Open access • Journal Article • DOI:10.1109/TGRS.2017.2765348

Detecting Changes Between Optical Images of Different Spatial and Spectral Resolutions: A Fusion-Based Approach — Source link

Vinicius Ferraris, Nicolas Dobigeon, Qi Wei, Marie Chabert

Institutions: University of Toulouse, Duke University

Published on: 01 Mar 2018 - IEEE Transactions on Geoscience and Remote Sensing (IEEE)

Topics: Image resolution, Image fusion, Hyperspectral imaging, Multispectral image and Change detection

Detecting Changes Between Optical Images of Different Spatial and Spectral Resolutions: A Fusion-Based Approach

Vinicius Ferraris, *Student Member, IEEE*, Nicolas Dobigeon, *Senior Member, IEEE*, Qi Wei, *Member, IEEE*, and Marie Chabert, *Member, IEEE*

***Abstract*— Change detection (CD) is one of the most challenging issues when analyzing remotely sensed images. Comparing several multidate images acquired through the same kind of sensor is the most common scenario. Conversely, designing robust, flexible, and scalable algorithms for CD becomes even more challenging when the images have been acquired by two different kinds of sensors. This situation arises in the case of emergency under critical constraints. This paper presents, to the best of our knowledge, the first strategy to deal with optical images characterized by dissimilar spatial and spectral resolutions. Typical considered scenarios include CD between panchromatic, multispectral, and hyperspectral images. The proposed strategy consists of a three-step procedure: 1) inferring a high spatial and spectral resolution image by fusion of the two observed images characterized one by a low spatial resolution and the other by a low spectral resolution; 2) predicting two images with, respectively, the same spatial and spectral resolutions as the observed images by the degradation of the fused one; and 3) implementing a decision rule to each pair of observed and predicted images characterized by the same spatial and spectral resolutions to identify changes. To quantitatively assess the performance of the method, an experimental protocol is specifically designed, relying on synthetic yet physically plausible change rules applied to real images. The accuracy of the proposed framework is finally illustrated on real images.**

***Index Terms*— Change detection (CD), different resolution, hyperspectral (HS) imagery, image fusion, multispectral (MS) imagery.**

I. Introduction

CHANGE detection (CD) is one of the most investigated issues in remote sensing [1]–[4]. As the name suggests, it consists in analyzing two or more multidate (i.e., acquired at different time instants) images of the same scene to detect potential changes. Applications are diverse, from natural disaster monitoring to long-term tracking of urban and forest growth. Optical images have been the most studied remote sensing data for CD. They are generally well suited to map land-cover types at large scales [5]. Multiband optical sensors use a spectral window with a particular width, often called spectral resolution, to sample part of the electromagnetic spectrum of the incoming light [6], [7]. *Panchromatic* (PAN) images are characterized by a low spectral resolution, sensing part of the electromagnetic spectrum with a single and generally wide spectral window. Conversely, *multispectral* (MS) and *hyperspectral* (HS) images have smaller spectral windows, allowing part of the spectrum to be sensed with higher precision. Multiband optical imaging has become a very common modality of remote sensing, boosted by the advent of new finer spectral sensors [8]. One of the major advantages of multiband images is the possibility of detecting changes by exploiting not only the spatial but also the spectral information. There is no specific convention regarding the numbers of bands that characterize MS and HS images. Yet, MS images generally consist of a dozen of spectral bands while HS may have a lot more than a hundred. In complement to spectral resolution taxonomy, one may describe multiband images in terms of their spatial resolution measured by the ground sampling interval (GSI), e.g., the distance, on the ground, between the center of two adjacent pixels [5], [7], [9]. Informally, it represents the smallest object that can be resolved up to a specific pixel size. Then, the higher the resolution, the smaller the recognizable details on the ground: a *high-resolution* (HR) image has smaller GSI and finer details than a *low-resolution* (LR) one, where only coarse features are observable.

Each image sensor is designed based on a particular signal-to-noise ratio (SNR). The reflected incoming light must be of sufficient energy to guarantee a sufficient SNR, and thus a proper acquisition. To increase the energy level of the arriving signal, either the instantaneous field of view or the spectral window width must be increased. However, these solutions are mutually exclusive. In other words, optical sensors suffer from an intrinsic energy tradeoff that limits the possibility of acquiring images of both high spatial and high spectral resolutions [9], [10]. This tradeoff prevents any simultaneous

This work was supported in part by the Coordenação de Aperfeiçoamento de Ensino Superior, Brazil, and in part by EU FP7 through the ERANETMED JC-WATER Program, MapInvPlnt under Project ANR-15-NMED-0002-02.*(Corresponding author: Vinicius Ferraris.)*

V. Ferraris and M. Chabert are with the IRIT/INP-ENSEEIHT, University of Toulouse, Toulouse, France (e-mail: vinicius.ferraris@enseeiht.fr; marie.chabert@enseeiht.fr).

N. Dobigeon is with the IRIT/INP-ENSEEIHT, University of Toulouse, Toulouse, France, and also with the Institut Universitaire de France, Paris, France (e-mail: nicolas.dobigeon@enseeiht.fr).

Q. Wei is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA (e-mail: qi.wei@duke.edu).

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

Digital Object Identifier 10.1109/TGRS.2017.2765348

decrease in both the GSI and the spectral window width. HS, MS, and PAN images are, in this order, characterized by an increasing spatial resolution and a decreasing spectral resolution.

Because of the common assumption of an additive Gaussian noise model for passive optical images, the most common CD techniques designed for single-band optical images are based on image differencing [1]–[4]. When dealing with multiband images, classical CD differencing methods have been adapted for such data through spectral change vectors (SCVs) [2], [11], [12] or transform analysis [13], [14]. Besides, most CD techniques assume that the multidate images have been acquired by sensors of the same type [4] with similar acquisition characteristics in terms of, e.g., angle-of-view, resolutions, or noise model [15], [16]. Nevertheless, in some specific scenarios, for instance, consecutive to natural disasters, such a constraint may not be ensured, e.g., images compatible with previously acquired ones may not be available in an acceptable timeframe. Such disadvantageous emergency situations yet require fast, flexible, and accurate methods able to handle images acquired by sensors of different kinds [17]–[22]. Facing the heterogeneity of data is a challenging task and must be carefully handled. However, since CD techniques for optical images generally rely on the assumption of data acquired by similar sensors, suboptimal strategies have been considered to make these techniques applicable when considering optical images of different spatial and spectral resolutions [13], [18]. In particular, interpolation and resampling are classically used to obtain a pair of images with the same spatial and spectral resolutions [18], [23]. However, such a compromise solution may remain suboptimal, since it considers each image individually without fully exploiting their joint characteristics and their complementarity. Besides, the approach presented in [24] also provides a suboptimal solution, since it preprocesses each observed image independently with resampling operations bringing both to the same (lower) spectral and spatial resolutions. Thus, this worst case strategy leads to a considerable waste of spatial and spectral information. In this paper, we address the problem of unsupervised CD technique of multiband optical images with different spatial and spectral resolutions. To the best of our knowledge, this is the first operational framework specifically designed to address this issue where the whole spectral and spatial information brought by the pairs of observed images is exploited jointly.

