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

Rogério Galante Negri^{ID}, Alejandro C. Frery^{ID}, *Senior Member, IEEE*, Wallace Casaca^{ID}, Samara Azevedo, Maurício Araújo Dias, Erivaldo Antônio Silva^{ID}, and Enner Herênio Alcântara^{ID}

Abstract—Change detection is a topic of great interest in remote sensing. A good similarity metric to compute the variations among the images is the key to high-quality change detection. However, most existing approaches rely on the fixed threshold values or the user-provided ground truth in order to be effective. The inability to deal 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 characterize and distinguish formally the change and nonchange 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 support-vector-machine-based approach in order to detect the change/nonchange areas. Three study cases using the images acquired with different sensors are given in order to compare the proposed method with other well-known unsupervised methods.

Index Terms—Classification, single-class support vector machine (SVM), stochastic distance, unsupervised change detection.

Manuscript received January 10, 2020; revised May 27, 2020; accepted July 12, 2020. The authors acknowledge the support from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Grants 2013/07375-0 and 2018/01033-3), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Grants 405364/2018-0, 304515/2013-2), and the Fundação de Amparo à Pesquisa do Estado de Alagoas (FAPEAL). (Corresponding author: Rogério Galante Negri.)

Rogério Galante Negri and Enner Herênio Alcântara are with the Department of Environmental Engineering, Sciences and Technology Institute, Universidade Estadual Paulista (UNESP), São José dos Campos 12245-000, Brazil (e-mail: rogerio.negri@unesp.br).

Alejandro C. Frery is with the Laboratório de Computação Científica e Análise Numérica, Universidade Federal de Alagoas, Maceió 57072-900, Brazil, and also with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, Xi'an 710071, China (e-mail: acfrery@laccan.ufal.br).

Wallace Casaca is with the Department of Energy Engineering, Universidade Estadual Paulista (UNESP), Rosana 19274-000, Brazil.

Samara Azevedo is with the Department of Natural Resources, Universidade Federal de Itajubá (UNIFEI), Itajubá 35903-087, Brazil.

Maurício Araújo Dias is with the Department of Mathematics and Computer Science, School of Sciences and Technology, Universidade Estadual Paulista (UNESP), Presidente Prudente 19060-900, Brazil.

Erivaldo Antônio Silva is with the Department of Cartography, School of Sciences and Technology, Universidade Estadual Paulista (UNESP), Presidente Prudente 19060-900, Brazil.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2020.3009483

I. INTRODUCTION

CHANGE detection is an active research field that seeks to track the land-cover differences in the 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 the spatial–temporal changes allows for a better understanding of how ecosystems behave along the time, elucidating the progressive interaction between the natural phenomena and the human activities.

Wu *et al.* [4] reviewed recent approaches devoted to identifying abrupt changes in the remotely sensed images. The authors grouped them into postclassification inspections, image transformations, and arithmetic with image bands and spectral indexes. Another approach commonly adopted to differentiate these methods is the well-known taxonomy of image classification, i.e., an algorithm is either supervised or unsupervised. More specifically, a change-detection method is 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 the 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 principal component analysis (PCA) and *k*-means (KM) clustering. The first method, CVA, comprises three stages: a preprocessing 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 the change and nonchange segments, by using a thresholding scheme such as Otsu [7] or Kittler and Illingworth [8]. Concerning PCA-KM, it relies on a simple but 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 coregistered 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 to these data to

extract the most relevant information from each block. The final representation, usually composed of the first principal components, is handed over to the KM algorithm in order to distinguish the change and nonchange regions.

Multivariate alteration detection (MAD) [9] is a data-transformation technique widely known and used for unsupervised change detection. MAD uses canonical correlation analysis in order to allow robust multitemporal comparisons in a common feature space. Nielsen [10] improved it and proposed the iterative reweighted MAD (IRMAD).

Recently, Wu *et al.* [4] proposed the use of slow feature analysis (SFA) [11] to handle irrelevant variations present in multitemporal data confrontation, a common issue in unsupervised change detection. The resulting method, referred by the authors as USFA, revealed promising results. Du *et al.* [12] integrated the SFA into a deep-learning approach as well as into a Bayesian framework [13] for, respectively, the unsupervised and supervised change detections.

Both IRMAD and USFA aim to transform an input pair of images into a new feature space, where the distinction between the changed and nonchanged areas is favorable. Such a differentiation usually employs thresholding and clustering.

Despite their flexibility, CVA and PCA-KM also have intrinsic weaknesses, which can undermine their accuracy and effectiveness. For instance, the use of input data without an adequate radiometric correction in the CVA may produce very unstable and inconsistent outputs. Furthermore, the use of only a single cutoff value to distinguish between the change and nonchange regions usually leads to the omission and inclusion errors. Regarding PCA-KM, the presence of high-contrast elements like clouds and background regions may impair the clustering process carried out by the KM algorithm. The accommodation of such outliers into the clusters may induce the omission and inclusion errors. Finally, concerning USFA and IRMAD, despite their more robust design compared with CVA and PCA-KM, they may also be trapped by the undesired effects in the thresholding or clustering stages.

Aiming at addressing these drawbacks without penalizing the computational cost, we propose a new unsupervised change-detection technique. Our proposal is grounded in determining a flexible decision rule adjusted accordingly to the response of the nonchange areas. Given an image pair, our approach has two main steps: the identification of the nonchange homogeneous regions and the learning of a decision rule based on the identified nonchange homogeneous regions. Such a rule allows the discrimination between the change and nonchange areas in the input pair. In our approach, we exploit the ideas of stochastic distances [14], [15] to drive the identification of the nonchange homogeneous regions under uncertainty and noise. Finally, the decision rule is computed by a single-class support-vector-machine (SVM) classification to label properly the change areas.

