UNSUPERVISED CHANGE DETECTION FOR MULTIMODAL REMOTE SENSING IMAGES VIA COUPLED DICTIONARY LEARNING AND SPARSE CODING

Vinicius Ferraris[(3)]*, *Nicolas Dobigeon*[(1)]†, *Yanna Cavalcanti*[(1)], *Thomas Oberlin*[(2)] *and Marie Chabert*[(1)]

[(1)] University of Toulouse, IRIT/INP-ENSEEIHT, Toulouse, France
[(2)] University of Toulouse, ISAE-SUPAERO, 31400 Toulouse, France
[(3)] University of São Paulo, Institute of Mathematics and Statistics, Brazil

firstname.lastname@enseeiht.fr, thomas.oberlin@isae-supaero.fr ,vinicius.ferraris@usp.br

ABSTRACT

Archetypal scenarios for change detection generally consider two images acquired through sensors of the same modality. The resolution dissimilarity is often bypassed though a simple preprocessing, applied independently on each image to bring them to the same resolution. However, in some important situations, e.g. a natural disaster, the only images available may be those acquired through sensors of different modalities and resolutions. Therefore, it is mandatory to develop general and robust methods able to deal with this unfavorable situation. This paper proposes a coupled dictionary learning strategy to detect changes between two images with different modalities and possibly different spatial and/or spectral resolutions. The pair of observed images is modelled as a sparse linear combination of atoms belonging to a pair of coupled overcomplete dictionaries learnt from the two observed images. Codes are expected to be globally similar for areas not affected by the changes while, in some spatially sparse locations, they are expected to be different. Change detection is then envisioned as an inverse problem, namely estimation of a dual code such that the difference between the estimated codes associated with each image exhibits spatial sparsity. A comparison with state-of-the-art change detection methods evidences the proposed method superiority.

***Index Terms*—** change detection, multimodal, unsupervised, resolution, dictionary learning.

1. INTRODUCTION

Change detection (CD) between remotely sensed images is a particularly effective mean to monitor the transformations occurring on the Earth surface over a period of time [1]. In general, unsupervised CD techniques are constrained to two images of the same modality with the same spatial and spectral resolutions acquired over the same geographical location at different dates [2]. This scenario is suitable for a straight comparison of homologous pixels such as pixel-wise differencing or rationing depending on the modality [3,4]. However, in some specific cases, e.g. in emergency situations, defense and security, the only available images may be of different modalities and resolutions. These dissimilarities introduce additional issues in the context of operational CD that are not addressed by most classical methods. In the case of same modality and different resolutions, state-of-the art methods come down to conventional CD methods after preprocessing steps applied independently on the two images, e.g., resampling operations intended to reach the same spatial and spectral resolutions [5, 6]. Nevertheless, these preprocessing steps may waste relevant information since they do not take into account the strong interplay existing between the two images. On the other hand, multimodal CD is not a simple task for unsupervised methods due to the lack of direct relation between modalities. Supervised CD tries to infer these relations, but with an usual unworkable cost of collecting ground data, which makes them not suitable for real applications [7, 8]. The literature about multimodal CD is very limited, even if it has always figured out as an important topic since the early works [9, 10] till nowadays with other few relevant references [11–16]. Although some of them present relatively high detection performance, they are often restrained to a specific modality or to a specific target application [16]. The other ones estimate some metrics from unchanged trained samples, which prevents their application within a fully unsupervised context [10, 14, 15]. Recently, some unsupervised multimodal CD methods based on coupled dictionary learning were proposed [17, 18]. Both methods exploits that high errors occur in zones affected by changes when the images are reconstructed using coupled dictionaries estimated from the observed images. Nevertheless, they run into difficulties because of the large variety of possible scenarios. Indeed, a different noise statistical model for each modality, possibly overlapping patches in reconstruction as well as the resolution dissimilarity [5,6] may negatively impact these methods performance.

This paper proposes an unsupervised CD method able to deal with images dissimilar in terms of modality and of spatial and/or spectral resolutions. The adopted methodology, similar to [17, 18], learns coupled dictionaries able to conveniently represent multimodal remote sensing images of the same geographical location. The problem is formulated as a joint estimation of the coupled dictionary and sparse code for each observed image. Additionally, appropriate statistical models are used to better fit the modalities of the pair of observed images. Overlapping patches are also taken into account during the estimation process. Finally, to better couple images with different resolutions, additional scaling matrices [19] are jointly estimated within the whole process. The overall estimation process is formulated as an inverse problem. Due to the nonconvex nature of the problem, it is solved iteratively using the proximal alternating linearized minimization (PALM) algorithm [20], which ensures convergence towards a critical point for some nonconvex non-smooth problems. CD is, then, envisaged through the differences between sparse codes estimated for each image using the estimated coupled dictionaries. This paper is organized as follows.