More precisely, this paper addresses the problem of CD between two optical images acquired over the same scene at different time instants, one with low spatial and high spectral resolutions (referred to as the LR image) and one with high spatial and low spectral resolutions (referred to as the HR image). Specifically, typical applicative scenarios considered in this paper differ by the HR and LR images to be compared: PAN and MS, PAN and HS, or MS and HS. The proposed approach consists in first fusing the two observed HR and LR images. If no change between the two observed images occurs, the fused product would be a high spatial and high spectral resolution image of the observed scene, as in any conventional fusion problem. Conversely, in the CD context addressed in this paper, the majority of the fused image pixels correspond to the truly observed scene while the few remaining ones, corresponding to locations impacted by the changes, are expected to suffer from spatial and spectral aberrations due to the inconsistency of the information between the two multidate images. The considered fusion process explicitly relies on a physically based sensing model that exploits the characteristics of the two sensors, following the frameworks in [25] and [26]. Then, these characteristics are subsequently resorted to obtain, by the degradation of the fusion result, two so-called predicted images with the same resolutions as the observed images, i.e., one with low spatial resolution and high spectral resolution and one with high spatial resolution and low spectral resolution. In the absence of changes, these two pairs of predicted and observed images should coincide, apart from residual fusion errors/inaccuracies. Conversely, any change between the two observed images is expected to produce spatial and/or spectral alterations in the fusion result, which will be passed on the predicted images. Finally, each predicted image can be compared with the corresponding observed image of the same resolution to identify possible changes. Since for each pair the images to be compared are of the same resolution, classical CD techniques dedicated to multiband images can be considered [3], [4]. The final result is composed of two CD maps with two different spatial resolutions.

It is worth noting that the fusion and prediction steps will inevitably lead to smoother changes than the ones that effectively occurred in the scene between the two observation times. Consequently, the changes between the fused and predicted images are weaker than the ones that would be observed between images acquired before and after the changes at the same (high) spectral and spatial resolutions. However, the applicative context considered in this paper implies that the spatial and/or spectral resolutions of the two images differ. In this case of interest, the comparison between the two images becomes a problem in itself. The main purpose of the proposed framework is to provide a solution to this nontrivial problem. The experimental results reported in Section VI demonstrate that the proposed fusion framework offers substantial performance improvements with respect to the existing naive methods. Moreover, visual inspection in the case of real images and comparison with ground truth when available show that the smoothing introduced by the fusion step does not prevent from high CD performance.

This paper is organized as follows. Section II formulates the problem of CD between two images with different spatial and spectral resolutions. This formulation relies on a forward model underlying the observed multiband optical images, which is described in Section III. Section IV introduces the main contribution of this paper, i.e., the proposed CD framework, and describes its three main steps. To quantitatively assess the performance of the proposed framework, a new experimental protocol is introduced in Section V. This protocol, inspired by the well-known Wald's protocol [25], allows to generate synthetic, yet realistic, pairs of images affected by physically motivated changes. Moreover, it ensures the availability of a ground truth essential to derive objective figures of merit for detection performance assessment.

In Section VI, the experimental results obtained thanks to this protocol as well as on real multidate LANDSAT 8 images demonstrate the efficiency of the proposed CD framework. Section VII concludes this paper.

II. Problem Formulation

Let us denote t_1 and t_2 the times of acquisition for two single-band or multiband optical images over the same scene of interest. Assume that the image acquired at time t_1 is a high spatial resolution (HR) image and the one acquired at time t_2 is a low spatial resolution (LR) image.[1] These images are denoted by $\mathbf{Y}_{\mathrm{HR}}^{t_1} \in \mathbb{R}^{n_\lambda \times n}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2} \in \mathbb{R}^{m_\lambda \times m}$, respectively, corresponding to matrices every line of which corresponds to the measurements in a given spectral band and every column corresponds to the measurements at a given pixel location, where the following holds.

1) $n = n_r \times n_c$ is the number of pixels in each band of the HR image.
2) $m = m_r \times m_c$ is the number of pixels in each band of the LR image, with $m < n$.
3) n_λ is the number of bands in the HR image.
4) m_λ is the number of bands in the LR image, with $n_\lambda < m_\lambda$.

Recall that three scenarios corresponding to three pairs of HR and LR images can be considered: HR-PAN and LR-MS, HR-PAN and LR-HS, or HR-MS and LR-HS. The main difficulty, which prevents any naive implementation of classical CD techniques, comes from the differences in spatial and spectral resolutions of the two observed images, i.e., $m \neq n$ and $n_\lambda \neq m_\lambda$.

Besides, in digital image processing, it is common to consider the image formation process as a sequence of transformations of the original scene into an output image. The output image of a given sensor is thus a particular limited representation of the original scene with characteristics imposed by the processing pipeline of that sensor, called an *image signal processor* (ISP). The original scene cannot be exactly represented because of its continuous nature. Nevertheless, to represent the ISP pipeline as a sequence of transformations, it is usual to consider a very fine digital approximation of the scene representation as the input image. Following this paradigm, the two observed images $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2}$ are assumed to be spectrally and spatially degraded versions of two corresponding latent (i.e., unobserved) high spatial and spectral resolution images $\mathbf{X}^{t_1}$ and $\mathbf{X}^{t_2}$, respectively:

$$\begin{aligned} \mathbf{Y}_{\mathrm{HR}}^{t_1} &= T_{\mathrm{HR}}[\mathbf{X}^{t_1}] \\ \mathbf{Y}_{\mathrm{LR}}^{t_2} &= T_{\mathrm{LR}}[\mathbf{X}^{t_2}] \end{aligned} \tag{1}$$

where $T_{\mathrm{HR}}[\cdot]$ and $T_{\mathrm{LR}}[\cdot]$, respectively, stand for spectral and spatial degradation operators and $\mathbf{X}^{t_j} \in \mathbb{R}^{m_\lambda \times n}$ $(j = 1, 2)$. Note that these two unobserved images $\mathbf{X}^{t_j} \in \mathbb{R}^{m_\lambda \times n}$ $(j = 1, 2)$ share the same spatial and spectral characteristics and, if they were available, they could be resorted as inputs of classical CD techniques operating on images of the same resolutions.

[1]Note that the modality-time order is not fixed, and without loss of generality, one may state either $t_1 \leq t_2$ or $t_2 \leq t_1$.

