Spectral-Spatial Aware Unsupervised Change Detection with Stochastic Distances and Support Vector Machines

Rogério Negri [1], Alejandro C. Frery [2], Wallace Casaca [2], Samara Azevedo [2], Maurício Araújo [2], Erivaldo Silva [2], and Enner Alcântara [2]

[1]UNESP
[2]Affiliation not available

October 30, 2023

Spectral-Spatial Aware Unsupervised Change Detection with Stochastic Distances and Support Vector Machines

Rogério G. Negri, Alejandro C. Frery, *IEEE Senior Member*, Wallace C. O. Casaca, Samara C. Oliveira, Maurício. A. Dias, Erivaldo. A. Silva and Enner. H. Alcântara

Abstract

Change detection is a topic of great interest in remote sensing. A good similarity metric to compute variations among the images is the key for high-quality change detection. However, most existing approaches rely on fixed threshold values or user-provided ground truth in order to be effective. The inability of dealing with artificial objects such as clouds and shadows is a significant difficulty for many change detection methods. We propose a new unsupervised change detection framework to address those critical points. The notion of homogeneous regions is introduced together with a set of geometric operations and statistic-based criteria to formally characterize and distinguish change and non-change areas in a pair of remote sensing images. Moreover, a robust and statistically well-posed family of stochastic distances is also proposed, which allows comparing the probability distributions of different regions/objects in the images. These stochastic measures are then used to train a SVM-based approach in order to detect the change/non-change areas. Three study cases using images acquired with different sensors are given in order to compare the proposed method with other well-known unsupervised methods.

Index Terms

Unsupervised change detection, stochastic distance, single-class SVM, classification.

I. Introduction

CHANGE DETECTION is an active research field that seeks to track land cover differences in images remotely acquired over the same region at different moments [1]. It has appeared in several applications, ranging from urban environmental monitoring [2] to vegetation mapping [3]. In forestry, for example, the identification of spatial-temporal changes allows for a better understanding of how ecosystems behave along the time, elucidating the progressive interaction between natural phenomena and human activities.

Wu et al. [4] review recent approaches devoted to identifying abrupt changes in remotely sensed images. The authors grouped them into post-classification inspections, Principal Component Analysis (PCA), and arithmetic with image bands and spectral indexes. Another approach commonly adopted to differentiate these methods is by employing the well-known taxonomy of image classification, i.e., a particular algorithm is either supervised or unsupervised. More specifically, a change detection method is

R. G. Negri is with Department of Environmental Engineering, Sciences and Technology Institute, So Paulo State University (UNESP), So Jos dos Campos, São Paulo, Brazil. e-mail: rogerio.negri@unesp.br

A. C. Frery is with the Universidade Federal de Alagoas, Maceió, Brazil, and the Key Lab of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, Xi'an, China. (e-mail: acfrery@laccan.ufal.br)

W. C. O. Casaca is with Department of Energy Engineering, So Paulo State University (UNESP), Rosana, São Paulo Brazil

S. C. Oliveira is with Natural Resources Department, Federal University of Itajuba (UNIFEI), Itajub, Minas Gerais, Brazil

M. A. Dias is with Department of Mathematics and Computer Science, School of Sciences and Technology, So Paulo State University (UNESP), Presidente Prudente, São Paulo, Brazil

E. A. Silva is with Department of Cartography, School of Sciences and Technology, So Paulo State University (UNESP), Presidente Prudente, São Paulo, Brazil

E. H. Alcântara is with Department of Environmental Engineering, Sciences and Technology Institute, So Paulo State University (UNESP), Campus So Jos dos Campos, São Paulo, Brazil.

Manuscript received XX YY, 20ZZ; revised WW UU, 20VV.

referred to as "supervised" if a set of labeled samples is given as input to the algorithm; otherwise, it is called "unsupervised." Since our approach relies on unsupervised learning, the forthcoming discussion is conducted on the basis of the unsupervised literature.

In remote sensing, the most representative precursors of unsupervised change detection methods are: the Change Vector Analysis (CVA) [5], and the framework described by Celik [6], the so-called PCA-KM, which integrates PCA and k-means clustering. The first method, CVA, comprises three stages: a pre-processing step (radiometric and geometric corrections), the computation of the change vector values for each pixel, usually through the norm of a feature vector difference for a pair of instants, and, finally, the binarization of the generated vectors into change and non-change segments, by using a thresholding scheme such as Otsu [7] or Kitller-Illinghworth [8]. Concerning PCA-KM, it relies on a simple, but highly-efficient algorithm to accomplish the identification of apparent changes, which is also robust to noise.

The change evidence is computed as a vector from a pair of co-registered images obtained at two different instants. Next, the change vector image is partitioned into disjoint blocks and their values are used to generate a high dimensional vector. PCA is applied on these data to extract the most relevant information from each block. The final representation, usually composed by the first principal components, is handed over the k-means algorithm in order to distinguish the changes and non-changes regions.

Despite its high flexibility in dealing with several applications, CVA and PCA-KM also have some intrinsic weaknesses which can undermine their accuracy and effectiveness. For instance, the use of input data without an adequate radiometric correction in CVA may produce very unstable and inconsistent outputs. Furthermore, the use of only a single cut-off value to distinguish between change and non-change regions usually leads to omission and inclusion errors. Regarding PCA-KM, the presence of high-contrast elements like clouds and background regions may impair the clustering process carried thought the k-means algorithm. The accommodation of such outliers into the clusters may induce omission and inclusion errors.

Aiming at addressing these drawbacks without penalizing the computational cost, a new unsupervised change detection technique is proposed in this paper. Given an image pair, our method takes two main steps: the identification of non-change homogeneous regions, and the decision rule modeling, based on the identified non-change homogeneous regions, which allows the discrimination between change and non-change areas in the input pair. In our approach, we exploit the ideas of stochastic distances [9], [10] to drive the identification of non-change homogeneous regions. Finally, the decision rule is computed by a Single-Class Support Vector Machine (SVM) classification in order to properly label the change areas.

Contributions. In summary, the main contributions of this paper are the following:

- A fully unsupervised change detection method which unifies a robust and statistically well-posed family of stochastic distances with a SVM-based approach.
- A set of geometric structures as well as a statistic criterion especially designed to characterize and identify homogeneous regions within a sequence of images.
- In contrast to CVA and PCA-KM, our method is capable of coping with outliers such as clouds and background regions, while still avoiding a fixed threshold to classify the images.
- The proposed method is modular and, thus, flexible regarding the use of other statistical measures and decision functions beyond those presented in the following formalization.

To assess our method while comparing it with CVA and PCA-KM, we study three cases of landscape changes occurred as consequence of rupture dams in Brazil. We employed images acquired from different satellites: Landsat-8, Sentinel-2, and ALOS-1.

This paper is organized as follows: Section II presents the the underlying concepts of hypotheses tests from stochastic distances, and the single-class classification based on SVM. Section III introduces the proposed unsupervised change detection method, formalizing it under the basis of stochastic distances. Section IV provides the details about the data, experiment design, the results, and discussions. Finally, Section V summarizes the findings of this paper.

II. Mathematical Background

A. Testing hypothesis from stochastic distances

Stochastic distance has its origin on divergence measures, i.e., based on the Information Theory, as established in the pivotal work by Shannon [11]. A divergence measure quantifies the level of complexity when the difference of two models needs to be computed.

Salicrú et al. [9] obtained a generalization of divergence measures, known as h-ϕ divergence family, with interesting statistical properties. Let X and Y be two random variables with probability density functions $f_X(\mathbf{x};\boldsymbol{\theta}_X)$ and $f_Y(\mathbf{x};\boldsymbol{\theta}_Y)$, defined over the same support Ξ, indexed by parameters $\boldsymbol{\theta}_X$ and $\boldsymbol{\theta}_Y$. The h-ϕ divergence between X and Y is:

$$d_\phi^h(X,Y) = h\left(\int_{\xi\in\Xi}\phi\left(\frac{f_X(\xi;\boldsymbol{\theta}_X)}{f_Y(\xi;\boldsymbol{\theta}_Y)}\right)f_Y(\xi;\boldsymbol{\theta}_Y)d\xi\right), \tag{1}$$

$\phi\colon (0,\infty)\to[0,\infty)$ is a convex function, $h\colon (0,\infty)\to[0,\infty)$ is a strictly increasing function with $h(0)=0$, and $h'(t)$ strictly positive for any value $t\in(0,\infty)$.

Several well-know divergence measures found in the literature can be obtained from (1) by taking appropriate choices for h and ϕ. Since divergence measures are not necessarily symmetric functions, we employ a straightforward symmetrization to obtain a distance $D(X,Y)$ from any divergence:

$$D(X,Y) = \frac{d_\phi^h(X,Y)+d_\phi^h(Y,X)}{2} \tag{2}$$

These measures are termed "Stochastic Distances" or h-ϕ distances. As we will see, every h-ϕ distance between members of the same family of distributions can be turned into a statistical goodness-of-fit test.

Test statistic-based metrics. Let us assume that X and Y belong to the same family of distributions. One may define a stochastic distance between X and Y as a function of its maximum likelihood estimators $\widehat{\boldsymbol{\theta}}_X$ and $\widehat{\boldsymbol{\theta}}_Y$. This gives rise to the notation $D(\widehat{\boldsymbol{\theta}}_X,\widehat{\boldsymbol{\theta}}_Y)$ in place of $D(X,Y)$. Salicr et al. [9] introduced the test statistic given by:

$$S_\phi^h(\widehat{\boldsymbol{\theta}}_X,\widehat{\boldsymbol{\theta}}_Y) = \frac{2n_Xn_Y}{n_X+n_Y}\frac{D(\widehat{\boldsymbol{\theta}}_X,\widehat{\boldsymbol{\theta}}_Y)}{h'(0)\phi''(1)}. \tag{3}$$

Under the null hypothesis (i.e., $\boldsymbol{\theta}_X=\boldsymbol{\theta}_Y$) and for $n_X,n_Y\to\infty$, where n_X and n_Y are the number of observations used to estimate $\widehat{\boldsymbol{\theta}}_X$ and $\widehat{\boldsymbol{\theta}}_Y$ so as to ensure that $n_X/(n_X+n_Y)\in(0,1)$, statistic S_ϕ^h converges to a χ^2_M distribution with M degrees of freedom, where M is the dimension of $\boldsymbol{\theta}_X$ and $\boldsymbol{\theta}_Y$. The hypothesis $\boldsymbol{\theta}_X=\boldsymbol{\theta}_Y$ can be then rejected at level α when $\Pr(\chi^2_M > S_\phi^h(\widehat{\boldsymbol{\theta}}_X,\widehat{\boldsymbol{\theta}}_Y))\le\alpha$ [10].

The Bhattacharya distance and test. The more recent use of stochastic distances has supported several Remote Sensing applications, including image classification [12]–[14], speckle filtering [15] and change detection [16]. Bhattacharya, Kullback-Leibler, Hellinger, Harmonic, and Triangular, are examples of such stochastic distances.

