patches are taken into consideration, whereas the rest of the image is colored in black, representing a mixture of the 4 manually labeled classes, or other classes that are not relevant for our analysis. The K-Means clustering reveals a strong correlation between the types of changes and the manually defined classes.



(a) Purity



(b) Adjusted Rand Index

Fig. 4. Performance measures for the clustering depending on the number of clusters and the similarity measure, i.e. image differencing, image rationing, or proposed method with several values for λ.

Acknowledgments. The work has been partially funded by the Sectoral Operational Program Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134398.

References

1. J. Théau, "Change detection," in *Springer Handbook of Geographic Information*, pp. 75–94. Springer Berlin Heidelberg, 2012.
2. B. Gong, F. Sha, and K. Grauman, "Overcoming Dataset Bias: An Unsupervised Domain Adaptation Approach," in *Big Data Meets Computer Vision: First International Workshop on Large Scale Visual Recognition and Retrieval*, December 2012.
3. Y. Shi, and F. Sha, "Information-Theoretical Learning of Discriminative Clusters for Unsupervised Domain Adaptation," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2012.
4. L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.
5. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge University Press, New York, NY, USA, 2008.
6. T. Celik, "Unsupervised Change Detection in Satellite Images Using Principal Component Analysis and k -Means Clustering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 6, no. 4, pp. 772–776, 2009.
7. L. Bruzzone, D.F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1171–1181, 2000.
8. Y. Zheng, X. Zhang, B. Hou, G. Liu, "Using Combined Difference Image and K-Means Clustering for SAR Image Change Detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 3, pp. 691–695, 2014.
9. C. Vaduva, T. Costachioiu, C. Patrascu, I. Gavat, V. Lazarescu, and M. Datcu, "A latent analysis of earth surface dynamic evolution using change map time series," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2105–2118, 2013.
10. B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by v1," *Vision Research*, vol. 37, pp. 3311–3325, 1997.
11. J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2010, pp. 3360–3367.