*Part of this work has been supported by the 2019/06254-0 project of FAPESP.

†Part of this work has been supported by the ANR-3IA Artificial and Natural Intelligence Toulouse Institute (ANITI).

Section 2 formulates the CD problem. In Section 3, the proposed CD algorithm is described. Section 4 analyzes the proposed method performance through experiments. Section 5 concludes the paper.

2. PROBLEM FORMULATION

Let consider that the two observed images $\mathbf{Y}_1 \in \mathbb{R}^{L_1 \times N_1}$ and $\mathbf{Y}_2 \in \mathbb{R}^{L_2 \times N_2}$ have been acquired by two sensors S_1 and S_2 at times t_1 and t_2, respectively, over the same geographical location. Note that the time ordering of the acquisitions is indifferent. Each image, consist of N_j voxels $\mathbf{y}_i \in \mathbb{R}^{L_j}$ (with $j = \{1, 2\}$) stacked lexicographically. The voxel dimension, L_j, represents different quantities depending on the modality of the data, e.g, it denotes the number of spectral bands in the case of multiband optical images [6]. Assuming that the scene has changed between the two acquisition times, our goal is to extract significant change information from this pair of images. This problem can be challenging in case of dissimilarity between S_1 and S_2 in terms of modality and of resolutions, since it prevents the use of classical CD algorithms [21] or of methods designed for specific image modalities [5, 6]. In order to alleviate this issue, we propose to model each image in terms of its latent (i.e., noise-free and unobserved) image, $\mathbf{X}_j \in \mathbb{R}^{L_j \times N_j}$. Each modality has a particular noise statistics that can been related to a particular divergence measure $\mathcal{D}_j(\mathbf{Y}_j|\mathbf{X}_j)$ derived from the likelihood function $p(\mathbf{Y}_j|\mathbf{X}_j)$ [8]. To account for the problem of resolution dissimilarity we resort to the paradigm of sparse representations over an overcomplete dictionary [22, 23] More precisely, each latent image $\mathbf{X}_j$ is decomposed into a set of N_p 3D-patches $\mathbf{p}_i \in \mathbb{R}^{K_j^2 L_j}$ with $i \in \{1, \ldots, N_\mathrm{p}\}$ using the binary extraction operator $\mathcal{R}_{ji} : \mathbb{R}^{L_j \times N_j} \to \mathbb{R}^{K_j^2 L_j}$. The ith $K_j^2 \times L_j$-pixel patch $\mathbf{p}_i = \mathcal{R}_{ji}\mathbf{X}_j$ is represented in its vectorized form with $K_j > 1$ defining the spatial size of the patches. The number of patches $N_\mathrm{p} \in \{1, \ldots, N\}$ is tailored by the end-user and patches may overlap. Considering the dictionary-based representation principles, these patches are assumed to be approximately and independently modelled as sparse combinations of atoms belonging to an overcomplete dictionary $\mathbf{D}_j = [\mathbf{d}_{j1}, \cdots, \mathbf{d}_{jN_\mathrm{d}}] \in \mathbb{R}^{K_j^2 L_j \times N_\mathrm{d}}$ with corresponding sparse code $\mathbf{a}_{ji} \in \mathbb{R}^{N_\mathrm{d}} | \ \forall i \in \{1, \ldots, N_\mathrm{p}\}$ as:

$$\mathbf{p}_{ji}|\mathbf{D}_j, \mathbf{a}_{ji} \sim \mathcal{N}\left(\mathbf{D}_j\mathbf{a}_{ji}, \sigma_j^2\mathbf{I}_{N_\mathrm{d}}\right) \tag{1}$$

with the dictionary constrained to the set

$$\mathcal{S} \triangleq \left\{\mathbf{D}_j \in \mathbb{R}_+^{K_j^2 L_j \times N_\mathrm{d}} \ | \ \forall l \in \{1, \ldots, N_\mathrm{d}, \} \ \|\mathbf{d}_{jl}\|_2^2 = 1\right\}, \tag{2}$$

and the code assigning to single-side exponential (i.e., Laplacian) prior distribution

$$\mathbf{a}_{ji} \sim \prod_{i=1}^{N_\mathrm{d}} \mathcal{L}(a_{ji}; \lambda_j). \tag{3}$$