As mentioned, the change-detection process is guided by a decision rule built so as to take advantage of the behavior of the nonchange area. The occurrence of clouds and other nonpermanent artifacts over the scenes are irrelevant changes, and consequently, they are discarded in the decision-rule

modeling. Hence, the results provided by the proposed method tend to be not influenced by such outliers.

Contributions: In summary, the main contributions of this article are as follows.

- 1) A fully unsupervised change-detection method that unifies a robust and statistically well-posed family of stochastic distances with an SVM-based approach.
- 2) A set of geometric structures as well as a statistical criterion especially designed to characterize and identify the homogeneous regions within a sequence of images.
- 3) In contrast to the approaches like 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.
- 4) 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 against CVA, PCA-KM, IRMAD, and USFA, we study three cases of landscape changes that occurred as a consequence of ruptured dams in Brazil. We employed the images acquired from different satellites: Landsat-8, Sentinel-2, and ALOS-1.

This article is organized as follows. Section II presents the underlying concepts of the hypothesis tests from stochastic distances and the single-class classification based on the 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, experimental design, results, and discussions. Finally, Section V summarizes the findings of this article.

II. MATHEMATICAL BACKGROUND

A. Testing Hypothesis From Stochastic Distances

The stochastic distance has its origin on the divergence measures, i.e., based on the Information Theory, as established in the pivotal work by Shannon [16]. A divergence measure quantifies the level of complexity when the difference between the two models needs to be computed.

Salicru *et al.* [14] obtained a generalization of divergence measures, known as h - ϕ divergence family, with interesting statistical properties. Let X and Y be the two random variables with probability density functions $f_X(\mathbf{x}; \theta_X)$ and $f_Y(\mathbf{x}; \theta_Y)$, defined over the same support Ξ , indexed by parameters θ_X and θ_Y . The h - ϕ divergence between X and Y is

$$d_\phi^h(X, Y) = h \left(\int_{\xi \in \Xi} \phi \left(\frac{f_X(\xi; \theta_X)}{f_Y(\xi; \theta_Y)} \right) f_Y(\xi; \theta_Y) d\xi \right). \quad (1)$$

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

Several well-known 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

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

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

1) *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 $\hat{\theta}_X$ and $\hat{\theta}_Y$. This gives rise to the notation $D(\hat{\theta}_X, \hat{\theta}_Y)$ in the place of $D(X, Y)$. Salicru *et al.* [14] introduced the test statistic given by

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

Under the null hypothesis (i.e., $\theta_X = \theta_Y$) and for $n_X, n_Y \rightarrow \infty$, where n_X and n_Y are the number of observations used to estimate $\hat{\theta}_X$ and $\hat{\theta}_Y$ so as to ensure that $n_X/(n_X + n_Y) \in (0, 1)$, statistic S_{ϕ}^h converges to a χ_M^2 distribution with M degrees of freedom, where M is the dimension of θ_X and θ_Y . The hypothesis $\theta_X = \theta_Y$ can be then rejected at level α when $\Pr(\chi_M^2 > S_{\phi}^h(\hat{\theta}_X, \hat{\theta}_Y)) \leq \alpha$ [15].

2) *Bhattacharya Distance and Test*: The more recent use of stochastic distances has supported several remote sensing applications, including image classification [17]–[19], speckle filtering [20], and change detection [21]. Bhattacharya, Kullback–Leibler, Hellinger, Harmonic, and Triangular are the examples of such stochastic distances.

Among them, under the multivariate Gaussian model, the Bhattacharya distance generalizes the Mahalanobis distance, which is the basilar to Fisher’s discriminant analysis. Another advantage of this dissimilarity measure is that it is simple to compute under such a model and that it does not require additional parameters.

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

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

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

$$D_B(X, Y) = \frac{1}{8}(\mu_X - \mu_Y)^{\dagger} \left(\frac{\Sigma_X + \Sigma_Y}{2} \right)^{-1} (\mu_X - \mu_Y) + \frac{1}{2} \ln \frac{|\Sigma_X + \Sigma_Y|}{\sqrt{|\Sigma_X| |\Sigma_Y|}}. \quad (5)$$

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

By setting (5) into (3), we obtain the following test statistic:

$$S_{\phi}^h(\hat{\theta}_X, \hat{\theta}_Y) = \frac{8n_X n_Y}{n_X + n_Y} D_B(\hat{\theta}_X, \hat{\theta}_Y) \quad (6)$$

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

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, (6) induces a theoretically well-defined comparison between the distributions, which drives our approach to compare the objects/regions from

their own probabilistic behaviors, including homogeneity and temporal changes.

B. Support Vector Machines

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

SVMs have been successfully used in the classification of the remotely sensed data. A solid mathematical foundation, simple algorithmic architecture, and high generalization capability are some of the benefits of using SVMs [22]. Furthermore, as reported in [23], SVMs have achieved similar or even better results compared with 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 [24], Transductive [25], context-sensitive [22], [26], and single-class [27] SVMs. The latter example, single-class SVM, presents an unsupervised approach that relies on the quantile estimation for pattern detection in the high-dimensional data.