We obtain the Bhattacharya distance setting $h(y)=-\log(1-y)$ and $\phi(x)=-\sqrt{x}+(x+1)/2$ in (1), and then (2):

$$D_{\mathrm{B}}(X,Y) = -\log\int_{\xi\in\Xi}\sqrt{f_X(\xi;\theta_X)f_Y(\xi;\theta_Y)}d\xi. \tag{4}$$

If f_X and f_Y are multivariate Gaussian distributions with means μ_X and μ_Y and covariance matrices Σ_X and Σ_Y, one obtains the expression:

$$\begin{aligned} D_{\mathrm{B}}(X,Y) = {} & \frac{1}{8}(\mu_X-\mu_Y)^\dagger\left(\frac{\Sigma_X+\Sigma_Y}{2}\right)^{-1}(\mu_X-\mu_Y) \\ & + \frac{1}{2}\ln\frac{|\Sigma_X+\Sigma_Y|}{\sqrt{|\Sigma_X||\Sigma_Y|}}. \end{aligned} \tag{5}$$

The symbols †, $|\cdot|$ and $(\cdot)^{-1}$ represent the transpose, determinant and inverse matrix, respectively.

By setting (5) into (3), the following test statistic is derived:

$$S_{\phi}^{h}(\widehat{\boldsymbol{\theta}}_X, \widehat{\boldsymbol{\theta}}_Y) = \frac{8n_X n_Y}{n_X + n_Y} D_{\mathrm{B}}(\widehat{\boldsymbol{\theta}}_X, \widehat{\boldsymbol{\theta}}_Y), \tag{6}$$

where $\widehat{\boldsymbol{\theta}}_X = (\widehat{\mu}_X, \widehat{\Sigma}_X)$ and $\widehat{\boldsymbol{\theta}}_Y = (\widehat{\mu}_Y, \widehat{\Sigma}_Y)$.

Statistic (6) is particularly important because it allows the computation of hypothesis tests with simple operations. It also allows quantifying the difference between two distributions in terms of its significance level. Moreover, Eq. (6) induces a theoretically well-defined comparison between distributions, which drives our approach to compare objects/regions from their own probabilistic behaviors, including homogeneity and temporal changes.

B. Support Vector Machines

Let $\mathcal{X}$ be a dataset whose elements $\mathbf{x}_i$ are feature vectors evaluated on a certain position/pixel of the image $\mathcal{I}$. Also, consider that $\mathcal{I}$ is defined on a support $\mathcal{S} \subset \mathbb{N}^2$. Then, classification aims at assigning to each $\mathbf{x}_i \in \mathcal{X}$ a particular class $w_k \in \Omega = \{\omega_1, \omega_2, \ldots, \omega_z\}$ from z possible classes by applying a labeling function $F\colon \mathcal{X} \to \Omega$. Classification methods differ in terms of the formulation of F and the learning strategy used to label data instances in $\mathcal{X}$.

Support Vector Machines (SVMs) have been successfully used in the classification of remotely sensed data. A solid mathematical foundation, simple algorithmic architecture, and high generalization capability are some of the benefits of using SVMs [17]. Furthermore, as reported in Ref. [18], SVMs have achieved similar or even better results compared to other influential classification methods such as maximum likelihood, k-nearest neighbor, fuzzy c-means, neural networks and decision trees.

Inspired on the seminal SVM formulation, diverse variants have been proposed to classifying data, for example, the Laplacian [19], Transductive [20], Context Sensitive [17], [21] and Single-Class [22] SVMs. The latter example, Single-Class SVM, presents an unsupervised approach that relies on quantile estimation for pattern detection in high-dimensional data.

From a set of unlabeled observations, the Single-Class SVM obtains a model which classifies elements as part of such set with a probability ν of false positive or negative occurrence. Formally, a labeling function $F\colon \mathcal{D} \subset \mathcal{X} \to \{+1, -1\}$ can de written, where $+1$ means that the input elements appear in $\mathcal{D}$, and -1 otherwise. The classifier F is given by (cf. Ref. [22]):

$$F(\mathbf{x}) = \operatorname{sgn}\Big(\sum_{i=1}^{m} \alpha_i K(\mathbf{x}, \mathbf{x}_i) - b \Big), \tag{7}$$

where $b = \sum_{j=1}^{m} \alpha_j K(\mathbf{x}_i, \mathbf{x}_j)$ for any $\mathbf{x}_i \in \mathcal{D}$, $i = 1, \ldots, m$, and $K(\cdot, \cdot)$ is a kernel function. Coefficients α_i, $i = 1, \ldots, m$, are computed as the solution of the following optimization problem:

$$\begin{aligned} &\min_{\alpha_1, \ldots, \alpha_m} \textstyle\sum_{i,j=1}^{m} \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j). \\ &s.t. \begin{cases} \alpha_i \in [0, \frac{1}{\nu m}] \\ \sum_{i=1}^{m} \alpha_i = 1. \end{cases} \end{aligned} \tag{8}$$

Notice that the Single-Class SVM is parameterized by $\nu \in [0, 1]$, in addition to the parameters that may be related to the choice of kernel function. For instance, if $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i, \mathbf{x}_j\|^2\right)$, then $\gamma \in (0, \infty)$ should be also handled. See Ref. [23] for a complete discussion about kernel functions.

In our approach, the Single-Class SMV has been adopted to perform the discrimination of changes and non-change events, thus ensuring that the classification will not rely on a global-fixed threshold, but instead through a more flexible and adaptive decision function.

III. Single-class classification of homogeneous unchanged areas

This section introduces our new framework for unsupervised change detection. Figure 1 shows a general pipeline, which is modulated into four main steps: (i) compute a band-wise difference image from a pair of images; (ii) search for homogeneous blocks in the band-wise difference image; (iii) remove homogeneous blocks related to areas wherein probably have occurred a temporal change while keeping the remain blocks as non-changed areas; and (iv) train and perform a single-class classification of band-wise difference image using the information obtained from homogeneous non-change areas, hence obtaining the definitive change/non-change map.

The framework outputs a binary classification where unchanged areas may occur or not. Notice that, in our approach, the lack of non-changed areas suggests a temporal change. Furthermore, the method learns in an unsupervised fashion, as the training process is fully performed using an automatic selection of unchanged area samples.

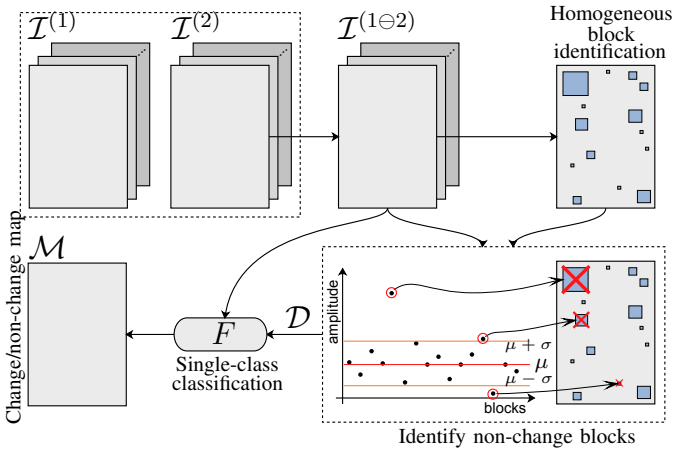


Fig. 1. Pipeline overview of proposed unsupervised change detection method.

Section III-A presents the notation and metrics employed, while the next Sections (III-B to III-D) discuss and formalize each step of the framework outlined in Figure 1.

A. Preliminaries

Let $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$ be images defined on the same support $\mathcal{S} \subset \mathbb{N}^2$, acquired over the same region scene in distinct instants. To preserve the consistency with the notations used in Section II, $\mathcal{X} \subset \mathbb{R}^n$ denotes the feature space, while $\mathbf{x}_i^{(j)} = \mathcal{I}^{(j)}(s_i)$ is the observation at position $s_i \in \mathcal{S}$ in the image $\mathcal{I}^{(j)}$, $j = 1, 2$. The elements of $\mathbf{x}_i^{(j)}$ are the values measured by the sensor, or derived features, over a specific Earth surface position.

Several measures may be applied to highlight the changes between two images, $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$. A commonly chosen measure is the L_2 norm between $\mathbf{x}_i^{(1)}$ and $\mathbf{x}_i^{(2)}$ [5] :

$$\mathcal{I}^{\|1-2\|}(s_i) = \|\mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}\|_2. \tag{9}$$

Another way to identify potential changes between $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$ is to compute the band-wise difference image:

$$\mathcal{I}^{(1\ominus 2)}(s_i) = \mathbf{x}_i^{(1)} - \mathbf{x}_i^{(2)}. \tag{10}$$

Notice that, while $\mathcal{I}^{\|1-2\|}$ returns a scalar as attribute, $\mathcal{I}^{(1\ominus 2)}$ remains in the feature space $\mathcal{X}$.

Region, block and geometric aspects. We introduce here the mathematical entities that support the following formalizations, especially regarding the scalability of the method concerning the size of the objects in the scene. This formalization also serves to define lower-bounds regrading the number

of observations for the statistical estimation. Additionally, this detailed notation also helps avoiding computational implementation issues.

Without loss of generality, assume that the support $\mathcal{S}$ is of the form $\{0,\ldots,\delta_1\}\times\{0,\ldots,\delta_2\}$. In our approach, the *characteristic points* of $\mathcal{S}$ are defined by the pairs $(c_\rho(i),\ell_\rho(j))$ such that $c_\rho(i)=\rho+(2\rho+1)i$, $i=0,1,\ldots,\lfloor(\delta_1-\rho)/(2\rho+1)\rfloor$, and $\ell_\rho(j)=\rho+(2\rho+1)j$, $j=0,1,\ldots,\lfloor(\delta_2-\rho)/(2\rho+1)\rfloor$, $\rho\in\mathbb{N}^*$. The set of characteristic points creates a regular grid on $\mathcal{S}$ whose minimum distance between any two points is always $2\rho+1$. Figure 2 depicts these introduced elements.

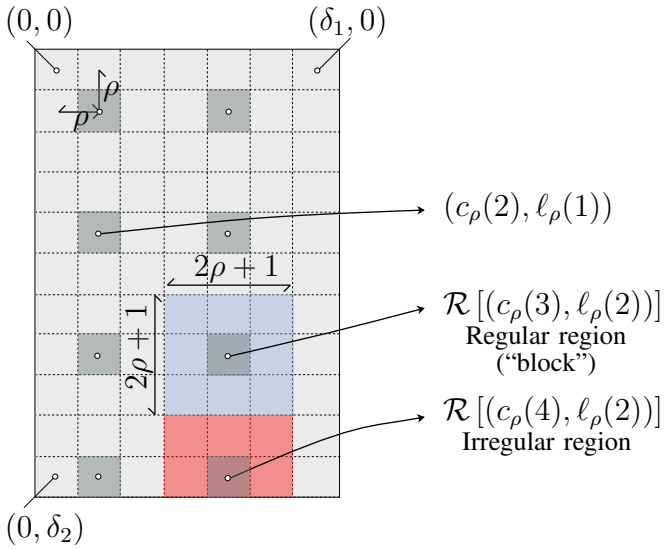


Fig. 2. The elements of $\mathcal{S}$, characteristic points, regions and blocks.