Recall that the two observed images represent the same geographical location. Consequently, two homologous patches extracted from each latent image represent the same geographical spot. In the absence of changes between the two acquisition times, the sparse codes associated with the corresponding latent images are expected to be approximately the same ($\mathbf{a}_{1i} \approx \mathbf{a}_{2i}$) when the two learned dictionaries are coupled [24–26], i.e., able to derive a joint representation for homologous multiple observations in a latent coupled space [17]. On the contrary, in case of changes between the acquisition times, pairs of homologous patches in the change location do not represent exactly the same scene and perfect reconstruction cannot be achieved using the same code. Relaxing the constraint of perfect reconstruction in these regions allows for an accurate reconstruction of both images while spatially mapping change locations. In the specific context of CD addressed in this work, this finding suggests to evaluate any change between the two observed, or equivalently latent, images by comparing the code change matrix $\Delta\mathbf{A}$:

$$\mathbf{A}_2 = \mathbf{A}_1 + \Delta\mathbf{A}. \tag{4}$$

The magnitude of the code change matrix accounts for significant changes in the observed scene and is expected to exhibit spatial sparsity. As a consequence, the CD problem can be formulated as the joint estimation of one code matrix, representing one of the scenes, and of the change code matrix, i.e. of $\{\mathbf{A}_1, \Delta\mathbf{A}\}$, as well as of the pair of coupled dictionary $\{\mathbf{D}_1, \mathbf{D}_2\}$. The next paragraph introduces the CD-driven optimization problem to be solved.

3. COUPLED DICTIONARY LEARNING FOR CD

3.1. Objective function

Constraining the dictionaries to belong to $\mathcal{S}$ defined by (2) allow to reconstruct a pair of unchanged homologous patches with exactly the same code, while changed patches would be associated to different codes, in good dictionary coupling condition. Nevertheless, it would not take into account differences in data dynamics or resolutions, a very current problem facing remote sensing multitemporal data. Therefore, we propose to use an additional diagonal scaling matrix, as in [19], constrained to the set $\mathcal{C} \triangleq \left\{\mathbf{S} \in \mathbb{R}_+^{N_\mathrm{d1} \times N_\mathrm{d1}} \ | \ \mathbf{S} = \mathrm{diag}(\mathbf{s}), \ \mathbf{s} \succeq 0\right\}$ to ensure that the sparse codes of the two observed images are directly comparable. It allow us to write the joint representation model for a pair of homologous patches $\{\mathbf{p}_{1i}, \mathbf{p}_{2i}\}$ as:

$$\begin{aligned} \mathbf{p}_{1i} &= \mathcal{R}_{1i}\mathbf{X}_1 \approx \mathbf{D}_1\mathbf{S}\mathbf{a}_{1i} \\ \mathbf{p}_{2i} &= \mathcal{R}_{2i}\mathbf{X}_2 \approx \mathbf{D}_2\mathbf{a}_{2i} = \mathbf{D}_2\left(\mathbf{a}_{1i} + \Delta\mathbf{a}_i\right) \end{aligned} \tag{5}$$

At this point, we successfully map the orignal space of the CD problem, in which it was not possible to do a direct comparison between pixels, into a representation space that allows element-wise comparisons. Naturally, in scenarios involving CD, most of the pixels are expected to remain unchanged between acquisitions. To account for possible changes in some specific locations while most of the patches remain unchanged, as in [6], we propose to adopt an $\ell_{2,1}$-norm regularization over the code change matrix to induce spatial sparsity while keeping the strong changes in the code energy,

$$\phi_2\left(\Delta\mathbf{A}\right) = \|\Delta\mathbf{A}\|_{2,1} = \sum_{i=1}^{N_\mathrm{p}} \|\Delta\mathbf{a}_i\|_2 . \tag{6}$$