From a set of unlabeled observations, the single-class SVM obtains a model that classifies elements as part of such a set with a probability ν of the false-positive (F.P) or false-negative (F.N.) occurrence. Formally, a labeling function $F: \mathcal{D} \subset \mathcal{X} \rightarrow \{+1, -1\}$ can be written, where $+1$ means that the input elements appear in $\mathcal{D}$ and -1 otherwise. The classifier F is given by (see [27])

$$F(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^m \alpha_i K(\mathbf{x}, \mathbf{x}_i) - b \right) \quad (7)$$

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

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

Note 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 the kernel function. For instance, if $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$, then $\gamma \in (0, \infty)$ should also be handled. See [28] for a complete discussion on the kernel functions.

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

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

III. SINGLE-CLASS CLASSIFICATION OF HOMOGENEOUS UNCHANGED AREAS

This section introduces our new framework for unsupervised change detection. Fig. 1 shows a general pipeline that is modulated into four main steps: 1) compute a bandwise difference image from a pair of images; 2) search for homogeneous blocks in the bandwise difference image; 3) remove homogeneous blocks related to the areas wherein probably have occurred a temporal change while keeping the remain blocks as nonchanged areas; and 4) train and perform a single-class classification of the bandwise difference image using the information obtained from the homogeneous nonchange areas, hence obtaining the definitive change/nonchange map.

The framework outputs a binary classification where unchanged areas may occur or not. Note that, in our approach, the lack of nonchanged 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.

Section III-A presents the notation and metrics employed, while Sections III-B–III-D discuss and formalize each step of the framework outlined in Fig. 1.

A. Preliminaries

Let $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$ be the 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 the derived features, over a specific Earth surface position.

Several measures may be applied to highlight the changes between the 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]

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

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

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

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

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

1) *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 regarding the number of observations for the statistical estimation. In addition, this detailed notation also helps to avoid the computational implementation issues.

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

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

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

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

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

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

2) *Homogeneous Block Characterization*: The proposed method relies on checking the homogeneity of the nonchange 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, following Zhong *et al.* [29], we consider both statistical and geographic data behaviors, allowing then more sound decisions about the region's homogeneity. To that aim, we will test homogeneity in six patches $B_1, \dots, B_6$ within each block of data B .

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 $\mathcal{B}[(c_\rho(i), \ell_\rho(j))]$, as illustrated in Fig. 3; the definition of the block-shaped structures is given in (13)–(18). While $B_1[(c_\rho(i), \ell_\rho(j))]$ and $B_2[(c_\rho(i), \ell_\rho(j))]$ are the vertical structures, $B_3[(c_\rho(i), \ell_\rho(j))]$ and $B_4[(c_\rho(i), \ell_\rho(j))]$ are horizontal, and $B_5[(c_\rho(i), \ell_\rho(j))]$ and $B_6[(c_\rho(i), \ell_\rho(j))]$ are halved templates

$$\begin{aligned} B_1[(c_\rho(i), \ell_\rho(j))] &= \{(p, q) \in \mathcal{B}[(c_\rho(i), \ell_\rho(j))]: \\ &\quad c_\rho(i) - \rho \leq p \leq 0; \ell_\rho(j) - \rho \leq q \leq \ell_\rho(j) + \rho\} \end{aligned} \quad (13)$$

$$\begin{aligned} B_2[(c_\rho(i), \ell_\rho(j))] &= \{(p, q) \in \mathcal{B}[(c_\rho(i), \ell_\rho(j))]: \\ &\quad 0 \leq p \leq c_\rho(i) + \rho; \ell_\rho(j) - \rho \leq q \leq \ell_\rho(j) + \rho\} \end{aligned} \quad (14)$$

$$\begin{aligned} B_3[(c_\rho(i), \ell_\rho(j))] &= \{(p, q) \in \mathcal{B}[(c_\rho(i), \ell_\rho(j))]: \\ &\quad c_\rho(i) - \rho \leq p \leq c_\rho(i) + \rho; \ell_\rho(j) - \rho \leq q \leq 0\} \end{aligned} \quad (15)$$

$$\begin{aligned} B_4[(c_\rho(i), \ell_\rho(j))] &= \{(p, q) \in \mathcal{B}[(c_\rho(i), \ell_\rho(j))]: \\ &\quad c_\rho(i) - \rho \leq p \leq c_\rho(i) + \rho; 0 \leq q \leq \ell_\rho(j) + \rho\} \end{aligned} \quad (16)$$

$$\begin{aligned} B_5[(c_\rho(i), \ell_\rho(j))] &= \{(p, q) \in \mathcal{B}[(c_\rho(i), \ell_\rho(j))]: \\ &\quad c_\rho(i) - \rho \leq p \leq c_\rho(i) + \rho \\ &\quad \ell_\rho(j) - \rho \leq q \leq \ell_\rho(j) + \rho; \quad p \geq q\} \end{aligned} \quad (17)$$

$$\begin{aligned} B_6[(c_\rho(i), \ell_\rho(j))] &= \{(p, q) \in \mathcal{B}[(c_\rho(i), \ell_\rho(j))]: \\ &\quad c_\rho(i) - \rho \leq p \leq c_\rho(i) + \rho \\ &\quad \ell_\rho(j) - \rho \leq q \leq \ell_\rho(j) + \rho; \quad p \leq q\}. \end{aligned} \quad (18)$$

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 $B_k[(c_\rho(i), \ell_\rho(j))]$, $k = 1, \dots, 6$. When this condition holds, the notation $\mathcal{H}[(c_\rho(i), \ell_\rho(j))]$ is used in the place of $B[(c_\rho(i), \ell_\rho(j))]$ to denote a homogeneous block.