When the two images $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2}$ have been acquired at the same time, i.e., $t_1 = t_2$, no change is expected and the latent images $\mathbf{X}^{t_1}$ and $\mathbf{X}^{t_2}$ should represent exactly the same scene, i.e., $\mathbf{X}^{t_1} = \mathbf{X}^{t_2} \triangleq \mathbf{X}$. In such a particular context, recovering an estimate $\hat{\mathbf{X}}$ of the high spatial and spectral resolution latent image $\mathbf{X}$ from the two degraded images $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2}$ can be cast as a fusion problem, for which efficient methods have been recently proposed [26]–[29]. Thus, in the case of a perfect fusion process, the no-change hypothesis $\mathcal{H}_0$ can be formulated as

$$\mathcal{H}_0 : \begin{cases} \mathbf{Y}_{\mathrm{HR}}^{t_1} = \hat{\mathbf{Y}}_{\mathrm{HR}}^{t_1} \\ \mathbf{Y}_{\mathrm{LR}}^{t_2} = \hat{\mathbf{Y}}_{\mathrm{LR}}^{t_2} \end{cases} \tag{2}$$

where

$$\begin{aligned} \hat{\mathbf{Y}}_{\mathrm{HR}}^{t_1} &\triangleq T_{\mathrm{HR}}[\hat{\mathbf{X}}] \\ \hat{\mathbf{Y}}_{\mathrm{LR}}^{t_2} &\triangleq T_{\mathrm{LR}}[\hat{\mathbf{X}}] \end{aligned} \tag{3}$$

are the two predicted HR and LR images from the estimated latent image $\hat{\mathbf{X}}$.

When there exists a time interval between acquisitions, i.e., when $t_1 \neq t_2$, a change may occur meanwhile. In this case, no common latent image $\mathbf{X}$ can be defined since $\mathbf{X}^{t_1} \neq \mathbf{X}^{t_2}$. However, since $\mathbf{X}^{t_1}$ and $\mathbf{X}^{t_2}$ represent the same area of interest, they are expected to keep a certain level of similarity. Thus, the fusion process does not lead to a common latent image, but to a pseudolatent image $\hat{\mathbf{X}}$ from the observed image pair $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2}$, which consists of the best joint approximation of latent images $\mathbf{X}^{t_1}$ and $\mathbf{X}^{t_2}$. Moreover, since $\hat{\mathbf{X}} \neq \mathbf{X}^{t_1}$ and $\hat{\mathbf{X}} \neq \mathbf{X}^{t_2}$, the forward model (1) does not hold to relate the pseudolatent image $\hat{\mathbf{X}}$ to the observations $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2}$. More precisely, when changes have occurred between the two time instants t_1 and t_2, the change hypothesis $\mathcal{H}_1$ can be stated as

$$\mathcal{H}_1 : \begin{cases} \mathbf{Y}_{\mathrm{HR}}^{t_1} \neq \hat{\mathbf{Y}}_{\mathrm{HR}}^{t_1} \\ \mathbf{Y}_{\mathrm{LR}}^{t_2} \neq \hat{\mathbf{Y}}_{\mathrm{LR}}^{t_2}. \end{cases} \tag{4}$$

More precisely, both the inequalities in (4) should be understood in a pixelwise sense, since any change occurring between t_1 and t_2 is expected to affect some spatial locations in the images. As a consequence, both diagnosis in (2) and (4) naturally induce pixelwise rules to decide between the no-change and change hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. This paper specifically proposes to derive a CD framework able to operate on the two observed images $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2}$. This framework implicitly relies on the forward model (1) and the degradation operators $T_{\mathrm{HR}}[\cdot]$ and $T_{\mathrm{LR}}[\cdot]$ introduced to relate the latent and observed images. The forward model adopted in this paper is described in Section III.

III. Forward Model

When dealing with optical images, the sequences of transformations $T_{\mathrm{HR}}[\cdot]$ and $T_{\mathrm{LR}}[\cdot]$ intrinsic to the sensors over the pseudolatent images $\mathbf{X}$ in (1) are generally classified as spectral and spatial degradations. Spatial degradations are related to the spatial characteristics of the sensor, such as the sampling scheme and the optical transfer function. Spectral degradations, on the other hand, are relative to the sensitivity to

wavelength and the spectral sampling. In this paper, following widely admitted assumptions [25], [26], these transformations are considered as linear degradations of the pseudolatent image. Thus, benefiting from convenient matrix representations, the observed images can be expressed as

$$\begin{aligned} \mathbf{Y}_{\mathrm{HR}}^{t_1} &\approx \mathbf{LX} \\ \mathbf{Y}_{\mathrm{LR}}^{t_2} &\approx \mathbf{XR}. \end{aligned} \tag{5}$$

The degradation resulting from the left multiplication by $\mathbf{L} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ models the combination of some spectral bands for each pixel. This degradation corresponds to a spectral resolution reduction with respect to the pseudolatent image $\mathbf{X}$ as in [28] and [30]. In practice, this degradation models an intrinsic characteristic of the sensor called spectral response. It can be either learned by cross calibration or known *a priori*.

Conversely, the right multiplication by $\mathbf{R} \in \mathbb{R}^{n \times m}$ degrades the pseudolatent image by linear combinations of pixels within a given spectral band, thus reducing the spatial resolution. The right degradation matrix $\mathbf{R}$ may model the combination of various transformations that are specific of sensor architectures and take into account external factors such as warp, blurring, translation, and decimation [28], [30], [31]. In this paper, only space-invariant blurring and decimation will be considered. Geometrical transformations, such as warp and translations, can be neglected if both observed images represent the same exact geographical region without strong perceptive distortions. A space-invariant blur can be modeled by a symmetric convolution kernel, yielding a sparse symmetric Toeplitz matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ [27]. It operates a cyclic convolution on the image bands individually. The decimation operation $\mathbf{S} \in \mathbb{R}^{n \times m}$ corresponds to a $d = d_r \times d_c$ uniform downsampling[2] operator with $m = n/d$ ones on the block diagonal and zeros elsewhere, such that $\mathbf{S}^T\mathbf{S} = \mathbf{I}_m$ [28]. Hence, the spatial degradation operation corresponds to the composition $\mathbf{R} = \mathbf{BS} \in \mathbb{R}^{n \times m}$.

The approximating symbol $\approx$ in (5) stands for any mismodeling effects or acquisition noise, which is generally considered as additive and Gaussian [4], [9], [26]–[29]. The full degradation model can thus be written as

$$\begin{aligned} \mathbf{Y}_{\mathrm{HR}}^{t_1} &= \mathbf{LX} + \mathbf{N}_{\mathrm{HR}} \\ \mathbf{Y}_{\mathrm{LR}}^{t_2} &= \mathbf{XBS} + \mathbf{N}_{\mathrm{LR}}. \end{aligned} \tag{6}$$

The additive noise matrices are assumed to be distributed according to matrix normal distributions[3] [32] as follows:

$$\begin{aligned} \mathbf{N}_{\mathrm{HR}} &\sim \mathcal{MN}_{m_\lambda, m}(\mathbf{0}_{m_\lambda \times m}, \mathbf{\Lambda}_{\mathrm{HR}}, \mathbf{I}_m) \\ \mathbf{N}_{\mathrm{LR}} &\sim \mathcal{MN}_{n_\lambda, n}(\mathbf{0}_{n_\lambda \times n}, \mathbf{\Lambda}_{\mathrm{LR}}, \mathbf{I}_n). \end{aligned}$$

Note that the row covariance matrices $\mathbf{\Lambda}_{\mathrm{HR}}$ and $\mathbf{\Lambda}_{\mathrm{LR}}$ carry the information of the spectral variance in-between bands. Since the noise is spectrally colored, these matrices are not necessarily diagonal. On the other hand, since the noise is assumed spatially independent, the column covariance matrices correspond to identity matrices, e.g., $\mathbf{I}_m$ and $\mathbf{I}_n$. In real applications, since the row covariance matrices are an intrinsic characteristic of the sensor, they are estimated by a prior calibration [30]. In this paper, to reduce the number of unknown parameters, we assume that $\mathbf{\Lambda}_{\mathrm{HR}}$ and $\mathbf{\Lambda}_{\mathrm{LR}}$ are both diagonal. This hypothesis implies that the noise is independent from one band to another and is characterized by a specific variance in each band [28].