Then, by incorporating all previously defined regularizations, the joint MAP estimator of the quantities of interest reads

$$\hat{\mathbf{\Theta}}_\mathrm{MAP} = \left\{\hat{\mathbf{X}}_{1,\mathrm{MAP}}, \hat{\mathbf{X}}_{2,\mathrm{MAP}}, \hat{\mathbf{D}}_{1,\mathrm{MAP}}, \hat{\mathbf{D}}_{2,\mathrm{MAP}}, \hat{\mathbf{S}}_\mathrm{MAP}, \hat{\mathbf{A}}_{1,\mathrm{MAP}}, \Delta\hat{\mathbf{A}}_\mathrm{MAP}\right\}$$

associated with the following minimization problem

$$\begin{aligned} \hat{\mathbf{\Theta}}_\mathrm{MAP} \in \underset{\mathbf{\Theta}}{\mathrm{argmin}} \ & \mathcal{D}(\mathbf{Y}_1|\mathbf{X}_1) + \frac{\sigma_1^2}{2}\sum_{i=1}^{N_\mathrm{p}} \|\mathcal{R}_{1i}\mathbf{X}_1 - \mathbf{D}_1\mathbf{S}\mathbf{a}_{1i}\|_\mathrm{F}^2 \\ & + \mathcal{D}(\mathbf{Y}_2|\mathbf{X}_2) + \frac{\sigma_2^2}{2}\sum_{i=1}^{N_\mathrm{p}} \left\|\mathcal{R}_{2i}\mathbf{X}_2 - \mathbf{D}_2\left(\mathbf{a}_{1i} + \lambda\|\mathbf{A}_1\|_1\Delta\mathbf{a}_i\right)\right\|_\mathrm{F}^2 \\ & + \lambda\|\mathbf{A}_1 + \Delta\mathbf{A}\|_1 + \gamma\|\Delta\mathbf{A}\|_{2,1} + \iota_\mathcal{S}(\mathbf{D}_1) + \iota_\mathcal{S}(\mathbf{D}_2) + \iota_\mathcal{C}(\mathbf{S}), \end{aligned} \tag{7}$$

where $\iota_{\cdot}(\cdot)$ represents the indicator function on a set.

3.2. Minimization algorithm

The nature of the problem (7) is nonconvex and nonsmooth. The adopted minimization strategy follows the proximal alternating linearized minimization (PALM) scheme, ensuring the convergence towards a local critical point $\mathbf{\Theta}^*$ [20]. It iteratively minimizes the objective function with respect to each block of variables from $\mathbf{\Theta}$ using descent gradient steps followed by proximal mapping associated to the corresponding nonsmooth functions (hereafter generically denoted $f_{\cdot}(\cdot)$). The overall algorithm is sketched in Algo. 1. In the following paragraphs the main steps of the proposed CD-driven coupled dictionary learning (CDL) are described in more details.

Algorithm 1: PALM-CDL

Data: $\mathbf{Y}$
Input: $\mathbf{A}_1^{(0)}, \Delta\mathbf{A}^{(0)}, \mathbf{D}_1^{(0)}, \mathbf{D}_2^{(0)}, \mathbf{S}^{(0)}, \mathbf{X}_1^{(0)}, \mathbf{X}_2^{(0)}$
$k \leftarrow 0$
begin
  while *stopping criterion not satisfied* **do**
    for $\mathbf{G} \in \Theta$ **do**
      $\text{prox}_{f_{\mathbf{G}}}^{L_{\mathbf{G}}}\left(\mathbf{G}^{(k)} - \frac{1}{L_{\mathbf{G}}^{(k)}}\nabla_{\mathbf{G}}H(\mathbf{\Theta})\right)$
    $k \leftarrow k+1$
  for $\mathbf{G} \in \Theta$ **do**
    $\hat{\mathbf{G}} \leftarrow \mathbf{G}^{(k+1)}$
Result: $\hat{\mathbf{A}}_1, \Delta\hat{\mathbf{A}}, \hat{\mathbf{D}}_1, \hat{\mathbf{D}}_2, \hat{\mathbf{S}}, \hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2$