B. Identifying Homogeneous Areas by Measuring Probability Distribution Similarity

In our work, we measure the similarity between $B[(c_\rho(i), \ell_\rho(j))]$ and $B_k[(c_\rho(i), \ell_\rho(j))]$ with a test statistic. More specifically, we use the concept of hypothesis testing derived from stochastic distance to compute the similarity, that is,

$$\begin{aligned} \mathcal{H}[(c_\rho(i), \ell_\rho(j))] &= \{(p, q) \in \mathcal{B}[(c_\rho(i), \ell_\rho(j))]: \\ &\quad \Pr(\chi_M^2 > S_\phi^h(\hat{\theta}, \hat{\theta}_k)) > \alpha; \\ &\quad k = 1, \dots, 6\} \end{aligned} \quad (19)$$

where $\hat{\theta}$ and $\hat{\theta}_k$ are the estimates of the parameters that index the distributions of the feature vectors in $\mathcal{B}[(c_\rho(i), \ell_\rho(j))]$ and $B_k[(c_\rho(i), \ell_\rho(j))]$, 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 $\hat{\theta}$ and $\hat{\theta}_k$. Consequently, the test statistic $S_\phi^h(\hat{\theta}, \hat{\theta}_k)$ is given by (6).

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

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

1) *Dealing With Block Dimensions and Scalability*: Once the lengths of the blocks 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 the elements of different dimensions. To cope with this issue, we take $\rho \in \{2^{-t}\rho_{\max} : t = 0, 1, \dots, k\}$, 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 $\theta = (\mu, \Sigma)$, the dimension of θ 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 θ when more than $(n^2 + 3n)/2$ observations are available.

As previously discussed, since any block has $(2\rho + 1)^2$ elements [see (12)] and its structures are half of a

Fig. 4. Geometric rationale behind the ρ values.

block [see (13)–(18)], we can impose $\rho_{\min}$ such that $(2\rho_{\min} + 1)^2/2 \geq (n^2 + 3n)/2$, and hence, $\rho_{\min} = \lceil ((n^2 + 3n)^{1/2} - 1)/2 \rceil$. 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 $\mathcal{S}$. Fig. 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 [30].

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

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

where

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

define the upper bound to the coordinates of characteristic points, as discussed in the beginning of this section.

C. Homogeneous Blocks on Change and Nonchange Areas

Although the identification process of homogeneous blocks proposed in Section III-B has been 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 bandwise difference image $\mathcal{I}^{(1\ominus 2)}$ in order to produce an initial representation of the changes and nonchanges between $\mathcal{I}^{(1)}$ and $\mathcal{I}^{(2)}$.

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

Assuming $\tilde{\mathcal{H}}$ from $\mathcal{I}^{(1\ominus 2)}$, a simple way for distinguishing homogeneous blocks from those related to the change and nonchange areas is defining statistic thresholds based on the

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

tendency and deviation of the attribute vectors of these blocks. To accomplish this task, first, the amplitude of the mean attribute vector is computed from each homogeneous block of $\tilde{\mathcal{H}}$. Such values are the elements of the following set:

$$\mathcal{Q} = \{ \|\mu[(c_{\rho(t)}(i), \ell_{\rho(t)}(j))]\| : t = 0, \dots, k; i = 0, \dots, z_1(t); j = 0, \dots, z_2(t) \} \quad (21)$$

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

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 the nonchange areas is given by

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

This interval of length $2\sigma_Q$ around the mean is short enough to reject suspicious observations, but may be changed by the user if needed.

D. One-Class Classification of Nonchanged Areas

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

Aiming at producing the change/nonchange 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 $\mathcal{D} = \{\mathbf{x}_s \in \mathcal{X} : \mathcal{I}^{(1\ominus 2)}(s) = \mathbf{x}_s; s = (p, q) \in \tilde{\mathcal{H}}\}$. This training set is then employed to build a single-class SVM classifier $F : \mathcal{X} \rightarrow \{0, 1\}$ that learns temporal changes from the images. Finally, we produce a change/nonchange mapping $\mathcal{M}$ by applying the obtained classifier on each pixel of $\mathcal{I}^{(1\ominus 2)}$. We should stress that although the 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 given by (11)–(22).

E. Modular Structure

The modular structure of our proposal is noteworthy. Spatial awareness is common to any type of data, with only possible changes in the size of the smallest acceptable sample. The spectral components that we treated as multivariate Gaussian observations may be modeled by any suitable distribution. Other examples of possible changes include: 1) the bandwise difference image (10); 2) other block-shaped structures used on the spectral–spatial comparisons [see (13)–(18)]; 3) the adoption of other (stochastic or deterministic) distances to test the similarities (19); and 4) other decision rules to distinguish the change and nonchange areas.

IV. EXPERIMENTS

In this section, we apply the proposed method on the study cases regarding the land-cover changes that 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. In addition, we consider well-established unsupervised change-detection methods as the competitors against the proposed method, and we compare the results in terms of accuracy and computational cost.

We compare our proposal with the methods discussed in Section I, i.e., CVA [5], a binary clustering of the PCA of spatial divergences (PCA-KM) [6], and two data-transformation-based methods: IRMAD [10] and USFA [4]. We employ the kappa coefficient [31] as an accuracy metric, as well the as true-positive (T.P.) / F.P. and true-negative (T.N.) / F.N. rates to analyze and compare the performance of the methods.