From the pairs $(c_\rho(i),\ell_\rho(j))$, $i=0,1,\ldots,\lfloor(\delta_1-\rho)/(2\rho+1)\rfloor$ and $j=0,1,\ldots,\lfloor(\delta_2-\rho)/(2\rho+1)\rfloor$, the following subsets $\mathcal{R}\left[(c_\rho(i),\ell_\rho(j))\right]\subset\mathcal{S}$ are determined, called here as *regions*:

$$\begin{aligned}\mathcal{R}\left[(c_\rho(i),\ell_\rho(j))\right]=\{&(p,q)\in\mathcal{S}:\\ &c_\rho(i)-\rho\le p\le c_\rho(i)+\rho;\\ &\ell_\rho(j)-\rho\le q\le\ell_\rho(j)+\rho\}\,.\end{aligned}\tag{11}$$

From (11), one can conclude that $(2\rho+1)^2$ is the maximum number of pairs in $\mathcal{R}\left[(c_\rho(i),\ell_\rho(j))\right]$. However, depending where $(c_\rho(i),\ell_\rho(j))$ is located on $\mathcal{S}$, it is possible that p and/or q, such that $c_\rho(i)-\rho\le p\le c_\rho(i)+\rho$ and $\ell_\rho(j)-\rho\le q\le\ell_\rho(j)+\rho$, defines a pair $(p,q)\notin\mathcal{S}$ (i.e., the pair falls out the bounds of $\mathcal{S}$). We, thus, define a *block* within $\mathcal{S}$ for every region $\mathcal{R}\left[(c_\rho(i),\ell_\rho(j))\right]$:

$$\begin{aligned}\mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right]=\{&(p,q)\in\mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right]:\\ &\#\mathcal{R}\left[(c_\rho(i),\ell_\rho(j))\right]=(2\rho+1)^2\}\,.\end{aligned}\tag{12}$$

Homogeneous block characterization. The proposed method relies on checking the homogeneity of non-change areas. However, it is reasonable to admit that assessing the region homogeneity just in terms of data values regardless of its geospatial distribution may lead to wrong conclusions. Herein, we introduce an approach that considers both statistical and geographic data behavior, allowing then more sound decisions about the region's homogeneity.

Let us consider an image $\mathcal{I}$ whose pixels are embedded in the feature space $\mathcal{X}\subset\mathbb{R}^n$. We propose multiple comparisons involving six block-shaped templates to assess the similarity of the feature vectors in $\mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right]$, as illustrated in Figure 3; the definition of block-shaped structures are given in Equations (13) to (18). While $\mathcal{B}_1\left[(c_\rho(i),\ell_\rho(j))\right]$ and $\mathcal{B}_2\left[(c_\rho(i),\ell_\rho(j))\right]$ are vertical structures, $\mathcal{B}_3\left[(c_\rho(i),\ell_\rho(j))\right]$ and $\mathcal{B}_4\left[(c_\rho(i),\ell_\rho(j))\right]$ are horizontal, and $\mathcal{B}_5\left[(c_\rho(i),\ell_\rho(j))\right]$ and $\mathcal{B}_6\left[(c_\rho(i),\ell_\rho(j))\right]$ are halved templates.

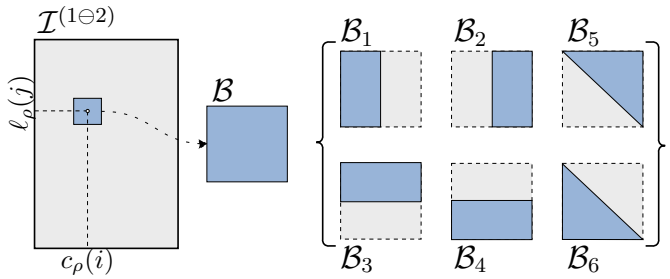


Fig. 3. The six block-shaped structures taken to assess the block homogeneity.

$$\mathcal{B}_1\left[(c_\rho(i),\ell_\rho(j))\right] = \{(p,q) \in \mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right] : c_\rho(i) - \rho \leq p \leq 0; \ell_\rho(j) - \rho \leq q \leq \ell_\rho(j) + \rho\}. \tag{13}$$

$$\mathcal{B}_2\left[(c_\rho(i),\ell_\rho(j))\right] = \{(p,q) \in \mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right] : 0 \leq p \leq c_\rho(i) + \rho; \ell_\rho(j) - \rho \leq q \leq \ell_\rho(j) + \rho\}. \tag{14}$$

$$\mathcal{B}_3\left[(c_\rho(i),\ell_\rho(j))\right] = \{(p,q) \in \mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right] : c_\rho(i) - \rho \leq p \leq c_\rho(i) + \rho; \ell_\rho(j) - \rho \leq q \leq 0\}. \tag{15}$$

$$\mathcal{B}_4\left[(c_\rho(i),\ell_\rho(j))\right] = \{(p,q) \in \mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right] : c_\rho(i) - \rho \leq p \leq c_\rho(i) + \rho; 0 \leq q \leq \ell_\rho(j) + \rho\}. \tag{16}$$

$$\mathcal{B}_5\left[(c_\rho(i),\ell_\rho(j))\right] = \{(p,q) \in \mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right] : c_\rho(i) - \rho \leq p \leq c_\rho(i) + \rho\,; \ell_\rho(j) - \rho \leq q \leq \ell_\rho(j) + \rho; p \geq q\}. \tag{17}$$

$$\mathcal{B}_6\left[(c_\rho(i),\ell_\rho(j))\right] = \{(p,q) \in \mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right] : c_\rho(i) - \rho \leq p \leq c_\rho(i) + \rho\,; \ell_\rho(j) - \rho \leq q \leq \ell_\rho(j) + \rho; p \leq q\}. \tag{18}$$

Given a block-shaped structure as defined above, it is called a *homogeneous block* in $\mathcal{I}$ if the statistical distribution of its feature vectors is similar to the ones observed in each structure $\mathcal{B}_k\left[(c_\rho(i),\ell_\rho(j))\right]$, $k = 1,\ldots,6$. When this condition holds, the notation $\mathcal{H}\left[(c_\rho(i),\ell_\rho(j))\right]$ is used in place of $\mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right]$ to denote a homogeneous block.

B. Identifying homogeneous areas by measuring probability distribution similarity

In our work, the similarity between $\mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right]$ and $\mathcal{B}_k\left[(c_\rho(i),\ell_\rho(j))\right]$ is determined from a statistical test. More specifically, we use the concept of hypothesis testing derived from stochastic distance to compute the similarity, i.e.:

$$\mathcal{H}\left[(c_\rho(i),\ell_\rho(j))\right] = \Big\{(p,q) \in \mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right] : \Pr(\chi^2_M > S^h_\phi(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}}_k)) > \alpha; k = 1,\ldots,6\Big\}, \tag{19}$$

where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_k$ are the estimates of the parameters that index the distributions of the feature vectors in $\mathcal{B}\left[(c_\rho(i),\ell_\rho(j))\right]$ and $\mathcal{B}_k\left[(c_\rho(i),\ell_\rho(j))\right]$, respectively, and $\alpha \in [0,1]$ sets the significance level of the

comparison. Without loss of generality, in this work we assume that the feature vectors follow multivariate Gaussian distributions, and that the Bhattacharya is a convenient distance to assess the similarity between $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_k$. Consequently, the test statistic $S_\phi^h(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}_k)$ is given by (6).

The rationale behind Equation (19) is that the probability $\Pr(\chi^2_M > S_\phi^h(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}_k)) > \alpha$ states that the null hypothesis (i.e., $H_0 : \widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_k$) should not be rejected with significance $1 - \alpha$. As a result, if $\alpha \to 1$, the similarity between $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}_k$ will be high in order to avoid rejecting H_0. Moreover, Equation (19) allows the identification of blocks (i.e., square regions) in $\mathcal{I}$ whose feature vectors exhibits a common statistical behavior.

The use of a hypothesis testing-based similarity metric as (19) provides not only a way of comparing blocks in the feature space, but also a significance value assigned to such comparison. Additionally, this significance is extended to the geometric space (i.e., the support $\mathcal{S}$) when the block structures are individually analyzed.

Dealing with block dimensions and scalability. Once the lengths of the blocks $\mathcal{B}$ are fixed in terms of ρ, they have to meet the scale of objects and targets in $\mathcal{I}$. However, it is usual that $\mathcal{I}$ be composed by elements of different dimensions. In order to cope with this issue, we take $\rho \in \{2^{-t}\rho_{\max} : t = 0, 1, \ldots, k\}$, with $k = \lfloor(\log(\rho_{\max}) - \log(\rho_{\min}))/\log 2\rfloor$ to ensure $(2^{-k}\rho_{\max}) \geq \rho_{\min}$. For simplicity, we will denote $\rho(t) = 2^{-t}\rho_{\max}$ when needed. Scalars $\rho_{\min}$ and $\rho_{\max}$ are determined from the $\mathcal{X}$ and $\mathcal{S}$ dimensions as well as the probability distribution family used to assess the homogeneity of the blocks.

As initially stated, $\mathcal{X}$ is an n-dimensional Euclidean space whose data follow a multivariate Gaussian distribution. Since such a distribution is parameterized by $\boldsymbol{\theta} = (\mu, \Sigma)$, the dimension of $\boldsymbol{\theta}$ is $n + (n(n-1))/2 = (n^2 + 3n)/2$, where n and $n(n-1)/2$ are the dimensions of μ and Σ, respectively. Therefore, it is possible to estimate $\boldsymbol{\theta}$ when more than $(n^2 + 3n)/2$ observations are available.

As previously discussed, since any block has $(2\rho + 1)^2$ elements, cf. (12), and its structures are half of a block (see Eqs. (13)–(18)), we can impose $\rho_{\min}$ such that $(2\rho_{\min} + 1)^2/2 \geq (n^2 + 3n)/2$, and hence $\rho_{\min} = \lceil(\sqrt{n^2 + 3n} - 1)/2\rceil$. Concerning $\rho_{\max}$, it can be upper bounded so as to ensure at least one block on $\mathcal{S}$. Thereby, $(2\rho_{\max} + 1) \leq \delta_{\max}$ leads to $\rho_{\max} = \lfloor(\delta_{\max} - 1)/2\rfloor$, where $\delta_{\max} = \min\{\delta_1, \delta_2\}$, and $(2\rho_{\max} + 1)$ arises from the maximum block side that fits in $\mathcal{S}$. Figure 4 illustrates the relation between the values of ρ as well as the resemblance between $\delta_{\min}$ and $2\rho_{\max} + 1$, where $\delta_{\min} = \delta_1$ in this pictorial example. It should highlight that the homogeneous block identification process uses different values for ρ, adopted in decreasing scale from $\rho_{\max}$ to $\rho_{\min}$. Thus, such a process encompasses a kind of multiscale verification based on the Quad-Tree representation structure [24].