Optimization with respect to $\mathbf{A}_1$ – Assuming that the remaining variables are fixed but $\mathbf{A}_1$, the PALM updating step can be written

$$\nabla_{\mathbf{A}_1}H(\mathbf{\Theta}) = \sigma_1^2\mathbf{S}^T\mathbf{D}_1^T\left(\mathbf{D}_1\mathbf{S}\mathbf{A}_1 - \mathbf{P}_1\right) + \sigma_2^2\mathbf{D}_2^T\left(\mathbf{D}_2\left(\mathbf{A}_1 + \Delta\mathbf{A}\right) - \mathbf{P}_2\right) + \lambda\frac{\left[\mathbf{A}_1 + \Delta\mathbf{A}\right]_i}{\sqrt{\left[\mathbf{A}_1 + \Delta\mathbf{A}\right]_i^2 + \epsilon_{\mathbf{A}_1}^2}}$$

where $[\cdot]_i/[\cdot]_i$ should be understood as an element-wise operation, $\mathbf{P}_j = \left[\mathbf{p}_{j1}, \cdots, \mathbf{p}_{jN_\text{p}}\right]$ is the $K_j^2L_j \times N_\text{p}$-matrix that stacks all extracted patches. The associated Lipschitz constant writes $L_{\mathbf{A}_1}^{(k)} = \sigma_1^2\left\|\mathbf{S}^T\mathbf{D}_1^T\mathbf{D}_1\mathbf{S}\right\| + \sigma_2^2\left\|\mathbf{D}_2^T\mathbf{D}_2\right\| + \frac{\lambda}{\epsilon_{\mathbf{A}_1}}$. Note that $f_{\mathbf{A}_1} = \lambda\left\|\cdot\right\|_1 + \geq 0$ can be simply computed by considering the positive part of the soft-thresholding operator [27].

Optimization with respect to $\Delta\mathbf{A}$ – Similarly, considering the single block optimization variable $\Delta\mathbf{A}$ and consistent notations, the PALM update can be derived considering

$$\nabla_{\Delta\mathbf{A}}H(\mathbf{\Theta}) = \sigma_2^2\mathbf{D}_2^T\left(\mathbf{D}_2\left(\mathbf{A}_1 + \Delta\mathbf{A}\right) - \mathbf{P}_2\right) + \lambda\frac{\left[\mathbf{A}_1 + \Delta\mathbf{A}\right]_i}{\sqrt{\left[\mathbf{A}_1 + \Delta\mathbf{A}\right]_i^2 + \epsilon_{\mathbf{A}_1}^2}}$$

with $L_{\Delta\mathbf{A}}^{(k)} = \sigma_2^2\left\|\mathbf{D}_2^T\mathbf{D}_2\right\| + \frac{\lambda}{\epsilon_{\mathbf{A}_1}}$. The proximal step $f_{\Delta\mathbf{A}} = \left\|\cdot\right\|_{2,1}$ can be simply computed as a group soft-thresholding operator [6].

Optimization with respect to $\mathbf{D}_j$ – Optimizing with respect to $\mathbf{D}_j$ with $j \in \{1, 2\}$, the PALM updating steps can be written

$$\nabla_{\mathbf{D}_j}H(\mathbf{\Theta}) = \sigma_j^2\left(\mathbf{D}_j\bar{\mathbf{A}}_j - \mathbf{P}_j\right)\bar{\mathbf{A}}_j^T$$

where $L_{\mathbf{D}_j}^{(k)} = \sigma_j^2\left\|\bar{\mathbf{A}}_j\bar{\mathbf{A}}_j^T\right\|$ with $\bar{\mathbf{A}}_1 = \mathbf{S}\mathbf{A}_1$ and $\bar{\mathbf{A}}_2 = \mathbf{A}_1 + \Delta\mathbf{A}$. Note that $f_{\mathbf{D}_j} = \iota_{\mathcal{S}}$ can be computed as in [23], keeping only the values greater than zero.

Optimization with respect to $\mathbf{S}$ – Updating the scaling matrix $\mathbf{S}$ can be written

$$\nabla_{\mathbf{S}}H(\mathbf{\Theta}) = \sigma_1^2\mathbf{D}_1^T\left(\mathbf{D}_1\mathbf{S}\mathbf{A}_1 - \mathbf{P}_1\right)\mathbf{A}_1^T \tag{8}$$

with $L_{\mathbf{S}}^{(k)} = \sigma_1^2\left\|\mathbf{D}_1^T\mathbf{D}_1\mathbf{A}_1\mathbf{A}_1^T\right\|$. The proximal update $f_{\mathbf{S}} = \iota_{\mathcal{C}}$ constrains all its diagonal element to be positive.