In order to make a fair comparison, we tested several parameter configurations for each technique and selected those that produced the best results. The *space-search* parameters used in the CVA, USFA, and IRMAD methods were Kittler–Illingworth (KI) [8] thresholding, with both Freedman–Diaconis' (FD) [32] and Scott's (SC) [33] rules to determine the size of the histogram bins before the thresholding step. In addition, we adopted the KM algorithm to cluster the outputs from these methods into the change and nonchange segments using the same approach conducted by PCA-KM. The use of the KM and thresholding schemes for change detection through the data-transformation-based methods is applied in the same sense of [34]. Regarding the PCA-KM method, we consider the neighborhood radii $\rho \in \{1, 2, 3\}$ 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 $\{0.1, 0.2, \dots, 0.9\}$, while the parameter ν of the single-class SVM ranges in $\{0.001, 0.0005, 0.001, 0.005\}$. Finally, we chose RBF as the kernel function, and the parameter γ was also tested in $\{0.001, 0.005, 0.01\}$. The standard grid-search with tenfold cross-validation procedures was used in our approach to tune ν and γ .

Fig. 6. Location of the study areas.

We used a computer with an Intel Xeon processor (16 core, 2.27 GHz), and 24 GB of RAM running the Debian Linux version 9 operating system. The main programming platform was interactive data language (IDL). We used LIBSVM [35] for single-class SVM classification and the MATLAB implementation of IRMAD available at <http://www.imm.dtu.dk/~alan/software.html>. 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 the environmental changes that occurred in the dam regions. Fig. 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 5, 2015. We used a pair of Landsat-8 images (OLI sensor), acquired on September 25, 2015 [Fig. 7(a)] and August 10, 2016 [Fig. 7(b)]. These images have 760×600 pixels, 30-m spatial resolution, and three multispectral bands from red to short-wave infrared.

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

The third area refers to the Curuá-Una's hydroelectric dam, State of Pará, Brazil. Represented by the pair of selected instants, on July 13, 2007 [Fig. 9(a)] and on November 6, 2010 [Fig. 9(b)], this study area is characterized by the 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 and 12.5-m spatial resolution, in the H.H. and H.V. polarizations.

Figs. 7(c), 8(c) and 9(c) show the spatial distribution of the change and nonchange 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

Fig. 7. Mariana's Dam data set. (a) and (b) Images represented in natural color composition. (c) Change and nonchange areas, labeled in green and red, respectively. (a) September 25, 2015. (b) August 10, 2016. (c) Ground-truth change/nonchange samples.

TABLE I
SUMMARY OF CHANGE AND NONCHANGE GROUND-TRUTH SAMPLES

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

to measure the accuracy of the analyzed change-detection methods.

It is worth mentioning that no additional atmospheric correction procedures were carried on the Sentinel-2 images.

Fig. 8. Brumadinho's Dam data set. (a) and (b) Images represented in natural color composition. (c) Change and nonchange areas, labeled in green and red, respectively. (a) January 17, 2019. (b) February 1, 2019. (c) Ground-truth change/nonchange samples.

The OLI images were obtained in “level-2 processing” [36], which includes a built-in atmospheric correction. A 7×7 average low-pass filtering was applied on the PALSAR images

Fig. 9. Curuá-Una's Dam data set. (a) and (b) Images represented in color composition through the HH-HV-HH polarization sequence assigned to red-green-blue color channels. (c) Change and nonchange areas, labeled in green and red, respectively. (a) July 13, 2007. (b) November 6, 2010. (c) Ground-truth change/nonchange samples.

as a simple approach to reduce the speckle noise and aid in temporal change detection.

Fig. 10. Effects of parameters on the proposed method over the analyzed data sets. (a) Mariana's Dam. (b) Brumadinho's Dam. (c) Curuá-Una's Dam.

B. Results and Discussion

1) *Sensitivity Analysis:* We assess the overall accuracy achieved by the decision rule in the identification of the homogeneous regions to measure the effect of the parametrization on our approach. The parameter configurations are in the set established at the beginning of Section IV.

Fig. 10 shows the profile common to all the study areas, with differences mostly in the accuracy levels. Regarding the single-class SVM parameters, for all data sets (i.e., study areas), the best accuracies arise with $\gamma = 0.01$. There is a lesser influence of ν , for which values below 0.01 tend to produce similar results.

Concerning parameter α , there is usually a drop in performance for values greater than 0.6. Such values restrict the learning to highly homogeneous blocks and, thus, reduce the generalization ability for change detection.

2) *Homogeneous Block Identification:* Fig. 11 illustrates the homogeneous block-identification process. Parameter α was set to 0.5 for Mariana's Dam area [Fig. 11(a)] and 0.6 for both Brumadinho [Fig. 11(b)] and Curuá-Una's [Fig. 11(c)] areas, respectively. Once the homogeneous blocks are identified, their average amplitudes are computed [see (21)] in order to obtain the nonchange blocks [see (22)]. The interval $[\mu_Q - \sigma_Q, \mu_Q + \sigma_Q]$ ranges for each study area lead to the nonchange blocks, shown in blue and the magenta blocks, which are the ones excluded as suspicious observations. From the detached subregion in Fig. 11, we observe that all blocks fall within the regions with a similar inner behavior. Moreover, note that there are suspicious blocks that have been excluded

Fig. 11. Identification of the homogeneous blocks in different study areas. Color compositions using the bandwise difference images and similar red-green-blue sequence in Figs. 7–9, respectively. (a) Mariana's Dam. (b) Brumadinho's Dam. (c) Curuá-Una's Dam.

even as a result of a slight amplitude variation that falls beyond the limits of $[\mu_Q - \sigma_Q, \mu_Q + \sigma_Q]$.