Capitalizing on this forward model, the proposed three-step CD framework is described in Section IV.

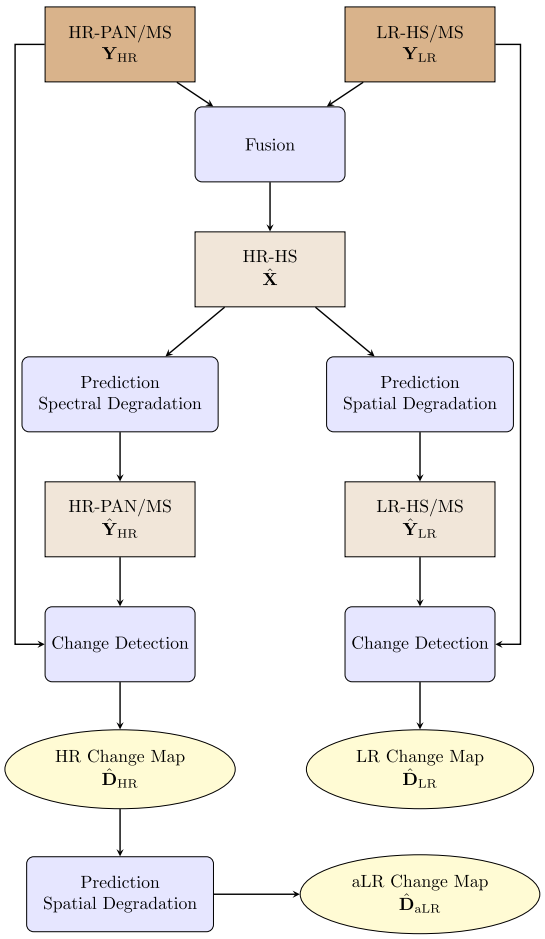


Fig. 1. CD framework.

IV. Proposed Three-Step Framework

This section describes the proposed CD framework that mainly consists of the following three steps, as shown in Fig. 1.

1) *Fusion:* Estimating the pseudolatent image $\hat{\mathbf{X}}$ from $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2}$.
2) *Prediction:* Reconstructing the two HR and LR images $\hat{\mathbf{Y}}_{\mathrm{HR}}^{t_1}$ and $\hat{\mathbf{Y}}_{\mathrm{LR}}^{t_2}$ from $\hat{\mathbf{X}}$.
3) *Decision:* Deriving HR and LR change maps $\hat{\mathbf{D}}_{\mathrm{HR}}$ and $\hat{\mathbf{D}}_{\mathrm{LR}}$ associated with the respective pairs of observed and predicted HR and LR images, namely

$$\Upsilon_{\mathrm{HR}} = \left\{\mathbf{Y}_{\mathrm{HR}}^{t_1}, \hat{\mathbf{Y}}_{\mathrm{HR}}^{t_1}\right\} \quad \text{and} \quad \Upsilon_{\mathrm{LR}} = \left\{\mathbf{Y}_{\mathrm{LR}}^{t_2}, \hat{\mathbf{Y}}_{\mathrm{LR}}^{t_2}\right\}.$$

[2]Note that $\mathbf{S}^T$ represents an upsampling transformation by zero interpolation from m to n.

[3]The probability density function $p(\mathbf{X}|\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c)$ of a matrix normal distribution $\mathcal{MN}_{r,c}(\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c)$ is given by

$$p(\mathbf{X}|\mathbf{M}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c) = \frac{\exp\left(-\frac{1}{2}\mathrm{tr}[\mathbf{\Sigma}_c^{-1}(\mathbf{X}-\mathbf{M})^T\mathbf{\Sigma}_r^{-1}(\mathbf{X}-\mathbf{M})]\right)}{(2\pi)^{rc/2}|\mathbf{\Sigma}_c|^{r/2}|\mathbf{\Sigma}_r|^{c/2}}$$

where $\mathbf{M} \in \mathbb{R}^{r \times c}$ is the mean matrix, $\mathbf{\Sigma}_r \in \mathbb{R}^{r \times r}$ is the row covariance matrix, and $\mathbf{\Sigma}_c \in \mathbb{R}^{c \times c}$ is the column covariance matrix.

An alternate LR (aLR) change map, denoted by $\hat{\mathbf{D}}_{\mathrm{aLR}}$, is also computed by spatially degrading the HR change map $\hat{\mathbf{D}}_{\mathrm{HR}}$ with respect to the spatial operator $T_{\mathrm{LR}}[\cdot]$. A $\hat{\mathbf{D}}_{\mathrm{aLR}}$ pixel then leads to a change decision if at least one of its $\hat{\mathbf{D}}_{\mathrm{HR}}$ parent pixels led to the same decision.

One should highlight the fact that this later *decision* step only requires to implement CD techniques within two pairs of optical images Υ_{HR} and Υ_{LR} of the same spatial and spectral resolutions, thus overcoming the initial issue raised by analyzing observed images $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2}$ with dissimilar resolutions.

To establish the rationale underlying this whole framework, one may refer to the two main properties required by any fusion procedure: consistency and synthesis [26]. The former one requires the reversibility of the fusion process: the original LR and HR images can be obtained by proper degradations of the fused HR-HS image. The latter requires that the fused HR-HS image must be as similar as possible to the image of the same scene that would be obtained by the sensor of the same resolution. Similarly, the generic framework proposed in [25] for fusion image quality assessment can also be properly stated by assigning the consistency and synthesis properties a greater scope.

Moreover, it is also worth noting that the proposed three-step CD framework has been explicitly motivated by the specific scenario of detecting changes between LR and HR optical images. However, it may be applicable for any other applicative context, provided that the following two assumptions hold: 1) a latent image can be estimated from the two observed images and 2) the latent and predicted images can be related through known transformations.

Particular instances of the three steps composing the proposed CD framework steps are proposed in the following paragraphs. The first two ones, i.e., fusion and prediction, explicitly rely on the forward model adopted in Section III. Standard CD techniques able to operate on image pairs of the same spatial and spectral resolutions are finally recalled. However, the main contribution of this paper lies in the whole framework. As a consequence, each step of this framework can be easily tailored by the end-user, provided the two assumptions stated above are ensured.