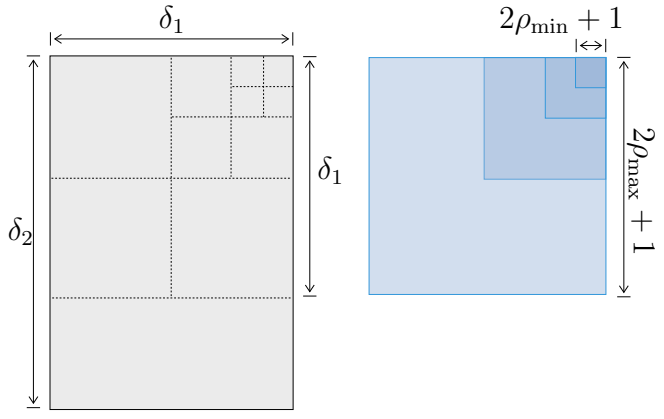


Fig. 4. The geometric rationale behind the ρ values.

Finally, we let $\bar{\mathcal{H}}$ be the set of all positions in $\mathcal{S}$ contained in a homogeneous block:

$$\bar{\mathcal{H}} = \bigcup_{t=0}^{k} \left\{ \bigcup_{i,j=0,0}^{z_1(t),z_2(t)} \mathcal{H}\left[c_{\rho(t)}(i), \ell_{\rho(t)}(j)\right] \right\}, \tag{20}$$

where

$$z_1(t) = \left\lfloor \frac{\delta_1 - \rho(t)}{2\rho_{\min} + 1} \right\rfloor \text{ and } z_2(t) = \left\lfloor \frac{\delta_2 - \rho(t)}{2\rho_{\min} + 1} \right\rfloor$$

define the upper-bound to the coordinates of characteristic points, as discussed at section beginning.

C. Homogeneous blocks on change and non-change areas

Although the identification process of homogeneous blocks proposed in Section III-B has useful for any image, we apply it to the task of detecting temporal changes between $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$. In this study, we seek for a band-wise difference image $\mathcal{I}^{(1\ominus 2)}$ in order to produce an initial representation of changes and non-changes between $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$.

It is expected that the feature vectors in $\mathcal{I}^{(1\ominus 2)}$ assigned to non-change areas lie around a central tendency. Oppositely, the feature vectors assigned to areas of potential changes should be far from such a tendency. Figure 5 illustrates this concept.

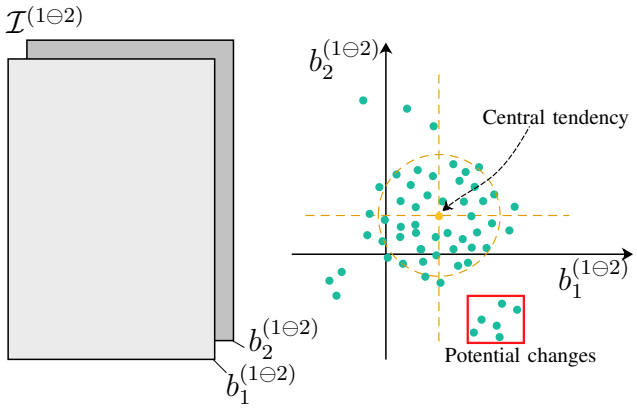


Fig. 5. Common tendency of attribute values on $\mathcal{I}^{(1\ominus 2)}$.

Assuming $\bar{\mathcal{H}}$ from $\mathcal{I}^{(1\ominus 2)}$, a simple way for distinguishing homogeneous blocks between those related to change and non-change areas is defining statistic thresholds based on the tendency and deviation of the attribute vectors of these blocks. To accomplish this task, firstly the amplitude of mean attribute vector is computed from each homogeneous block of $\bar{\mathcal{H}}$. Such values are elements of the following set:

$$\begin{aligned}\mathcal{Q} = \{ &\| \mu \left[\left(c_{\rho(t)}(i), \ell_{\rho(t)}(j) \right) \right] \| : \\ &t = 0, \ldots, k; i = 0, \ldots, z_1(t); j = 0, \ldots, z_2(t) \} \ , \end{aligned} \tag{21}$$

where $\mu \left[\left(c_{\rho(t)}(i), \ell_{\rho(t)}(j) \right) \right]$ is the mean attribute vector of the homogeneous block $\mathcal{H} \left[\left(c_{\rho(t)}(i), \ell_{\rho(t)}(j) \right) \right]$.

Then, we use the mean and standard deviation of $\mathcal{Q}$, denoted as μ_Q and σ_Q, respectively, to establish the interval $[\mu_Q - \sigma_Q, \mu_Q + \sigma_Q]$. Finally, the set of homogeneous blocks related to non-change areas is given by:

$$\begin{aligned}\mathcal{E} = \{ &\mathcal{H} \left[\left(c_{\rho(t)}(i), \ell_{\rho(t)}(j) \right) \right] \in \bar{\mathcal{H}} : \\ &\| \mu \left[\left(c_{\rho(t)}(i), \ell_{\rho(t)}(j) \right) \right] \| \in [\mu_Q - \sigma_Q, \mu_Q + \sigma_Q] \ ; \\ &t = 0, \ldots, k; i = 0, \ldots, z_1(t); j = 0, \ldots, z_2(t) \} \ . \end{aligned} \tag{22}$$

D. One-class classification of non-changed areas

From the formalization given in Sections III-B and III-C, we are now able to perform the detection of non-change areas for a subject pair of multi-temporal images. Notice that the initial identification of non-change homogeneous areas does not have the purpose of building a change/non-change map, but instead allow us to understand the behavior of non-change areas regarding the analyzed pair of images.

Aiming at producing change/non-change mappings from a pair of images, we employ the concept of single-class classification (see Section II-B). From $\mathcal{I}^{(1\ominus 2)}$, which embeds the two instants, $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$ into its representation, we obtain a training set $\mathcal{D} = \left\{\mathbf{x}_s \in \mathcal{X} : \mathcal{I}^{(1\ominus 2)}(s) = \mathbf{x}_s; s = (p, q) \in \bar{\mathcal{H}}\right\}$. This training set is then employed to build a Single-Class SVM classifier $F\colon \mathcal{X} \to \{0, 1\}$ which learns temporal changes from the images. Finally, we produce a change/non-change mapping $\mathcal{M}$ by applying the obtained classifier on each pixel of $\mathcal{I}^{(1\ominus 2)}$. We should stress that although Single-Class SVM uses a training set to model a decision function, such a set is not provided by any human/analyst intervention, but instead by an automatic process as given by Equations (11)-(22).

E. Modular structure

The modular structure of our proposal is noteworthy. The spatial awareness is common to any type of data, with only possible changes in the size of the smallest acceptable sample. The spectral components, which we treated as multivariate Gaussian observations, may be modeled by any suitable distribution. Other examples of possible changes include (i) the band-wise difference image (10); (ii) other block-shaped structures used on the spectral-spatial comparisons, Eqs. (13)–(18); (iii) the adoption of other (stochastic or deterministic) distances to test similarities (19); and (iv) other decision rules to distinguish change and non-change areas.

IV. EXPERIMENTS

In this section, we apply the proposed method on study cases regarding land cover changes occurred in three dams in Brazil. Moreover, we employ different remote sensors for each study case in order to assess the sensitivity of the presented framework regarding distinct scenes. Also, we take well-established unsupervised change detection methods as competitors against the proposed method, and we compare the results in terms of accuracy and computational cost.

The compared change detection methods are the ones previously discussed in Section I, i.e., Change Vector Analysis (CVA) [5], and binary clustering of PCA of spatial divergences (PCA-KM) [6]. Finally, we employ the kappa coefficient [25] as well as F1-Score [26] as accuracy metrics.

Several parameter configurations have been exhaustively tested, and the best-achieved results, taken from these comparisons. The parameters w.r.t. *space-search* for CVA were the Otsu (OT) [7] and Kittler-Illingworth (KI) [8] thresholding options, with Freedman-Diaconis' (FD) [27] and Scott's (SC) [28] rules to determine the size of the histograms bins before the thresholding step. Regarding the PCA-KM method, we consider neighborhood radius $\rho \in \{1, 2, 3\}$ to promote a spectral expansion (each value generates squared neighborhoods of sizes 3, 5 and 7), and $1 \leq P_c \leq \rho$ to define the number of principal components. Concerning the proposed method, the values for α ranged in $\{0.1, 0.2, \ldots, 0.9\}$, while the parameter ν of the Single-Class SVM ranges in $\{0.001, 0.0025, 0.005\}$. Finally, we chose RBF as kernel function, and the parameter γ was also tested in $\{0.001, 0.0025, 0.005\}$.

The experiments were run on a computer with an Intel Xeon processor ($16\,\mathrm{core}$, $2.27\,\mathrm{GHz}$), and $24\,\mathrm{GB}$ of RAM running the Debian Linux version 9 operating system. The platform was IDL (Interactive Data Language), and the LIBSVM [29] implementation was used to run the Single-Class SVM classification. The code of the proposed framework is freely available at https://github.com/rogerionegri/HBSC.

A. Data description

This section presents three study cases regarding the detection of environmental changes occurred in dam regions. Figure 6 depicts the spatial location of these regions.

The first area refers to Mariana's dam region, state of Minas Gerais, Brazil. This area became known since the rupture of its tailings dam on November 5th, 2015. We used a pair of Landsat-8 images (OLI sensor), acquired on September 25th, 2015 (Figure 7(a)) and August 10th, 2016 (Figure 7(b)). These images have 760×600 pixels, $30\,\mathrm{m}$ spatial resolution and three multispectral bands from red to short wave infrared.

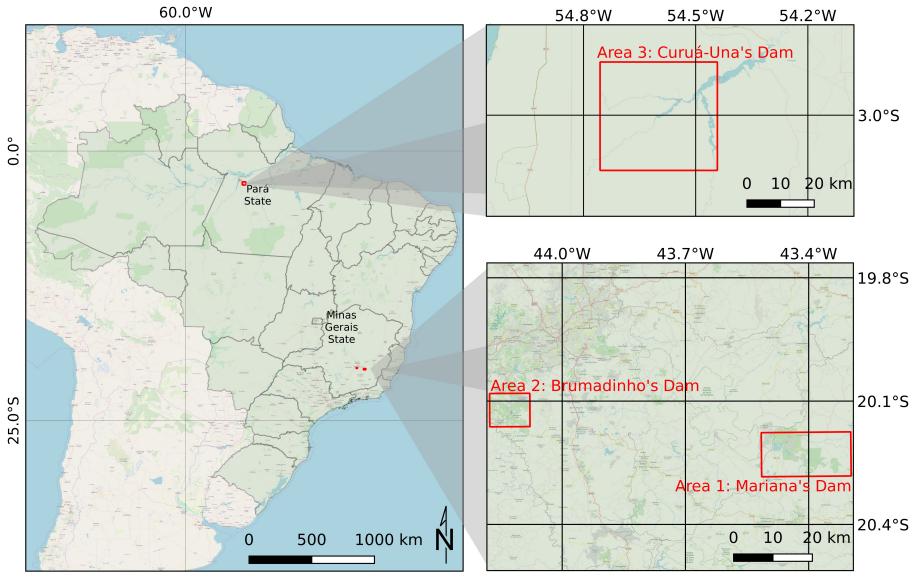


Fig. 6. Study areas location.

Similarly, the second area is also known due to a rupture of a tailings dam in the state of Minas Gerais. Such rupture occurred on January 25th, 2019. We used two images acquired by the Sentinel-2 satellite on January 17th, 2019 (Figure 8(a)) and February 1st, 2019 (Figure 8(b)) to map the environmental changes. The images have 1026×897 pixels, $10\,\mathrm{m}$ spatial resolution and four multispectral bands from blue to near-infrared wavelength.