Fig. 12. Accuracy of the analyzed methods for different parameter settings.

TABLE II

BEST KAPPA VALUES AND STANDARD DEVIATIONS ACHIEVED BY THE ANALYZED METHODS AND BEST PARAMETER CONFIGURATIONS

Method	Kappa coefficient (standard deviation $\times 10^{-3}$)		
	Mariana's Dam	Brumadinho's Dam	Curuá-Una's Dam
CVA	0.602 (6.35)	0.159 (2.09)	0.839 (0.80)
PCA-KM	0.675 (5.90)	0.102 (2.70)	0.747 (0.96)
IRMAD	0.644 (6.12)	0.901 (1.65)	0.798 (0.90)
USFA	0.682 (5.94)	0.860 (1.33)	0.833 (0.82)
Proposed	0.768 (5.16)	0.837 (1.48)	0.875 (0.71)
	Parameter configuration		
	Mariana's Dam	Brumadinho's Dam	Curuá-Una's Dam
CVA	<i>FD</i>	<i>SC</i>	<i>SC</i>
PCA-KM	$(\rho, P_c) = (7, 3)$	$(\rho, P_c) = (7, 5)$	$(\rho, P_c) = (3, 3)$
IRMAD	<i>SC</i>	<i>FD</i>	<i>SC</i>
USFA	<i>FD</i>	<i>FD</i>	<i>SC</i>
Proposed	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.6$

3) *Quantitative Analysis*: Fig. 12 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. The boxplot depicts the accuracy ranges and dots the individual accuracy values.

We verify that the proposed method delivers the highest accuracy scores in the first and third study areas. For instance, the worst parameter configuration is comparable with the best settings for the other techniques in the first study area. Concerning the second study area, IRMAD provides the best kappa when it is tuned with the best parameter setting, followed by USAF and the proposed method, which delivers very similar scores for the best case. Although IRMAD performs better in Brumadinho's area, it is still surpassed by the proposed method and by USAF in Mariana's and Curuá-Una's areas. It is worth mentioning that the proposed approach is more stable than the others, showing that it is less sensitive to the parameter tuning; such a trait is exclusive of our proposal and of CVA.

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 values listed in Table II, the statistical significance of the results was verified based on a statistic test to compare the values of kappa [31]. It was

Fig. 13. Proportion of true/false-positives/negatives of the best mapping results for the analyzed methods.

Fig. 14. Runtime of the analyzed methods in logarithmic scale.

attested that the proposed framework performs significantly better than the other competitors in almost all cases, at least at the 1% confidence level. The only exception, as previously discussed, occurs in Brumadinho's Dam area.

We also compare the results in terms of true/false-positive/negative percentages. In such an analysis, while a T.P. represents the accuracy percentage in the task of identifying a land-cover change, a F.P. quantifies the false alarms for the land-cover changes. In a similar fashion, T.N. and F.N. account for the percentage of areas truly and falsely identified as areas where a land-cover change does not occur.

Fig. 13 illustrates these values for each method and study area. We see that, regardless of the study area, the proposed method produces small F.P. and F.N. values. In particular, for F.P., our proposal is more precise in identifying the changes. Another consequence of low F.P. rates is the definition of more regularized (i.e., homogeneous) change-detection maps. Regarding Mariana's Dam, all competitors present similar behavior, but with F.P. error rates much higher than the one achieved by the proposed method. Concerning Curuá-Una's Dam, CVA, IRMAD, and USFA provide similar error rates, but the proposed method shows a better tradeoff among the T.P., TN, and lowest error rate sum. Now, focusing on Brumadinho's Dam, we observe that there is a slightly lower overall accuracy (i.e., the summation of the T.P. and T.N. rates)

(a)

(b)

(c)

Fig. 15. Three best change/nonchange maps obtained by the analyzed methods for Mariana's Dam data set (a) PCA-KM. (b) USFA. (c) Proposed.

of the proposed method when than the IRMAD and USFA. However, the almost-nil F.P. and F.N. values given by IRMAD justifies its higher kappa on this data set (Fig. 12 and Table II).

4) *Computational Cost*: Fig. 14 illustrates, in logarithmic scale, the boxplots of the time each procedure requires for each data set. The proposed method presents an intermediate runtime between PCA-KM and the other methods. As expected for any classification method, the runtime increases with the input data size. CVA has the lowest time cost because of its simplicity. The data transformation required by both IRMAD and USFA combines elegant, straightforward algebraic procedures with low computational costs. CVA, IRMAD, and USFA only incur long runtimes when the strategy to distinguish

Fig. 16. Three best change/nonchange maps obtained by the analyzed methods for Brumadinho's Dam data set. (a) IRMAD. (b) USFA. (c) Proposed.

change from nonchange areas is the KM algorithm. In contrast, PCA-KM requires the computation of principal components on a high-dimensional space, followed by a data clustering, hence demanding large amounts of computational resources

Fig. 17. Three best change/nonchange maps obtained by the analyzed methods for Curuá-Una's Dam data set. (a) CVA. (b) IRMAD. (c) Proposed.

proportionally to the input images. Finally, in 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 significantly reduce the computational cost, especially for the computation of the block distances.