A. Fusion

The forward observation model (6) has been exploited in many applications involving optical multiband images, especially those related to image restoration such as fusion and super-resolution [28], [30]. Whether the objective is to fuse multiband images from different spatial and spectral resolutions or to increase the resolution of a single one, it consists in compensating the energy tradeoff of optical multiband sensors to get a higher spatial and spectral resolution image compared with the observed image set. One popular approach to conduct fusion consists in solving an inverse problem, formulated through the observation model. In the specific context of HS pansharpening (i.e., fusing PAN and HS images), such an approach has proven to provide the most reliable fused product, with a reasonable computational complexity [26]. For these reasons, this is the strategy followed in this paper and it is briefly sketched in what follows. Nevertheless, the same approach can be used as well for MS pansharpening (i.e., fusing PAN and MS images). Note that the fusion technique described in what follows is considered here for its genericity, and thus can easily be replaced by any more competitive technique by the end-user.

Because of the additive nature and the statistical properties of the noise $\mathbf{N}_{\mathrm{HR}}$ and $\mathbf{N}_{\mathrm{LR}}$, both observed images $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ and $\mathbf{Y}_{\mathrm{LR}}^{t_2}$ are assumed to be distributed according to matrix normal distributions

$$\begin{aligned}\mathbf{Y}_{\mathrm{HR}}^{t_1}|\mathbf{X} &\sim \mathcal{MN}_{m_\lambda,m}(\mathbf{LX}, \mathbf{\Lambda}_{\mathrm{HR}}, \mathbf{I}_m)\\ \mathbf{Y}_{\mathrm{LR}}^{t_2}|\mathbf{X} &\sim \mathcal{MN}_{n_\lambda,n}(\mathbf{XBS}, \mathbf{\Lambda}_{\mathrm{LR}}, \mathbf{I}_n).\end{aligned} \tag{7}$$

Since the noise can be reasonably assumed sensor-dependent, the observed images can be assumed statistically independent. Consequently, the joint likelihood function of the statistically independent observed data can be written

$$p\big(\mathbf{Y}_{\mathrm{HR}}^{t_1}, \mathbf{Y}_{\mathrm{LR}}^{t_2}|\mathbf{X}\big) = p\big(\mathbf{Y}_{\mathrm{HR}}^{t_1}|\mathbf{X}\big)\, p\big(\mathbf{Y}_{\mathrm{LR}}^{t_2}|\mathbf{X}\big) \tag{8}$$

and the negative log-likelihood, defined up to an additive constant, is

$$\begin{aligned}-\log p(\mathbf{\Psi}|\mathbf{X}) = \frac{1}{2}\big\|\mathbf{\Lambda}_{\mathrm{HR}}^{-\frac{1}{2}}\big(\mathbf{Y}_{\mathrm{HR}}^{t_1} - \mathbf{LX}\big)\big\|_F^2 \\ + \frac{1}{2}\big\|\mathbf{\Lambda}_{\mathrm{LR}}^{-\frac{1}{2}}\big(\mathbf{Y}_{\mathrm{LR}}^{t_2} - \mathbf{XBS}\big)\big\|_F^2\end{aligned} \tag{9}$$

where $\mathbf{\Psi} = \{\mathbf{Y}_{\mathrm{HR}}^{t_1}, \mathbf{Y}_{\mathrm{LR}}^{t_2}\}$ denotes the set of observed images and $\|\cdot\|_F^2$ stands for the Frobenius norm.

Computing the maximum likelihood estimator $\hat{\mathbf{X}}_{\mathrm{ML}}$ of $\mathbf{X}$ from the observed image set $\mathbf{\Psi}$ consists in minimizing (9). The aforementioned derivation intends to solve a linear inverse problem that can have a unique solution depending on the properties of the matrices $\mathbf{B}$, $\mathbf{S}$, and $\mathbf{L}$ defining the forward model (6). When the resulting of this inverse problem is ill-posed or ill-conditioned, complementary regularization is needed to promote a relevant and reliable solution. To overcome this issue, additional prior information can be included, setting the estimation problem into the Bayesian formalism [33]. Following a maximum *a posteriori* (MAP) estimation, recovering the estimated pseudolatent image $\hat{\mathbf{X}}$ from the linear model (6) consists in minimizing the negative log-posterior:

$$\begin{aligned}\hat{\mathbf{X}} \in \underset{\mathbf{X}\in\mathbb{R}^{m_\lambda \times n}}{\operatorname{argmin}} \Bigg\{ \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{HR}}^{-\frac{1}{2}}\big(\mathbf{Y}_{\mathrm{HR}}^{t_1} - \mathbf{LX}\big)\right\|_F^2 \\ + \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{LR}}^{-\frac{1}{2}}\big(\mathbf{Y}_{\mathrm{LR}}^{t_2} - \mathbf{XBS}\big)\right\|_F^2 + \lambda\phi(\mathbf{X})\Bigg\}\end{aligned} \tag{10}$$

where $\phi(\cdot)$ defines an appropriate regularizer derived from the prior distribution assigned to $\mathbf{X}$ and λ is a parameter that tunes the relative importance of the regularization and data terms. Computing the MAP estimator (10) is expected to provide the best approximation $\hat{\mathbf{X}}$ with the minimum distance to the latent images $\mathbf{X}^{t_1}$ and $\mathbf{X}^{t_2}$ simultaneously. This optimization problem is challenging because of the high dimensionality of the data $\mathbf{X}$. Nevertheless, Wei *et al.* [28] have proved that

its solution can be efficiently computed for various relevant regularization terms $\phi(\mathbf{X})$. In this paper, a Gaussian prior is considered, since it provides an interesting tradeoff between accuracy and computational complexity, as reported in [26].

B. Prediction

The prediction step relies on the forward model (6) proposed in Section III. As suggested by (3), it merely consists in applying the respective spectral and spatial degradations to the estimated pseudolatent image $\hat{\mathbf{X}}$, leading to

$$\begin{aligned}\hat{\mathbf{Y}}_{\mathrm{HR}}^{t_2} &= \mathbf{L}\hat{\mathbf{X}} \\ \hat{\mathbf{Y}}_{\mathrm{LR}}^{t_1} &= \hat{\mathbf{X}}\mathbf{B}\mathbf{S}.\end{aligned} \tag{11}$$

C. Decision

This section presents the third and last steps of the proposed CD framework, which consists in implementing decision rules to identify possible changes between the images composing the two pairs $\Upsilon_{\mathrm{HR}} = \{\mathbf{Y}_{\mathrm{HR}}^{t_1}, \hat{\mathbf{Y}}_{\mathrm{HR}}^{t_2}\}$ and $\Upsilon_{\mathrm{LR}} = \{\mathbf{Y}_{\mathrm{LR}}^{t_2}, \hat{\mathbf{Y}}_{\mathrm{LR}}^{t_1}\}$. As noted in Section II, these CD techniques operate on observed $\mathbf{Y}_{\triangle R}^{t_i}$ and predicted $\hat{\mathbf{Y}}_{\square \mathrm{R}}^{t_j}$ images of the same spatial and spectral resolutions, with $(i, \triangle) \in \{(1, H), (2, L)\}$ and $(j, \square) \in \{(2, H), (1, L)\}$, as in [2], [3], [34], and [35]. Unless specifically tailored by the end-user, these techniques can be *a priori* employed whatever the number of spectral bands. As a consequence, $\mathbf{Y}_{\triangle R}^{t_i}$ and $\hat{\mathbf{Y}}_{\square R}^{t_j}$ could refer to either PAN, MS, or HS images and the two resulting CD maps are either of HR, either of LR, associated with the pairs Υ_{HR} and Υ_{LR}, respectively. To lighten the notations, without loss of generality, the pairs $\mathbf{Y}_{\triangle R}^{t_i}$ and $\hat{\mathbf{Y}}_{\square R}^{t_j}$ will be denoted by $\mathbf{Y}^{t_i} \in \mathbb{R}^{\ell \times \eta}$ and $\mathbf{Y}^{t_j} \in \mathbb{R}^{\ell \times \eta}$, which can be set as follows.