The third area refers to the Curuá-Una's hydroelectric dam, state of Par, Brazil. Represented by the pair of selected instants, on July 13th, 2007 (Figure 9(a)) and on November 6th, 2010 (Figure 9(b)), this study area is characterized by changes caused by forest regeneration, crop stages, and variation level of Curuá-Una river. The images for this study were acquired by the PALSAR sensor, onboard the ALOS satellite, with 2797×2581 pixels, $12.5\,\mathrm{m}$ spatial resolution, in HH and HV polarizations.

Figures 7(c), 8(c) and 9(c) show the spatial distribution of change and non-change samples for each study area regarding the considered periods: polygons in green and red, respectively. Table I summarizes the ground truth samples sizes. These samples, obtained by visual inspection, were taken to measure the accuracy of the analyzed change detection methods via kappa coefficient and F1-Score.

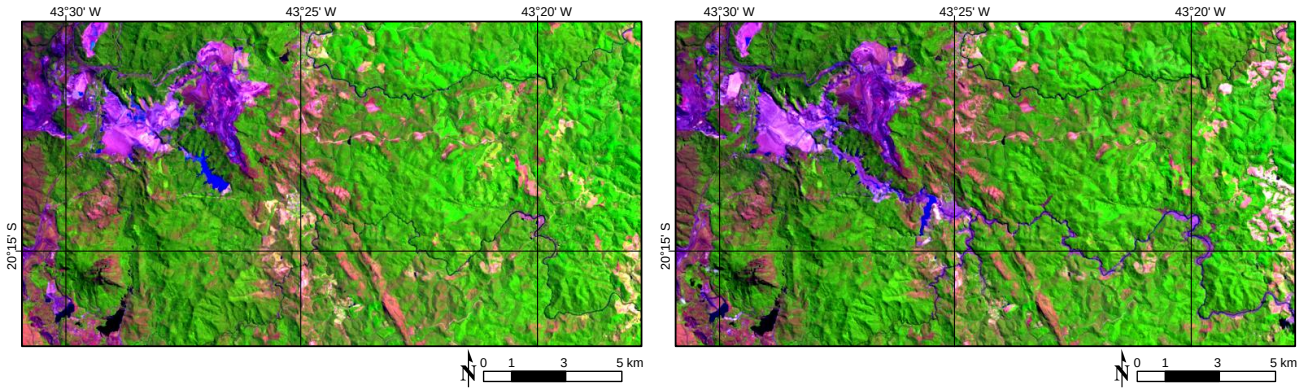
TABLE I
SUMMARY OF CHANGE AND NON-CHANGE GROUND TRUTH SAMPLES.

	Mariana	Brumadinho	Curuá-Una
	Landsat-8 OLI	Sentinel-2	ALOS PALSAR
Change	7712	67876	213588
Non-change	7258	67670	249903

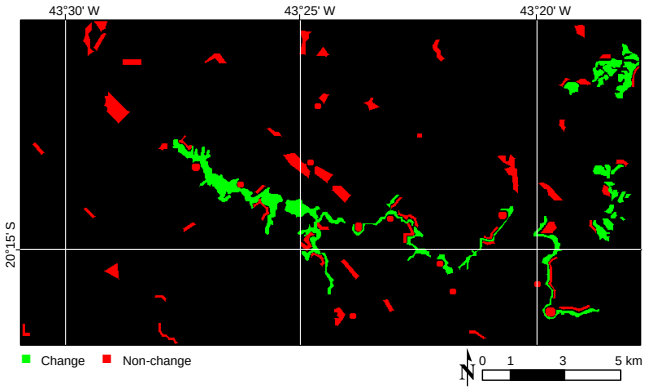
It is worth mentioning that no additional atmospheric correction procedures were carried on the Sentinel-2 images. The OLI images were obtained in "level-2 processing" [30], which includes a built-in atmospheric correction. A 7×7 average low-pass filtering was applied on PALSAR images as a simple approach to reduce the speckle noise and aid temporal changes detection.

B. Results and Discussions

1) Quantitative analysis: Figure 10 shows the performance of the analyzed methods to the kappa coefficient, for each study area when different parameters are tested, as previously discussed in Section IV.



(a) September 25th, 2015 (b) August 10th, 2016



(c) Groud thruth change/non-change samples

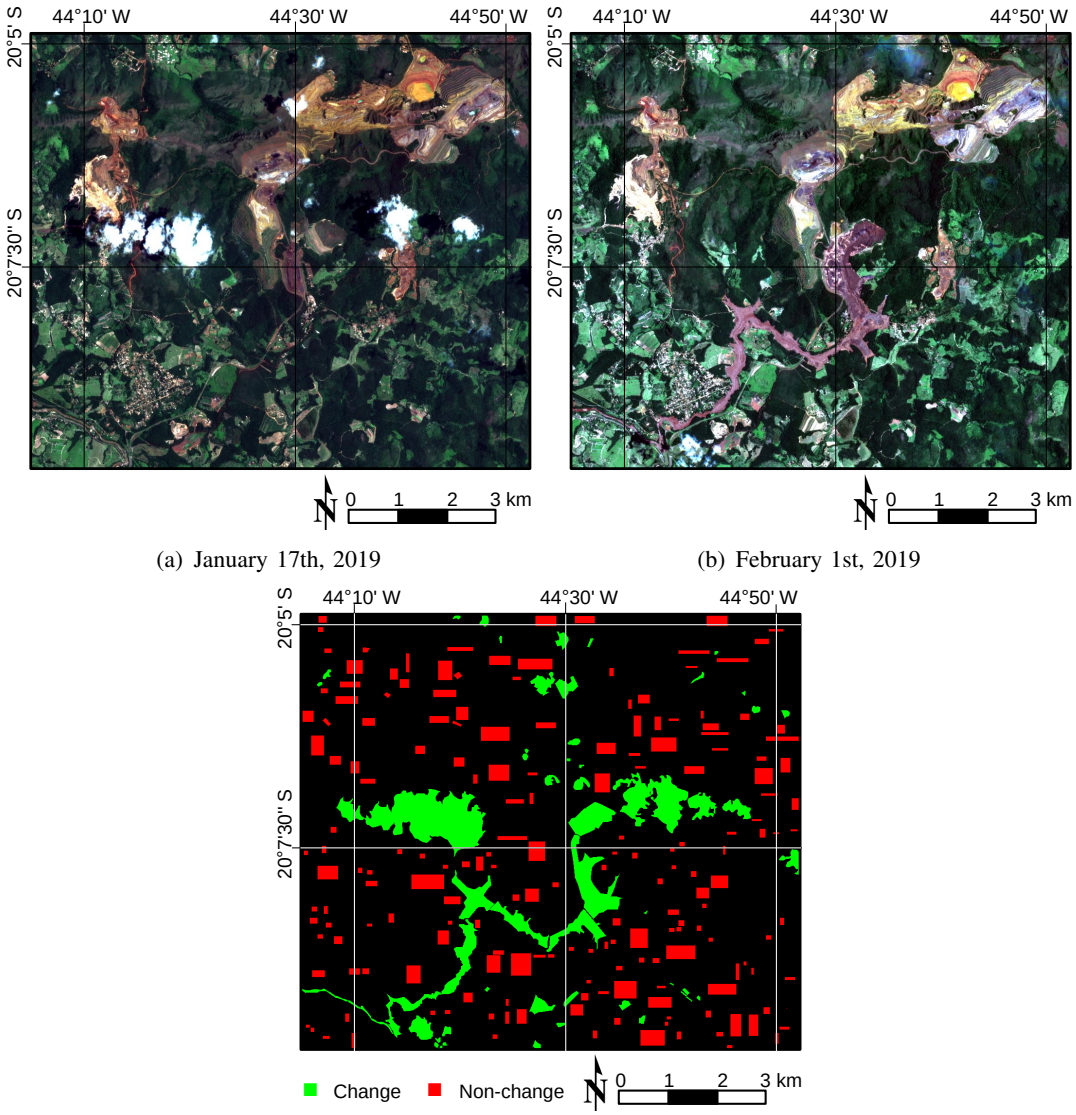
Fig. 7. Mariana's Dam dataset. (a) and (b) Images represented in natural color composition. (c) Change and non-change areas, labeled in green and red, respectively.

The box plot depicts the accuracy ranges and dots the individual accuracy values. One can observe that the proposed method deliveries higher accuracy values when compared to the other methods. Furthermore, our change detection framework is less sensitive to parameter tuning, especially in comparison to CVA. On the other hand, the most competitive results are assigned to CVA in the third study area (Curuá-Unas Dam area), but still numerically lower than our classification results.

TABLE II
BEST KAPPA VALUES (AND RESPECTIVE STANDARD DEVIATIONS) ACHIEVED BY THE ANALYZED METHODS AND RESPECTIVE BEST PARAMETER CONFIGURATION.

Method	Kappa coefficient (standard deviation $\times 10^{-3}$)		
	Mariana's Dam	Brumadinho's Dam	Curuá-Una's Dam
CVA	0.599 (6.35)	0.159 (2.09)	0.842 (0.82)
PCA-KM	0.675 (5.90)	0.102 (2.70)	0.722 (1.01)
Proposed	0.768 (5.16)	0.837 (1.48)	0.811 (0.88)
	Parameter configuration		
	Mariana's Dam	Brumadinho's Dam	Curuá-Una's Dam
CVA	KI, FD	KI, SC	KI, SC
PCA-KM	$(\rho, P_c) = (5, 5)$	$(\rho, P_c) = (3, 3)$	$(\rho, P_c) = (3, 3)$
Proposed	$\alpha = 0.3$	$\alpha = 0.6$	$\alpha = 0.7$

Table II shows the highest kappa values found for each method and the respective parameter configuration. The standard deviation of kappa values is also included in Table II. From the accuracy and deviation