Optimization with respect to $\mathbf{X}_j$ – The updates of the latent images $\mathbf{X}_j$ ($j \in \{1, 2\}$) are achieved considering

$$\nabla_{\mathbf{X}_j}H(\mathbf{\Theta}) = \sigma_j^2\sum_{i=1}^{N_\text{p}}\mathcal{R}_{ji}^T\left(\mathcal{R}_{ji}\mathbf{X}_j - \mathbf{D}_j\bar{\mathbf{a}}_{ji}\right) \tag{9}$$

with $L_{\mathbf{X}_j}^{(k)} = \sigma_j^2\left\|\sum_{i=1}^{N_\text{p}}\mathcal{R}_{ji}^T\mathcal{R}_{ji}\right\|$. Note that, $f_{\mathbf{X}_j} = \mathcal{D}_j(\mathbf{Y}_j|\cdot)$ represents the proximal mapping for the divergence measure associated with the likelihood function characterizing the modality of the observed image $\mathbf{Y}_j$ [8].

4. EXPERIMENTAL RESULTS

Real images with synthetic changes – This section analyzes the performance of the proposed CD method. Real data for CD is rarely available. Thus, to test the proposed method, a simulation protocol inspired by the Wald's protocol [5, 28] has been used to generate observed images from reference images. Two multimodal reference images acquired at the same date have been selected as change-free observed images. Then, by conducting simple copy-paste of regions, changes between images have been generated as well as the corresponding ground-truth maps. This process allows synthetic yet realistic changes to be incorporated within one of these images, w.r.t. a pre-defined binary reference change mask locating the pixels affected by these changes and further used to assess the performance of the CD algorithms. The reference image pair is one $540 \times 525 \times 3$ multispectral Sentinel-2 image (Bands 2 to 4) and a 540×525 multi-looked SAR intensity Sentinel-1 image acquired over the same geographical area, i.e., the Mud Lake region in Lake Tahoe, at the same date on April 12th 2016. The combination of reference and a generated changed image allow us to divide three different scenarios to assess the performance of the proposed CD method. Scenario 1 considering two optical images, Scenario 2 considering two SAR images and Scenario 3 considering a SAR image and an optical image. A set of 10 predefined change masks has been designed according to specific copy-paste change rules similar as the ones introduced by [5]. The proposed coupled dictionary learning approach (PALM-CDL) is compared to two CD methods. The first one, proposed by [17] and denoted $\hat{\mathbf{m}}_\text{F}$, also peforms a coupled dictionary learning strategy, but represents changes according to the final reconstruction error grouped into change/no changed class by a Fuzzy-C-means clustering method. The second one is robust fusion-based method, denoted $\hat{\mathbf{m}}_\text{RF}$, proposed [6] that is able to deal exclusively with multi-band optical images of different resolutions. The performance of CD methods is evaluated through the receiving operator characteristics (ROC) curves, that display the probability of false alarm (PFA) as a function of the probability of detection (PD). Figure 1 presents the averaged ROC curves obtained with the three methods in the three predefined scenarios. Additionally, two quantitative measures of detection performance can be extracted from