1) $\{\mathbf{Y}^{t_i}, \mathbf{Y}^{t_j}\} = \Upsilon_{\mathrm{LR}}$ to derive the estimated CD binary map $\hat{\mathbf{D}}_{\mathrm{LR}}$ at LR.
2) $\{\mathbf{Y}^{t_i}, \mathbf{Y}^{t_j}\} = \Upsilon_{\mathrm{HR}}$ to derive the estimated CD binary map $\hat{\mathbf{D}}_{\mathrm{HR}}$ at HR and its spatially degraded aLR counterpart $\hat{\mathbf{D}}_{\mathrm{aLR}}$.

In this seek of generality, the numbers of bands and pixels are denoted by ℓ and η, respectively. The spectral dimension ℓ depends on the considered image sets Υ_{HR} or Υ_{LR}, i.e., $\ell = n_\lambda$ and $\ell = m_\lambda$ for HR and LR images, respectively.[4] Similarly, the spatial resolution of the CD binary map generically denoted by $\hat{\mathbf{D}} \in \mathbb{R}^{\eta}$ depends on the considered set of images Υ_{HR} or Υ_{LR}, i.e., $\eta = n$ and $\eta = m$ for HR and LR images, respectively.

As stated before, the main contribution of this paper lies in the proposed three-step CD framework able to deal with optical images of different spatial and spectral resolution. Thus, it does not aim at selecting the most powerful technique for each step of this framework. As a consequence, in what follows, three particular yet well-admitted and efficient CD techniques operating on images of the same spatial and spectral resolutions are discussed. Obviously, any other concurrent CD technique can be alternatively considered.

[4]Note, in particular, that $\ell = n_\lambda = 1$ when the set of HR images are PAN images.

1) Change Vector Analysis: When considering multiband optical images that have been atmospherically precalibrated and that represent exactly the same geographical region without strong perceptive distortions, one may consider that, for a pixel at spatial location p

$$\begin{aligned}\mathbf{Y}^{t_i}(p) &\sim \mathcal{N}(\boldsymbol{\mu}^{t_i}, \boldsymbol{\Sigma}^{t_i}) \\ \mathbf{Y}^{t_j}(p) &\sim \mathcal{N}(\boldsymbol{\mu}^{t_j}, \boldsymbol{\Sigma}^{t_j})\end{aligned} \tag{12}$$

where $\boldsymbol{\mu}^{t_i} \in \mathbb{R}^{\ell}$ and $\boldsymbol{\mu}^{t_j} \in \mathbb{R}^{\ell}$ correspond to the pixel spectral mean and $\boldsymbol{\Sigma}^{t_i} \in \mathbb{R}^{\ell \times \ell}$ and $\boldsymbol{\Sigma}^{t_j} \in \mathbb{R}^{\ell \times \ell}$ are the spectral covariance matrices (here, they were obtained using the maximum likelihood estimator) [11]. A comparison of the pixels $\mathbf{Y}^{t_i}(p)$ and $\mathbf{Y}^{t_j}(p)$ can be naturally conducted by deriving the SCV magnitude $\mathbf{V}_{\mathrm{CVA}}(p)$. According to a generalized formulation, it is defined by the squared Mahalanobis distance between the two pixels [2], [36], which can be computed from the pixelwise spectral difference operator $\Delta\mathbf{Y}(p) = \mathbf{Y}^{t_i}(p) - \mathbf{Y}^{t_j}(p)$, that is

$$\mathbf{V}_{\mathrm{CVA}}(p) = \|\Delta\mathbf{Y}(p)\|^2_{\boldsymbol{\Sigma}^{-1}} = \Delta\mathbf{Y}(p)^T \boldsymbol{\Sigma}^{-1} \Delta\mathbf{Y}(p) \tag{13}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{t_i} + \boldsymbol{\Sigma}^{t_j}$. For a given threshold τ, the pixelwise statistical test can be formulated as

$$\mathbf{V}_{\mathrm{CVA}}(p) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{14}$$

and the final CD map denoted by $\hat{\mathbf{D}}_{\mathrm{CVA}} \in \{0, 1\}^{\eta}$ can be pixelwise derived as

$$\hat{\mathbf{D}}_{\mathrm{CVA}}(p) = \begin{cases} 1 & \text{if } \mathbf{V}_{\mathrm{CVA}}(p) \geq \tau \ (\mathcal{H}_1) \\ 0 & \text{otherwise } (\mathcal{H}_0). \end{cases} \tag{15}$$

For a pixel that has not been affected by a change (hypothesis $\mathcal{H}_0$), the spectral difference operator is expected to be statistically described by $\Delta\mathbf{Y}(p) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. As a consequence, the threshold τ can be related to the probability of false alarm (PFA) of the test

$$P_{\mathrm{FA}} = \mathbb{P}[\mathbf{V}_{\mathrm{CVA}}(p) > \tau | \mathcal{H}_0] \tag{16}$$

or equivalently

$$\tau = F^{-1}_{\chi^2_\ell}(1 - P_{\mathrm{FA}}) \tag{17}$$

where $F^{-1}_{\chi^2_\ell}(\cdot)$ is the inverse cumulative distribution function of the χ^2_ℓ distribution. Note that jointly considering the magnitude and angle of the SCV may provide additional relevant information, and thus may allow changes to be better detected and classified [11].

2) Spatially Regularized Change Vector Analysis: Since change vector analysis (CVA) in its simplest form, as presented in Section IV-C1, is a pixelwise procedure, it significantly suffers from low robustness with respect to noise. To overcome this limitation, spatial information can be exploited by considering the neighborhood of a pixel to compute the final distance criterion, which is expected to make the change map spatially smoother. Indeed, changed pixels are generally gathered together into regions or clusters, which means that there is a high probability to observe changes in the neighborhood of an identified changed pixel [3]. Let Ω_p^L denote the set of indexes of neighboring spatial locations of a given

pixel p defined by a surrounding regular window of size L centered on p. The spatially smoothed energy map $\mathbf{V}_{\text{sCVA}}$ of the spectral difference operator can be derived from its pixelwise counterpart $\mathbf{V}_{\text{CVA}}$ defined by (13) as

$$\mathbf{V}_{\text{sCVA}}(p) = \frac{1}{|\Omega_p^L|} \sum_{k \in \Omega_p^L} \omega(k) \mathbf{V}_{\text{CVA}}(k) \tag{18}$$

where the weights $\omega(k) \in \mathbb{R}$, $k \in \{1, \ldots, |\Omega_p^L|\}$, implicitly define a spatial smoothing filter. In this paper, they have been chosen as $\omega(k) = 1$, $\forall k$. Then, a decision rule similar to (15) can be followed to derive the final CD map $\hat{\mathbf{D}}_{\text{sCVA}}$. Note that the choice of the window size L is based on the strong hypothesis of the window homogeneity. This choice thus may depend upon the kind of observed scenes.