(a) January 17th, 2019

(b) February 1st, 2019

(c) Groud thruth change/non-change samples

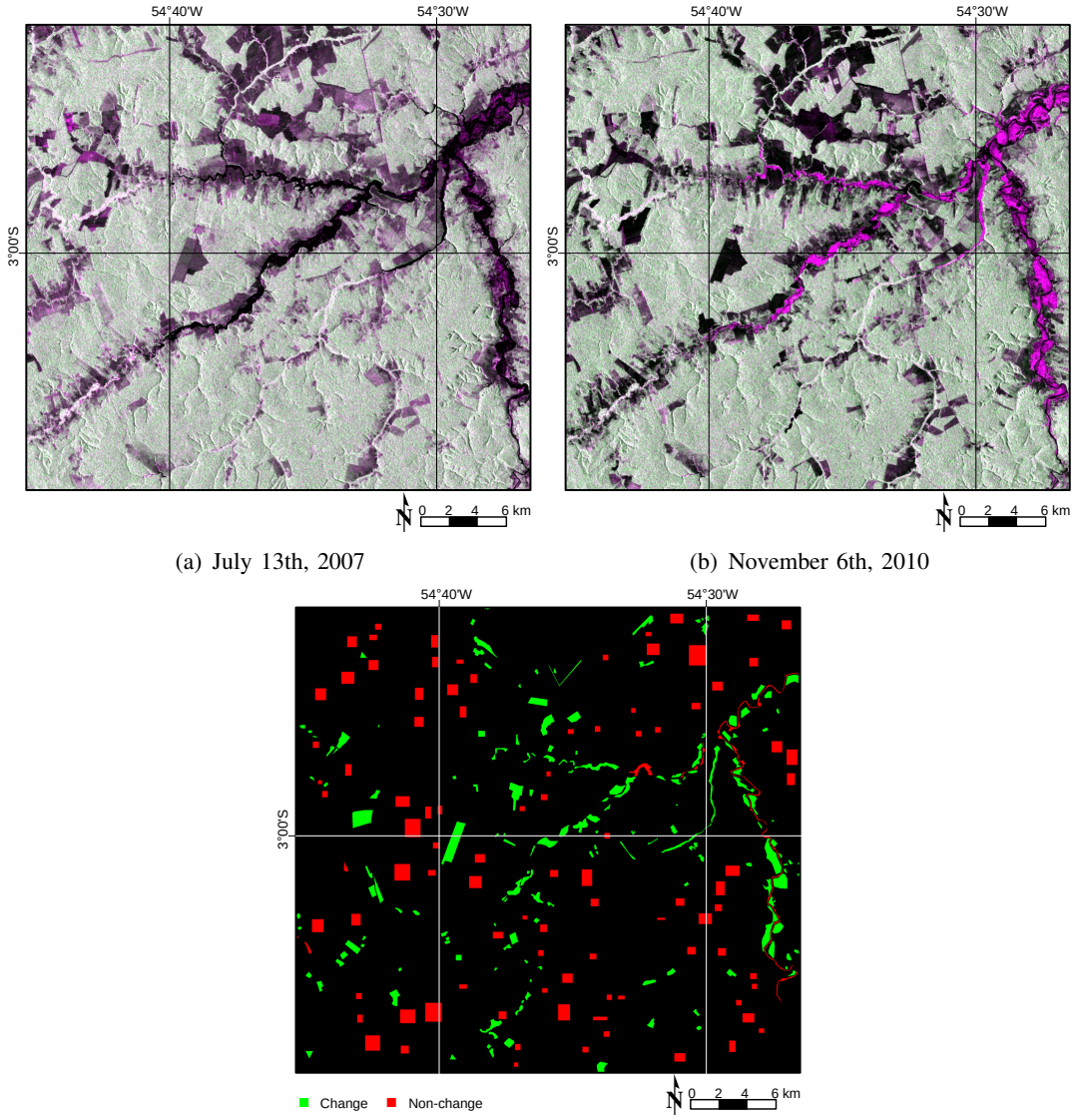
Fig. 8. Brumadinho's Dam dataset. (a) and (b) Images represented in natural color composition. (c) Change and non-change areas, labeled in green and red, respectively.

values listed in Table II, the statistical significance of the results was verified based on a statistic test to compare the values of kappa [25]. It was attested that the proposed framework performs significantly better than other competitors in all cases at least at the $1\,\%$ confidence level.

Additionally, the results are also compared in terms of true/false–positive/negative percentages as well as F1-Score values. Figure 11 is a graphical representation for these values for each method and study area.

In this analysis, while a true-positive (TP) represents the accuracy percentage in the task of identifying a land cover change, a false-positive (FP) quantifies false alarms for land cover changes. In a similar fashion, true-negative (TN) and false-negative (FN) account for the percentage of areas truly and falsely identified as areas where do not occur a land cover change. Finally, the F1-Score summarizes these four percentages into a single value.

From the depicted results, we verify that, independently of the study area, the proposed method produces small FP and FN values. Although CVA and PCA-KM provide highest TP values for Marianas Dam study area, the FP values are also high, i.e., the methods over-estimate land cover changes. This behavior is also observed for Marianas Dam area. When these accuracy indexes are combined and expressed by F1-Scores, the proposed method still reach the highest accuracy scores in all the evaluations.



(a) July 13th, 2007 (b) November 6th, 2010

(c) Groud thruth change/non-change samples

Fig. 9. Curuá-Una's Dam dataset. (a) and (b) Images represented in natural color composition. (c) Change and non-change areas, labeled in green and red, respectively.

2) Computational cost: Regarding the computational cost, the proposed method presents an intermediate run-time between CVA and PCA-KM. As expected for any method, the run-time increases as the input data dimension grows. However, the low-cost assigned to CVA method rises from its simple architecture, which requires an arithmetic operation on the input images followed by histogram thresholding. In contrast, besides the same initial arithmetic operation of CVA, PCA-KM requires the computation of principal components on a high-dimensional space, followed by a data clustering process, hence demanding large amounts of computational resources proportionally to the input images.

Considering the proposed method, the identification of homogeneous blocks under different radius values represents the main computational bottleneck. As our prototype does not take any boosting strategy to accelerate the processing, the use of parallel computing schemes may highly reduce the computational cost, especially for the computation of block distances.

Fig. 12 summarizes these results.

3) Qualitative analysis: Now, considering the change/non-change maps obtained for Marianas Dam area, one can see that radiometric variations of the same target with respect to the considered instants lead

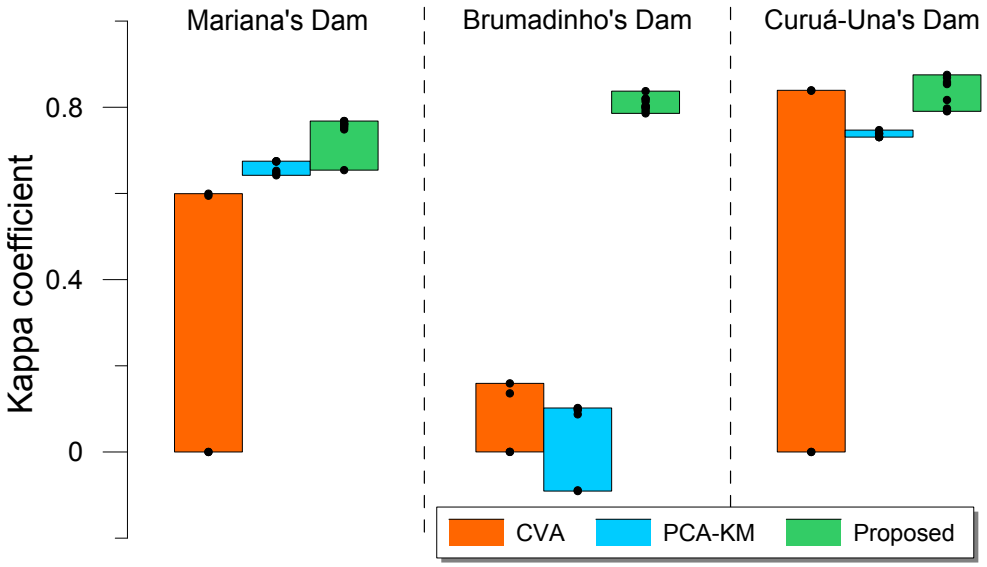


Fig. 10. Accuracy of the analyzed methods for different parameter settings.

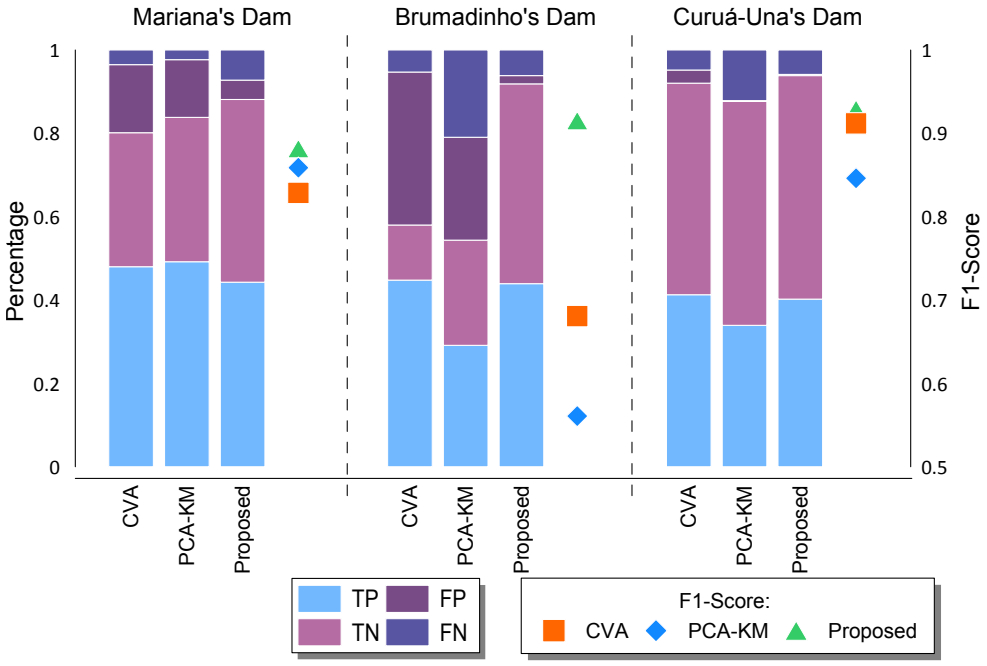


Fig. 11. True/false–positives/negatives proportion and F1-Score of best mapping results for the analyzed methods.

the CVA and PCA-KM methods to detect such variations as false positives (see Figures 13(a) and 13(b)). These variations appear to impair the adequate definition of threshold and clusters for CVA and PCA-KM methods. In contrast, a better delimitation of changed regions is achieved by the proposed method, as shown along the northwest-southeast track, where is located the dam that affected the river after the dams collapse, as well as vegetation suppression on the right side of the study region (see Figure 13(c)).

Concerning the experiment on the Brumadinhos Dam area (Figure 14), the presence of atmospheric factors (i.e., cloud/shadow and haze – see Figures 8(a) and 8(b)) result in inadequate change/non-change maps for CVA and PCA-KM. Reversely, the proposed method was able to detect the region affected by the rupture of Brumadinhos Dam with very fine details beyond could/shadow and haze presence, as small soil and vegetation changes.

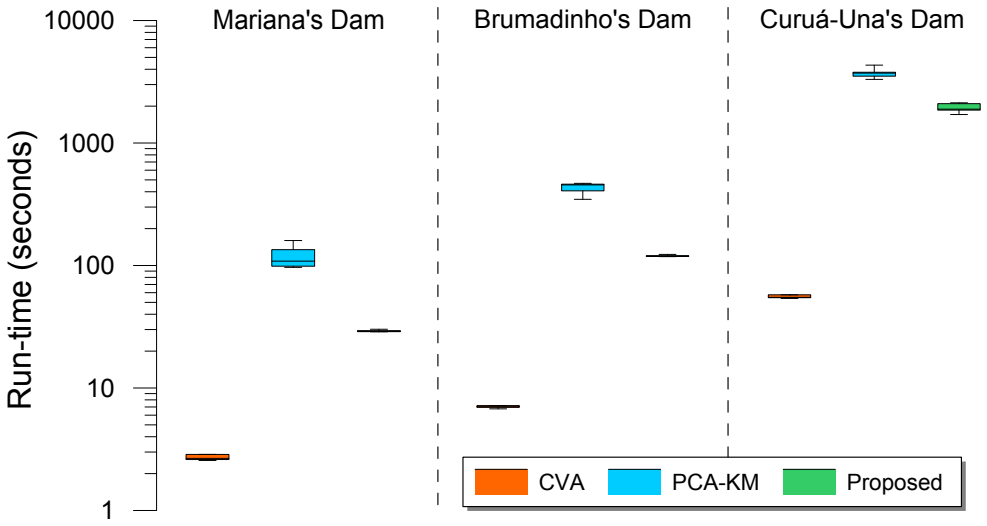


Fig. 12. Run-time of the analyzed methods in logarithmic scale.