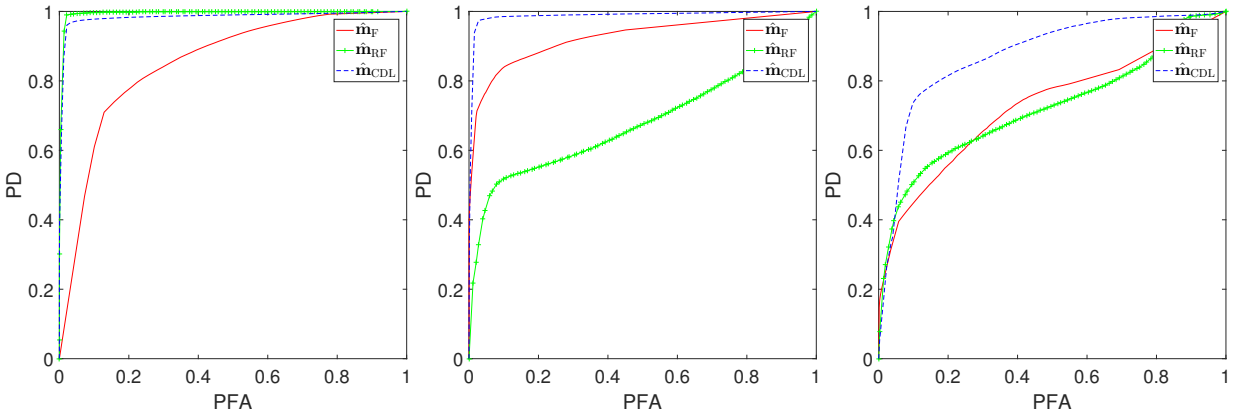


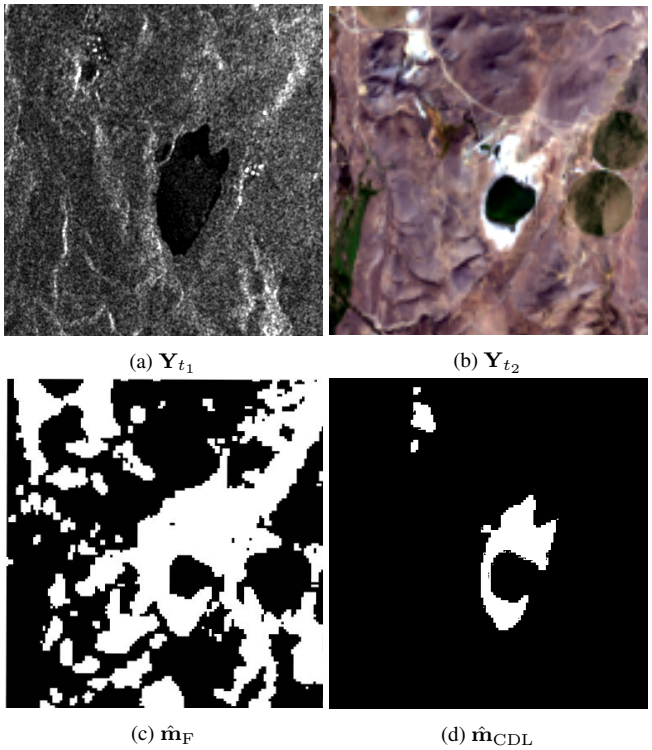
Fig. 1: Real images affected by synthetic changes: ROC curves for Scenario 1 (left), Scenario 2 (middle) and Scenario 3 (right).

these ROC curves: the area under the curve (AUC), corresponding to the integration of the ROC curve and the distance (Dist.) between the interception of the ROC curve with the diagonal line, $\mathrm{PFA} = 1 - \mathrm{PD}$, and the no detection point $(\mathrm{PFA} = 1, \mathrm{PD} = 0)$. In both cases, the better the detection the closer to one the measure. For both scenarios, the proposed method shows overall higher detection performance than the other methods. Only in the Scenario 1, the RF method is specialized in, the performance of this method is slightly better than the proposed one. Nevertheless, the RF method cannot generalize to other cases.

Real images with real changes – To illustrate the high precision and the benefits of the proposed algorithm, Fig. 2 presents a Sentinel-1 SAR image with 10m spatial resolution acquired on April 12th 2016 and a Landsat 8 MS (RGB) image acquired on September 22th 2015 with 30m spatial resolution. The proposed method was compared with the Fuzzy method, which was not capable to output reliable results, while the proposed one shows very high accuracy in the most relevant change, named the draugh of the lake.

Table 1: Real images affected by synthetic changes for Scenarios 1–3: quantitative detection performance (AUC and distance).

		$\hat{\mathbf{m}}_{\mathrm{F}}$	$\hat{\mathbf{m}}_{\mathrm{RF}}$	$\hat{\mathbf{m}}_{\mathrm{CDL}}$
Scenario 1	AUC	0.8520	**0.9946**	0.9838
	Dist.	0.7867	**0.9802**	0.9677
Scenario 2	AUC	0.9251	0.6819	**0.9871**
	Dist.	0.8587	0.6185	**0.9727**
Scenario 3	AUC	0.7277	0.7227	**0.8755**
	Dist.	0.6758	0.6604	**0.8097**



(a) $\mathbf{Y}_{t_1}$ (b) $\mathbf{Y}_{t_2}$

(c) $\hat{\mathbf{m}}_{\mathrm{F}}$ (d) $\hat{\mathbf{m}}_{\mathrm{CDL}}$

Fig. 2: Real images affected by real changes without ground truth, Scenario 3 (different spatial resolutions): (a) Sentinel-1 SAR image $\mathbf{Y}_{t_1}$ acquired on 04/12/2016, (b) Landsat 8 MS image $\mathbf{Y}_{t_2}$ acquired on 09/22/2015, (c) change map $\hat{\mathbf{m}}_{\mathrm{F}}$ of the fuzzy method and (d) change map $\hat{\mathbf{m}}_{\mathrm{CDL}}$ of the proposed method.