3) Iteratively Reweighted Multivariate Alteration Detection: The multivariate alteration detection (MAD) technique introduced in [13] has been shown to be a robust CD technique due its suitability for analyzing the multiband image pair $\{\mathbf{Y}_1, \mathbf{Y}_2\}$ with possible different intensity levels. Similar to the CVA and sCVA methods, it exploits an image differencing operator while better concentrating information related to changes into auxiliary variables. More precisely, the MAD variate is defined as $\Delta\tilde{\mathbf{Y}}(p) = \tilde{\mathbf{Y}}^{t_i}(p) - \tilde{\mathbf{Y}}^{t_j}(p)$ with

$$\begin{aligned} \tilde{\mathbf{Y}}^{t_1}(p) &= \mathbf{U}\mathbf{Y}^{t_i}(p) \\ \tilde{\mathbf{Y}}^{t_2}(p) &= \mathbf{W}\mathbf{Y}^{t_j}(p) \end{aligned} \tag{19}$$

where $\mathbf{U} = [\mathbf{u}_\ell, \mathbf{u}_{\ell-1}, \ldots, \mathbf{u}_1]^T$ is an $\ell \times \ell$-matrix composed of the $\ell \times 1$-vectors $\mathbf{u}_k$ identified by canonical correlation analysis and $\mathbf{W} = [\mathbf{w}_\ell, \mathbf{w}_{\ell-1}, \ldots, \mathbf{w}_1]^T$ is defined similarly. As in (13), the MAD-based change energy map can then be derived as

$$\mathbf{V}_{\text{MAD}}(p) = \|\Delta\tilde{\mathbf{Y}}(p)\|^2_{\mathbf{\Lambda}^{-1}}$$

where $\mathbf{\Lambda}$ is the diagonal covariance matrix of the MAD variates. Finally, the MAD CD map $\hat{\mathbf{D}}_{\text{MAD}}$ can be pixelwisely computed using a decision rule similar to (15) with a threshold τ related to the PFA by (17). In this paper, the iteratively reweighted version of MAD (IR-MAD) has been considered to better separate the change pixels from the no-change pixels [14].

V. Experimental Protocol for Performance Assessment

A. General Overview

Assessing the performance of CD algorithms requires image pairs with particular characteristics, which makes them rarely freely available. Indeed, CD algorithms require images acquired at two different dates, presenting changes, representing exactly the same geographical region without strong perceptive distortions, which have been radiometrically precorrected and, for the specific problem addressed in this paper, coming from different optical sensors. Moreover, to properly and statistically assess the performance of the proposed CD framework, these image pairs need to be accompanied by a ground-truth information in the form of a validated CD mask.

To overcome this issue, this section proposes a modified version of the simulation protocol introduced in [25] that was initially proposed to assess the performance of image fusion algorithms. This protocol relies on a unique reference HS image $\mathbf{X}_{\text{ref}}$, also considered as HR. It avoids the need of coregistered and geometrically corrected images by generating a pair of synthetic but realistic HR and LR images from this reference image and by including changes within a semantic description of this HR-HS image. In this paper, this description is derived by spectral unmixing [37] and the full proposed protocol can be summarized as follows.

1) Unmixing the Reference Image: Given a reference image $\mathbf{X}_{\text{ref}} \in \mathbb{R}^{m_\lambda \times n}$, conduct linear unmixing to extract K endmember signatures $\mathbf{M}^{t_1} \in \mathbb{R}^{m_\lambda \times K}$ and the associated abundance matrix $\mathbf{A}^{t_1} \in \mathbb{R}^{K \times n}$ such that $\mathbf{X}_{\text{ref}} \approx \mathbf{M}^{t_1}\mathbf{A}^{t_1}$.

2) Generating the Before-Change HR-HS Image: Define the HR-HS image $\mathbf{X}^{t_1}$ before change as

$$\mathbf{X}^{t_1} = \mathbf{M}^{t_1}\mathbf{A}^{t_1}. \tag{20}$$

3) Generating HR and LR Change Masks: Define a reference HR change mask $\mathbf{D}_{\text{HR}}$ by selecting particular regions (i.e., pixels) in the latent image $\mathbf{X}^{t_1}$ where changes occur. The corresponding LR change mask $\mathbf{D}_{\text{LR}}$ is computed according to the spatial degradations relating the two modalities. Both the change masks will be considered as the ground truth and will be compared with the estimated CD HR map $\hat{\mathbf{D}}_{\text{HR}}$ and LR maps $\hat{\mathbf{D}}_{\text{LR}}$ and $\hat{\mathbf{D}}_{\text{aLR}}$, respectively, to evaluate the performance of the CD.

4) Implementing Change Rules: According to this reference HR change mask, realistic change rules are implemented on the reference abundances $\mathbf{A}^{t_1}$ associated with pixels affected by changes. The abundance matrix after the changes can be written $\mathbf{A}^{t_2} = \vartheta_{\mathbf{A}}(\mathbf{A}^{t_1}, \mathbf{D}_{\text{HR}})$, where $\vartheta_{\mathbf{A}}(\cdot, \mathbf{D}_{\text{HR}})$ stands for an abundance change-inducing function associated with the HR change mask $\mathbf{D}_{\text{HR}}$. Several change rules applied to the before-change abundance matrix will be discussed in Section V-C. Note that these rules may also require the use of additional endmembers that are not initially present in the latent image $\mathbf{X}^{t_1}$. Thus, with similar notations, the endmember matrix after changes can be denoted by $\mathbf{M}^{t_2} = \vartheta_{\mathbf{M}}(\mathbf{M}^{t_1}, \mathbf{D}_{\text{HR}})$.

5) Generating the After-Change HR-HS Image: Define the HR-HS latent image $\mathbf{X}^{t_2}$ after changes by linear mixing such that

$$\mathbf{X}^{t_2} = \mathbf{M}^{t_2}\mathbf{A}^{t_2}. \tag{21}$$

6) Generating the Observed HR Image: Generate a simulated observed HR image $\mathbf{Y}_{\text{HR}}$ by applying the spectral degradation $T_{\text{HR}}[\cdot]$ either to the before-change HR-HS latent image $\mathbf{X}^{t_1}$ or to the after-change HR-HS latent image $\mathbf{X}^{t_2}$. The observed HR image can be subsequently corrupted by noise according to (6).