In our last experiment (Curuá-Unas Dam region, as shown in Figure 15), the evaluated change detection methods provided consistent results in detecting changes on vegetation areas and, especially, the changes highlighted by the drought on Curuá-Unas river and affluents in 2010. Nevertheless, it is worth observing that CVA is more affected by speckle effects than the other methods.

Similar results were obtained with the PCA-KM approach, however with less influence of speckle noise and more omission rate on detect vegetation changes. The best trade-off between speckle noise influence and small omission rate is achieved by the proposed method, as one can see in Figure 15(c).

V. Conclusions

We proposed a new unsupervised change detection framework which combines stochastic distances and single-class classification concepts. The core idea consists in identifying homogeneous areas where no changes occur on a pair of images acquired in different instants, and to extract the information from these areas to create a hypothesis testing-guided function capable of assessing the similarity among regions.

Three study cases were carried out to assess the accuracy of the current approach, including comparisons against well-established unsupervised change detection methods. In all analyzed study cases, our model overcomes the concurrent methods in terms of quantitative assessments, as well as qualitative results. However, its main drawback lies on the computational cost when larger images are taken.

As future work, we plan to (i) investigate alternative schemes to reduce the computational burden; (ii) analyze other stochastic distances beyond Bathacharyya; (iii) extend our method to deal with multiple images in a time series, by simply taking into account all the band-wise difference images from the sequence; and (iv) apply the technique on polarimetric synthetic aperture radar data using stochastic distances between Wishart models.

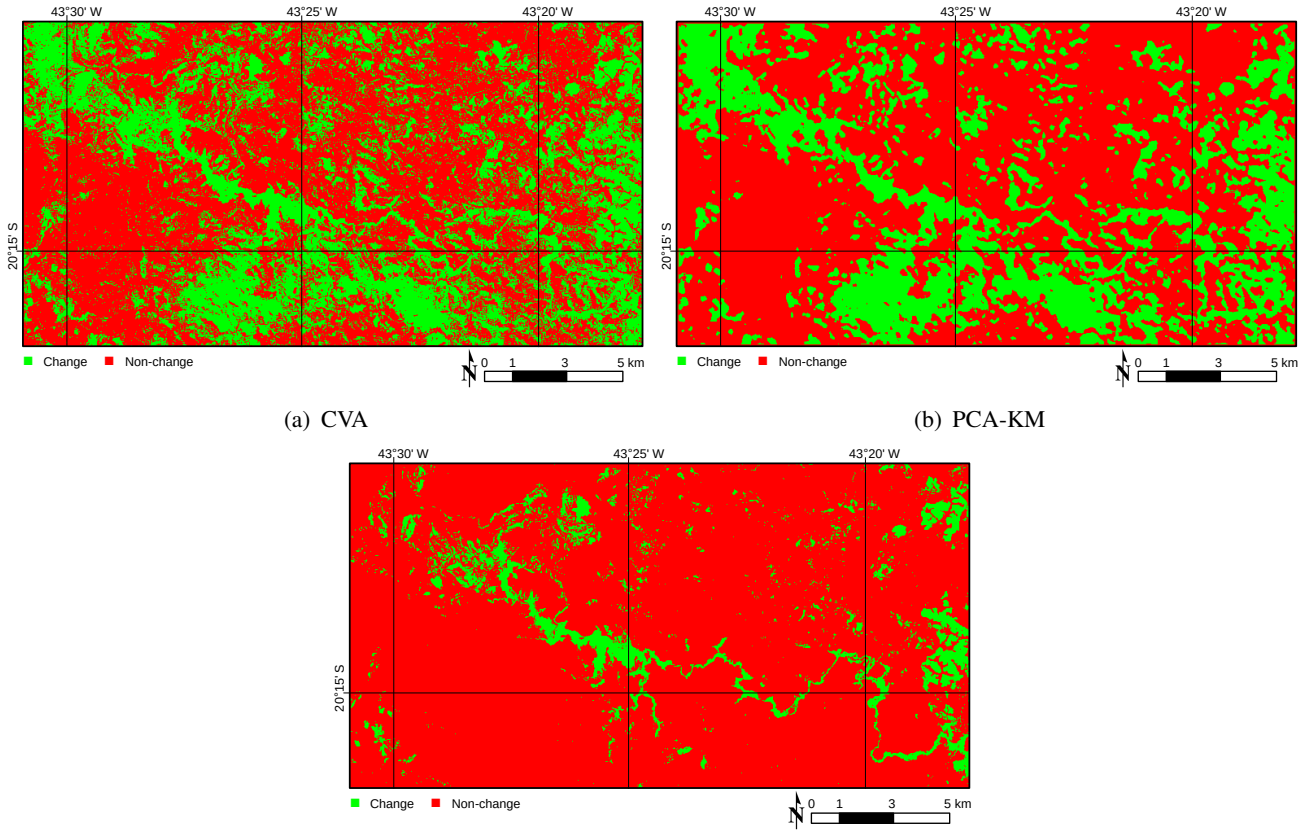
Acknowledgment

The authors acknowledge the support from FAPESP (Grant 2018/01033-3), CNPq (Grants 405364/2018-0, 304515/2013-2), and Fapeal.

References

[1] K. C. Seto, C. E. Woodcock, C. Song, X. Huang, J. Lu, and R. K. Kaufmann, "Monitoring land-use change in the Pearl River Delta using Landsat TM," *International Journal of Remote Sensing*, vol. 23, no. 10, pp. 1985–2004, 2002.

[2] H. S. Salah, S. Ait-Aoudia, A. Rezgui, and S. E. Goldin, "Change detection in urban areas from remote sensing data: a multidimensional classification scheme," *International Journal of Remote Sensing*, vol. 40, no. 17, pp. 6635–6679, 2019.



(a) CVA

(b) PCA-KM

(c) Proposed

Fig. 13. Best change/non-change maps obtained by each analyzed method for Mariana's Dam dataset.

[3] Z. Wang, W. Yao, Q. Tang, L. Liu, P. Xiao, X. Kong, P. Zhang, F. Shi, and Y. Wang, "Continuous change detection of forest/grassland and cropland in the Loess Plateau of China using all available Landsat data," *Remote Sensing*, vol. 10, no. 11, p. 1775, 2018.

[4] C. Wu, B. Du, and L. Zhang, "Slow Feature Analysis for Change Detection in Multispectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2858–2874, 2014.

[5] R. D. Johnson and E. S. Kasischke, "Change vector analysis: a technique for the multispectral monitoring of land cover and condition," *International Journal of Remote Sensing*, vol. 19, no. 3, pp. 411–426, 1998.

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

[7] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[8] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognition*, vol. 19, no. 1, pp. 41 – 47, 1986.

[9] M. Salicru, D. Morales, M. Menendez, and L. Pardo, "On the applications of divergence type measures in testing statistical hypotheses," *Journal of Multivariate Analysis*, vol. 51, no. 2, pp. 372 – 391, 1994.

[10] A. C. Frery, A. D. C. Nascimento, and R. J. Cintra, "Analytic expressions for stochastic distances between relaxed Complex Wishart distributions," *IEEE Trans. Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 1213–1226, 2014.

[11] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, 1948.

[12] W. B. Silva, C. C. Freitas, S. J. S. Sant'Anna, and A. C. Frery, "Classification of segments in PolSAR imagery by minimum stochastic distances between Wishart distributions," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1263–1273, 2013.

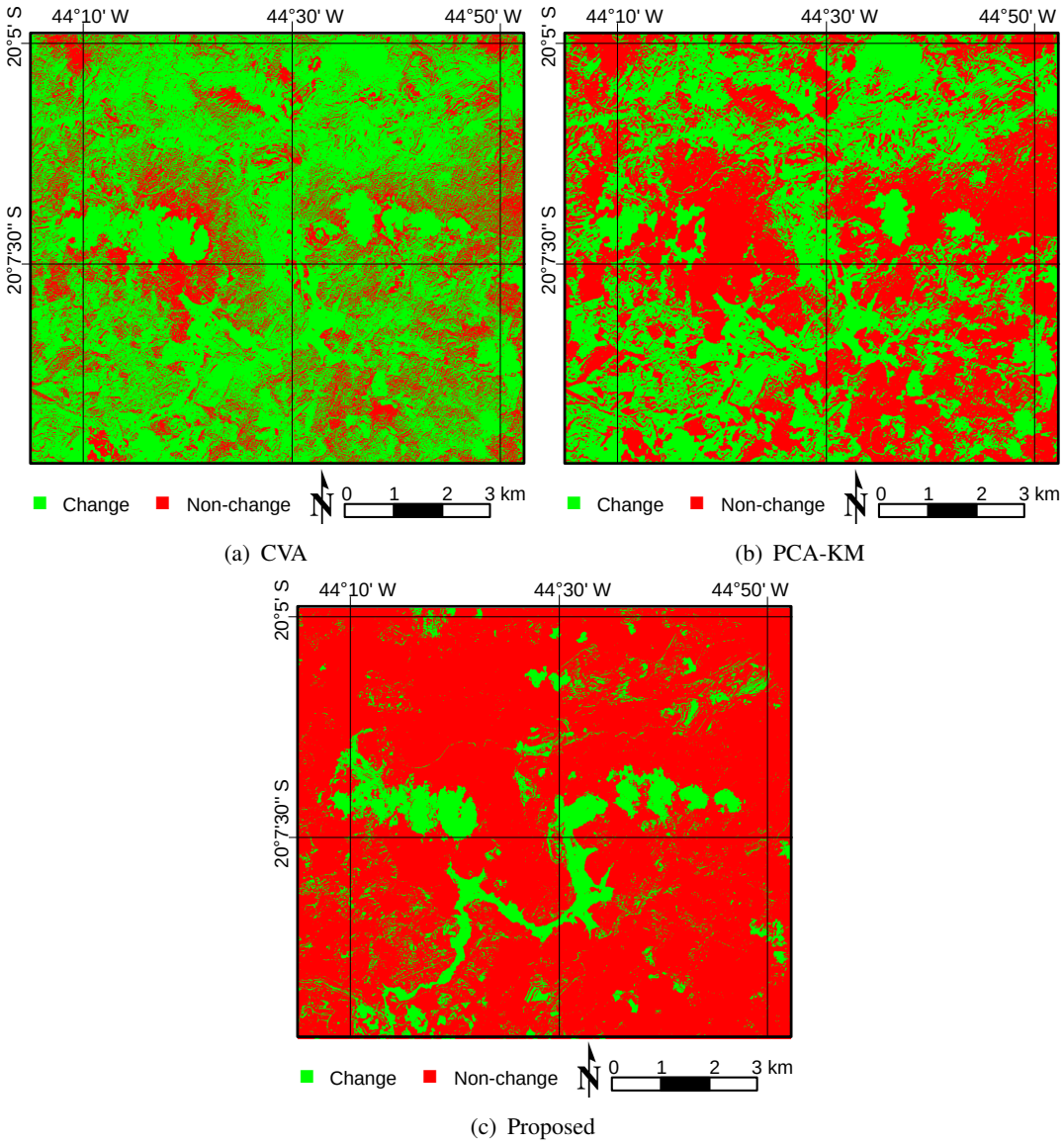
[13] R. G. Negri, L. V. Dutra, S. J. S. Sant'Anna, and D. Lu, "Examining region-based methods for land cover classification using stochastic distances," *International Journal of Remote Sensing*, vol. 37, no. 8, pp. 1902–1921, 2016.