5) *Qualitative Analysis*: A visual inspection of the change/nonchange maps for Mariana's Dam area reveals that temporal radiometric variations of the same target make the PCA-KM and USFA methods falsely detect them as changes [see Fig. 15(a) and (b)]. For example, it is unlikely that there was a change in the region between the longitudes 43°25'W to 43°20'W in only 46 days. Such variations appear to prevent an adequate definition of the threshold and clusters for USFA and PCA-KM. In contrast, the proposed method achieves a better delimitation of the changed regions, as shown along the NW-SE track, where the dam whose collapse affected the river, as well as vegetation suppression on the right-hand side of the study region [see Fig. 15(c)].

Concerning Brumadinho's Dam area (Fig. 16), not only the IRMAD and USFA techniques but also the proposed method provide precise mappings of the region affected by the rupture of the dam with very fine details beyond cloud/shadow and haze presence, as small soil and vegetation changes. Atmospheric factors, i.e., cloud/shadow and haze, exert a strong influence on CVA and PCA-KM [see Fig. 8(a) and (b)], thus producing inadequate change/nonchange maps.

Finally, in our last experiment (Curuá-Una's Dam region; see Fig. 17), the evaluated change-detection methods provide consistent results in detecting changes on the vegetation areas and, especially, the changes highlighted by the drought on Curuá-Una's river and effluents in 2010. Nevertheless, it is worth observing that the proposed method is less affected by the speckle effects compared with the other methods, as shown in Fig. 17(c).

V. CONCLUSION

We proposed a new unsupervised change-detection framework that 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 in extracting 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 the well-established unsupervised change-detection methods. The proposed method overcame the other ones or at least provided results closer to the best method in terms of quantitative assessments, as well as qualitative results.

A limitation of our approach, which is shared by CVA, PCA-KM, USFA, and IRMAD, is that it assumes that the changes represent a relatively small area with respect to the whole data set.

As future work, we plan to: 1) investigate alternative schemes to reduce the computational burden; 2) analyze other stochastic distances beyond Bhattacharyya; 3) extend our method to deal with multiple images in a time series, by simply considering all the bandwise difference images from the

sequence; and 4) apply the technique on polarimetric synthetic aperture radar data using stochastic distances between the Wishart models.

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," *Int. J. Remote Sens.*, vol. 23, no. 10, pp. 1985–2004, Jan. 2002.
- [2] H. Si Salah, S. Ait-Aoudia, A. Rezgui, and S. E. Goldin, "Change detection in urban areas from remote sensing data: A multidimensional classification scheme," *Int. J. Remote Sens.*, vol. 40, no. 17, pp. 6635–6679, Sep. 2019.
- [3] Z. Wang *et al.*, "Continuous change detection of forest/grassland and cropland in the loess plateau of China using all available Landsat data," *Remote Sens.*, vol. 10, no. 11, p. 1775, Nov. 2018.
- [4] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, 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," *Int. J. Remote Sens.*, vol. 19, no. 3, pp. 411–426, Jan. 1998.
- [6] T. Celik, "Unsupervised change detection in satellite images using principal component analysis and *k*-means clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 772–776, Oct. 2009.
- [7] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
- [8] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognit.*, vol. 19, no. 1, pp. 41–47, Jan. 1986.
- [9] A. A. Nielsen, K. Conradsen, and J. J. Simpson, "Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies," *Remote Sens. Environ.*, vol. 64, no. 1, pp. 1–19, Apr. 1998.
- [10] A. A. Nielsen, "The regularized iteratively reweighted MAD method for change detection in Multi- and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, Feb. 2007.
- [11] L. Wiskott and T. J. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Comput.*, vol. 14, no. 4, pp. 715–770, Apr. 2002.
- [12] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.
- [13] C. Wu, B. Du, X. Cui, and L. Zhang, "A post-classification change detection method based on iterative slow feature analysis and Bayesian soft fusion," *Remote Sens. Environ.*, vol. 199, pp. 241–255, Sep. 2017.
- [14] M. Salicrú, D. Morales, M. L. Menéndez, and L. Pardo, "On the applications of divergence type measures in testing statistical hypotheses," *J. Multivariate Anal.*, vol. 51, no. 2, pp. 372–391, Nov. 1994.
- [15] A. C. Frery, A. D. C. Nascimento, and R. J. Cintra, "Analytic expressions for stochastic distances between relaxed complex Wishart distributions," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1213–1226, Feb. 2014.
- [16] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, 1948.
- [17] 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 J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1263–1273, Jun. 2013.
- [18] 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," *Int. J. Remote Sens.*, vol. 37, no. 8, pp. 1902–1921, Apr. 2016.
- [19] 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," *Int. J. Digit. Earth*, vol. 12, no. 6, pp. 699–719, Jun. 2019.
- [20] L. Torres, S. J. S. Sant'Anna, C. da Costa Freitas, and A. C. Frery, "Speckle reduction in polarimetric SAR imagery with stochastic distances and nonlocal means," *Pattern Recognit.*, vol. 47, no. 1, pp. 141–157, Jan. 2014.
- [21] A. D. C. Nascimento, A. C. Frery, and R. J. Cintra, "Detecting changes in fully polarimetric SAR imagery with statistical information theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1380–1392, Mar. 2019.