5. CONCLUSIONS

This paper proposed a new CD method to deal with multimodal images with possible different resolutions. The CD problem was tackled as a coupled dictionary estimation problem formulated as an inverse problem. Changes were supposed to correspond to the differences between the sparse codes estimated for each homologous pair of patches extracted from two images acquired at the same geographical location. The overall estimation process was solved using an iterative solution based on alternate proximal gradient steps which presented guarantees of convergence to a critical point. The proposed method showed far higher performance and flexibility than the state-of-the-art most relevant methods in various different scenarios involving multimodal remote sensing images with different resolutions.

6. REFERENCES

[1] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing," *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, Sept. 2015.

[2] J. B. Campbell and R. H. Wynne, *Introduction to remote sensing*, 5th ed. New York: Guilford Press, 2011.

[3] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 294–307, 2005.

[4] F. Bovolo and L. Bruzzone, "The Time Variable in Data Fusion: A Change Detection Perspective," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 8–26, Sept. 2015.

[5] V. Ferraris, N. Dobigeon, Q. Wei, and M. Chabert, "Detecting Changes Between Optical Images of Different Spatial and Spectral Resolutions: A Fusion-Based Approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1566–1578, March 2018.

[6] ——, "Robust Fusion of Multiband Images With Different Spatial and Spectral Resolutions for Change Detection," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 175–186, 2017.

[7] F. Bovolo and L. Bruzzone, "A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 218–236, Jan. 2007.

[8] V. Ferraris, N. Dobigeon, Y. Cavalcanti, T. Oberlin, and M. Chabert, "Coupled dictionary learning for unsupervised change detection between multimodal remote sensing images," *Computer Vision and Image Understanding*, 2019.

[9] J. G. Kawamura, "Automatic Recognition of Changes in Urban Development from Aerial Photographs," *IEEE Trans. Systems, Man, Cybernet.*, vol. SMC-1, no. 3, pp. 230–239, 1971.

[10] L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1350–1359, 1999.

[11] J. Inglada, "Similarity measures for multisensor remote sensing images," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, vol. 1. IEEE, 2002, pp. 104–106.

[12] D. Lu, P. Mausel, E. Brondízio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, 2004.

[13] V. Alberga, M. Idrissa, V. Lacroix, and J. Inglada, "Comparison of similarity measures of multi-sensor images for change detection applications," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. IEEE, 2007, pp. 2358–2361.

[14] G. Mercier, G. Moser, and S. Serpico, "Conditional Copulas for Change Detection in Heterogeneous Remote Sensing Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, 2008.

[15] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, 2015.

[16] M. Chabert, J.-Y. Tourneret, V. Poulain, and J. Inglada, "Logistic regression for detecting changes between databases and remote sensing images," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. IEEE, 2010, pp. 3198–3201.

[17] M. Gong, P. Zhang, L. Su, and J. Liu, "Coupled Dictionary Learning for Change Detection From Multisource Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7077–7091, 2016.

[18] X. Lu, Y. Yuan, and X. Zheng, "Joint Dictionary Learning for Multispectral Change Detection," *IEEE Transactions on Cybernetics*, vol. 47, no. 4, pp. 884–897, 2017.

[19] N. Seichepine, S. Essid, C. Fevotte, and O. Cappe, "Soft Nonnegative Matrix Co-Factorization," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5940–5949, 2014.

[20] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459–494, 2014.

[21] F. Bovolo and L. Bruzzone, "The Time Variable in Data Fusion: A Change Detection Perspective," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 8–26, 2015.

[22] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[23] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. Int. Conf. Machine Learning (ICML)*. ACM, 2009, pp. 689–696.

[24] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image Super-Resolution Via Sparse Representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.

[25] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled Dictionary Training for Image Super-Resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, 2012.

[26] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International Conference on Curves and Surfaces*. Springer, 2010, pp. 711–730.

[27] N. Parikh, S. Boyd, and others, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[28] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images," *Photogrammetric engineering and remote sensing*, vol. 63, no. 6, pp. 691–699, 1997.