[14] R. G. Negri, A. C. Frery, W. B. Silva, T. S. G. Mendes, and L. V. Dutra, "Region-based classification of PolSAR data using radial basis kernel functions with stochastic distances," *International Journal of Digital Earth*, p. in press, 2019.

[15] L. Torres, S. J. Sant'Anna, C. da Costa Freitas, and A. C. Frery, "Speckle reduction in polarimetric sar imagery with stochastic distances and nonlocal means," *Pattern Recognition*, vol. 47, no. 1, pp. 141–157, Jan. 2014.

[16] A. D. C. Nascimento, A. C. Frery, and R. J. Cintra, "Detecting changes in fully polarimetric SAR imagery with Statistical Information Theory," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1380–1392, March 2019.

[17] L. Bruzzone and C. Persello, "A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2142–2154, 2009.



(a) CVA (b) PCA-KM

(c) Proposed

Fig. 14. Best change/non-change maps obtained by each analyzed method for Brumadinho's Dam dataset.

[18] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in Remote Sensing: a review," *ISPRS Journal of Photogrammetry and Remote Sensing Society*, vol. 66, no. 3, pp. 247–259, 2011.

[19] Y. Gu and K. Feng, "Optimized laplacian SVM with distance metric learning for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1109–1117, June 2013.

[20] Y. Li, Y. Wang, C. Bi, and X. Jiang, "Revisiting transductive support vector machines with margin distribution embedding," *Knowledge-Based Systems*, vol. 152, pp. 200 – 214, 2018.

[21] R. G. Negri, L. V. Dutra, and S. J. S. Sant'Anna, "An innovative support vector machine based method for contextual image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, no. 0, pp. 241 – 248, 2014.

[22] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

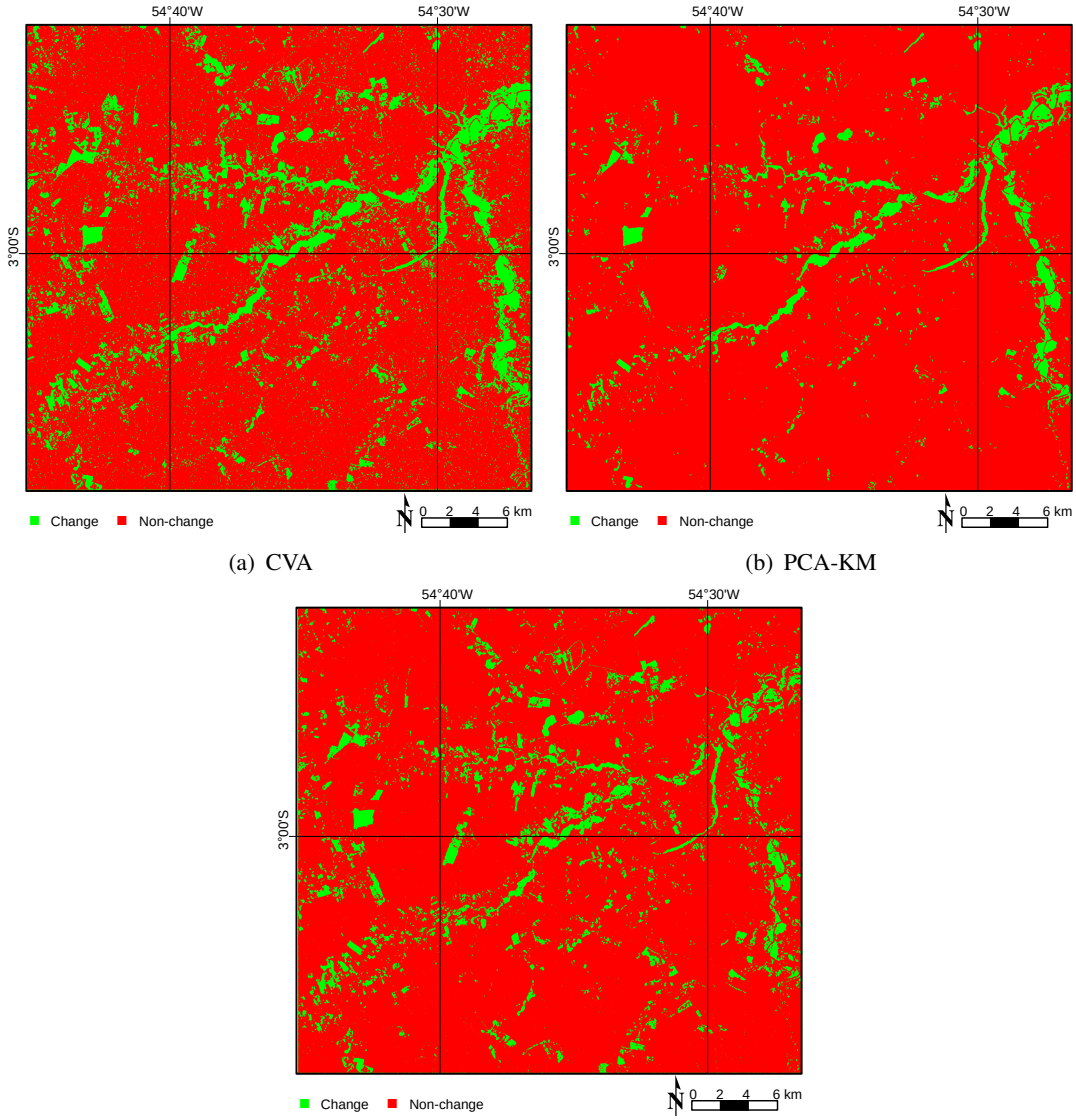
[23] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. New York, NY, USA: Cambridge University Press, 2004.

[24] R. A. Finkel and J. L. Bentley, "Quad trees: a data structure for retrieval on composite keys," *Acta Informatica*, vol. 4, no. 1, pp. 1–9, Mar 1974.

[25] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data*. Boca Raton: CRC Press, 2009.

[26] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-Score, with implication for evaluation," in *Advances in Information Retrieval*, D. E. Losada and J. M. Fernández-Luna, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 345–359.

[27] D. Freedman and P. Diaconis, "On the histogram as a density estimator: L_2 theory," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 57, no. 4, pp. 453–476, Dec 1981.



(a) CVA

(b) PCA-KM

(c) Proposed

Fig. 15. Best change/non-change maps obtained by each analyzed method for Curuá-Unas Dam dataset.

[28] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–610, 12 1979. [Online]. Available: https://doi.org/10.1093/biomet/66.3.605

[29] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[30] E. Vermote, C. Justice, M. Claverie, and B. Franch, "Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product," *Remote Sensing of Environment*, vol. 185, pp. 46 – 56, 2016.

Rogério G. Negri received the graduated degree in Mathematics (2006) at Universidade Estadual Paulista (UNESP), and the M.Sc. (2009) and Ph.D. (2013) in Applied Computation at Instituto Nacional de Pesquisas Espaciais (INPE). His main research interests include pattern recognition and image processing. He is currently Professor at UNESP, São José dos Campos, São Paulo, Brazil.

Alejandro C. Frery (S'92–SM'03) received a B.Sc. degree in Electronic and Electrical Engineering from the Universidad de Mendoza, Mendoza, Argentina. His M.Sc. degree was in Applied Mathematics (Statistics) from the Instituto de Matemática Pura e Aplicada (IMPA, Rio de Janeiro) and his Ph.D. degree was in Applied Computing from the Instituto Nacional de Pesquisas Espaciais (INPE, São José dos Campos, Brazil). He is Professor at the Universidade Federal de Alagoas, and holds a Huashan Scholar position (2019–2021) with the Key Lab of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, Xi'an, China. His research interests are statistical computing and stochastic modeling.

Wallace C. O. Casaca obtained the B.Sc. and Master degree in Pure and Applied Mathematics from Universidade Estadual Paulista (UNESP), Brazil, in 2008 and 2010, respectively. During 2010-2014, he pursued his Ph.D. in Computer Sciences and Applied Mathematics at Universidade de São Paulo(USP), Brazil. As part of his doctoral studies, he also worked as a visiting researcher at Brown University, School of Engineering. He is currently a professor of Computer Science at UNESP. He is also a research associate at the CEPID-FAPESP CeMEAI - Center for Research in Mathematical Sciences Applied to Industry. His research interests include data clustering, computer vision, remote sensing applications, machine learning, data visualization, and numerical methods.

Samara C. Azevedo received the B.S. degree in cartographic engineering from Universidade Estadual Paulista (UNESP), Brazil, in 2011, the M.S. and Ph.D. degree in cartographic sciences from UNESP, in 2014 and 2018, respectively. She is currently a professor of Geomatics at Institute of Natural Resources from Universidade Federal de Itajubá (UNIFEI), Brazil. Her research interests include Remote Sensing, Cartography, Image Processing, GIS and Geodesy.

Maurcio A. Dias received the M.Eng and the Ph.D. degrees from Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil, in 2002 and 2007, respectively. He was a Postdoctoral Researcher (Visiting Researcher in Sabbatical Break) twice: (1) with the Centre for Vision, Speech and Signal Processing (CVSSP) of the Faculty of Electronics and Physical Sciences of the University of Surrey, Guildford, England, in 2018; (2) with the Electronics Department of the Polytechnic School of the University of Alcalá (UAH), Alcalá de Henares, Spain. He is currently an Assistant Professor with UNESP, Presidente Prudente, Brazil. His research interests include digital image processing, digital image analysis, and mathematical morphology.

Erivaldo. A. Silva graduated in Cartographic Engineering from Universidade Estadual Paulista (UNESP), in 1985, Masters in Remote Sensing from Instituto Nacional de Pesquisas Espaciais (INPE), in 1989 and Ph.D in Transport Engineering from Universidade de São Paulo (USP), in 1995. Currently work as Full Professor at UNESP at the Faculty of Science and Technology in the area of Remote Sensing. Has experience in Geosciences with emphasis on Remote Sensing, acting on the following subjects: mathematical morphology, cartography, feature extraction and cartographic update.

Enner H. Alcntara received the BS degree in aquatic sciences from the Universidade Federal do Maranhão (UFMA) in 2005 and Ph.D. degree in remote sensing from the Instituto Nacional de Pesquisas Espaciais (INPE) in 2010. He is currently Associate Professor at Universidade Estadual Paulista (UNESP), São José dos Campos, São Paulo, Brazil.