- [22] L. Bruzzone and C. Persello, "A novel context-sensitive semisupervised SVM classifier robust to mislabeled training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2142–2154, Jul. 2009.
- [23] G. Mountrakis, J. Im, and C. Ogoale, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, May 2011.
- [24] Y. Gu and K. Feng, "Optimized Laplacian SVM with distance metric learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1109–1117, Jun. 2013.
- [25] Y. Li, Y. Wang, C. Bi, and X. Jiang, "Revisiting transductive support vector machines with margin distribution embedding," *Knowl.-Based Syst.*, vol. 152, pp. 200–214, Jul. 2018.
- [26] R. G. Negri, L. V. Dutra, and S. J. S. Sant'Anna, "An innovative support vector machine based method for contextual image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 241–248, Jan. 2014.
- [27] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.
- [28] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. New York, NY, USA: Cambridge Univ. Press, 2004.
- [29] H. Zhong, J. Zhang, and G. Liu, "Robust polarimetric SAR despeckling based on nonlocal means and distributed lee filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4198–4210, Jul. 2014, doi: [10.1109/TGRS.2013.2280278](https://doi.org/10.1109/TGRS.2013.2280278).
- [30] R. A. Finkel and J. L. Bentley, "Quad trees a data structure for retrieval on composite keys," *Acta Inf.*, vol. 4, no. 1, pp. 1–9, 1974.
- [31] R. G. Congalton and K. Green, *Assessing Accuracy Remotely Sensed Data*. Boca Raton, MA, USA: CRC Press, 2009.
- [32] D. Freedman and P. Diaconis, "On the histogram as a density estimator: L_2 theory," *Probab. Theory Rel. Fields*, vol. 57, pp. 453–476, Dec. 1981.
- [33] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–610, 1979.
- [34] B. Du, Y. Wang, C. Wu, and L. Zhang, "Unsupervised scene change detection via latent Dirichlet allocation and multivariate alteration detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4676–4689, Dec. 2018.
- [35] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 1–27, Feb. 2011. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [36] E. Vermote, C. Justice, M. Claverie, and B. Franch, "Preliminary analysis of the performance of the landsat 8/OLI land surface reflectance product," *Remote Sens. Environ.*, vol. 185, pp. 46–56, Nov. 2016.

Rogério Galante Negri received the B.Ed. degree in mathematics from the Universidade Estadual Paulista (UNESP), Presidente Prudente, Brazil, in 2006, and the M.Sc. and Ph.D. degrees in applied computing from the Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil, in 2009 and 2013, respectively.

He is currently a Professor with UNESP, São José dos Campos. His main research interests include pattern recognition and image processing.

Alejandro C. Frery (Senior Member, IEEE) received the B.Sc. degree in electronic and electrical engineering from the Universidad de Mendoza, Mendoza, Argentina, in 1985, the M.Sc. degree in applied mathematics (statistics) from the Instituto de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, Brazil, in 1990, and the Ph.D. degree in applied computing from the Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil, in 1993.

He is currently a Professor with 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 Casaca received the B.Sc. and M.Sc. degrees in pure and applied mathematics from Universidade Estadual Paulista (UNESP), São José do Rio Preto, Brazil, in 2008 and 2010, respectively, and the Ph.D. degree in computer sciences and applied mathematics from the Universidade de São Paulo (USP), São Carlos, Brazil, in 2014.

As a part of his doctoral studies, he also worked as a Visiting Researcher at the School of Engineering, Brown University, Providence, RI, USA. He is currently a Professor with UNESP, Rosana, Brazil. He is also a Research Associate at the CEPID-FAPESP Center for Research in Mathematical Sciences Applied to Industry (CeMEAI). His research interests include data clustering, computer vision, remote sensing applications, machine learning, data visualization, and numerical methods.

Samara Azevedo received the B.S. degree in cartographic engineering, and the M.Sc. and Ph.D. degrees in cartographic sciences from the Universidade Estadual Paulista (UNESP), Presidente Prudente, Brazil, in 2011, 2014, and 2018, respectively.

She is currently a Professor of geomatics with the Institute of Natural Resources, Universidade Federal de Itajubá (UNIFEI), Itajubá, Brazil. Her research interests include remote sensing, cartography, image processing, GIS, and geodesy.

Maurício Araújo Dias received the M.Sc. and Ph.D. degrees from the Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil, in 2002 and 2007, respectively.

He was a Post-Doctoral Researcher (Visiting Researcher in Sabbatical Break) with Electronics Department, Polytechnic School, University of Alcalá (UAH), Alcalá de Henares, Spain, in 2008, and the Centre for Vision, Speech and Signal Processing (CVSSP), Faculty of Electronics and Physical Sciences, University of Surrey, Guildford, U.K., in 2018. He is currently a Professor with the Universidade Estadual Paulista (UNESP), Presidente Prudente, Brazil. His research interests include digital image processing, digital image analysis, and mathematical morphology.

Erivaldo Antônio Silva received the B.S. degree in cartographic engineering from the Universidade Estadual Paulista (UNESP), Presidente Prudente, Brazil, in 1985, the M.Sc. degree in remote sensing from the Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil, in 1989, and the Ph.D. degree in transport engineering from Universidade de São Paulo (USP), São Paulo, Brazil, in 1995.

He is currently a Full Professor in remote sensing with the Faculty of Science and Technology, UNESP. He has experience in geosciences with an emphasis on remote sensing, acting on the following subjects: mathematical morphology, cartography, feature extraction, and cartographic update.

Enner Herênio Alcântara received the B.S. degree in aquatic sciences from the Universidade Federal do Maranhão (UFMA), São Luís, Brazil, in 2005, and the Ph.D. degree in remote sensing from the Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, Brazil, in 2010.

He is currently an Associate Professor with UNESP, São José dos Campos.