7) Generating the Observed LR Image: Conversely, generate a simulated observed LR image $\mathbf{Y}_{\text{LR}}$ by applying the spatial degradation $T_{\text{LR}}[\cdot]$ either to the after-change HR-HS latent image $\mathbf{X}^{t_2}$ or to the before-change HR-HS latent image $\mathbf{X}^{t_1}$. The observed LR image can be subsequently corrupted by noise according to (6).

The full protocol is schematically illustrated in Fig. 2. Complementary information on how these steps have been

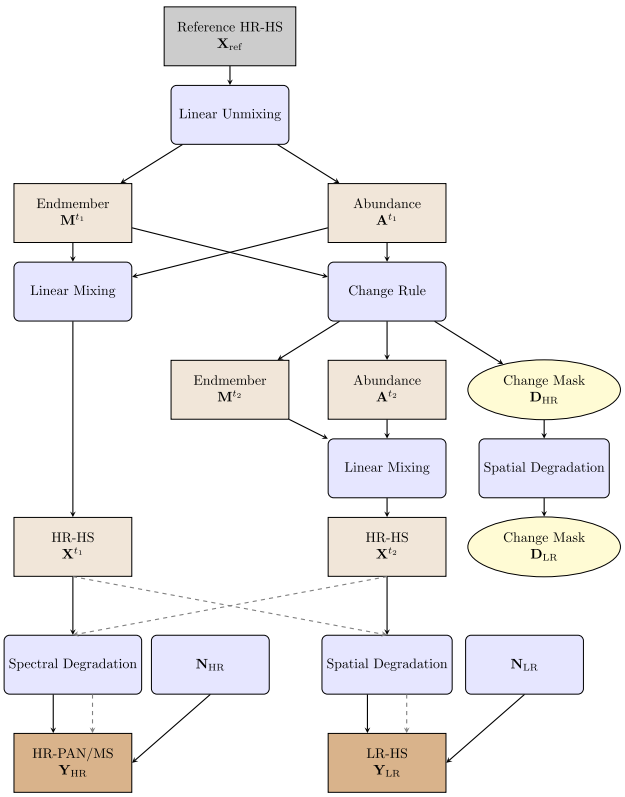


Fig. 2. Simulation protocol: two HR-HS latent images $\mathbf{X}^{t1}$ (before changes) and $\mathbf{X}^{t2}$ (after changes) are generated from the reference image. In temporal configuration 1 (black box), the observed HR image $\mathbf{Y}_{\text{HR}}$ is a spectrally degraded version of $\mathbf{X}^{t_1}$, while the observed LR image $\mathbf{Y}_{\text{LR}}^{t_2}$ is a spatially degraded version of $\mathbf{X}^{t_2}$. In temporal configuration 2 (gray dashed lines), the degraded images are generated from reciprocal HR-HS images.

implemented in this paper is provided in the following paragraphs.

B. Reference Image

The HR-HS reference image used in the simulation protocol is an HS image of the Pavia University in Italy acquired by the reflective optics system imaging spectrometer sensor with 610×340 pixels. The number of spectral bands in this image is 103 with a spectral coverage ranging from 0.43 to 0.86 μm and a spatial resolution of 1.3 m/pixel. A precorrection has been conducted to smooth the atmospheric effects due to vapor water absorption by removing corresponding spectral bands. Then the final HR-HS reference image is of size $610 \times 330 \times 93$.

C. Generating the HR-HS Latent Images: Unmixing, Change Mask, and Change Rules

To produce the HR-HS latent image $\mathbf{X}^{t_1}$ before change, the reference image $\mathbf{X}_{\text{ref}}$ has been linearly unmixed, which provides the endmember matrix $\mathbf{M}^{t_1} \in \mathbb{R}^{m_\lambda \times K}$ and the matrix of abundances $\mathbf{A}^{t_1} \in \mathbb{R}^{K \times n}$, where K is the number of endmembers. This number K can be obtained by investigating the dimension of the signal subspace, for instance, by conducting principal component analysis [37]. In this paper, the linear unmixing has been conducted by coupling the vertex component analysis [38] as an endmember extraction algorithm and the fully constrained least squares algorithm [39] to obtain $\mathbf{M}^{t_1}$ and $\mathbf{A}^{t_1}$, respectively.

Given the HR-HS latent image $\mathbf{X}^{t_1} = \mathbf{M}^{t_1}\mathbf{A}^{t_1}$, the HR change mask $\mathbf{D}_{\text{HR}}$ has been produced by selecting spatial regions in the HR-HS image affected by changes. This selection can be made randomly or by using prior knowledge on the scene. In this paper, manual selection is performed.

Then, the change rules applied to the abundance matrix $\mathbf{A}^{t_1}$ to obtain the changed abundance matrix $\mathbf{A}^{t_2}$ are chosen such that they satisfy the standard positivity and sum-to-one constraints

$$\text{Nonnegativity } a_k^{t_2}(p) \geq 0 \quad \forall p \in \{1, \ldots, n\} \quad \forall k \in \{1, \ldots, K\} \tag{22}$$

$$\text{Sum-to-one } \sum_{k=1}^{K} a_k^{t_2}(p) = 1 \quad \forall p \in \{1, \ldots, n\}. \tag{23}$$

More precisely, three distinct change rules have been considered.

1) *Zero Abundance:* Find the most present endmember in the selected region, set all corresponding abundances to zero, and rescale abundances associated with remaining endmembers in order to fulfill (22). This change can be interpreted as a total disappearance of the most present endmember.
2) *Same Abundance:* Choose a pixel abundance vector at random spatial location, and set all abundance vectors inside the region affected by changes to the chosen one. This change consists in filling the change region by the same spectral signature.
3) *Block Abundance:* Randomly select a region with the same spatial shape as the region affected by changes and replace original region abundances by the abundances of the second one. This produces a "copy-paste" pattern.

Note that other change rules on the abundance matrix $\mathbf{A}^{t_1}$ could have been investigated; in particular, some of them could require to include additional endmembers in the initial endmember matrix $\mathbf{M}^{t_1}$. The updated abundance $\mathbf{A}^{t_2}$ and endmember $\mathbf{M}^{t_2}$ matrices allow to define the after-change HR-HS latent image $\mathbf{X}^{t_2}$ as

$$\mathbf{X}^{t_2} = \mathbf{M}^{t_2}\mathbf{A}^{t_2}.$$

Fig. 3 shows an example of $\mathbf{X}^{t_2}$ using the three different change rules for one single selected region.

D. Generating the Observed Images: Spectral and Spatial Degradations

To produce spectrally degraded versions $\mathbf{Y}_{\text{HR}}$ of the HR-HS latent image $\mathbf{X}^{t_j}$ ($j = 1$ or $j = 2$), two particular spectral responses have been used to assess the performance of the proposed algorithm when analyzing an HR-PAN or a four-band HR-MS image. The former has been obtained by uniformly averaging the first 43 bands of the HR-HS pixel spectra. The later has been obtained by filtering the HR-HS latent image $\mathbf{X}^{t_j}$ by a four-band LANDSAT-like spectral response.

To generate a spatially degraded image, the HR-HS latent image $\mathbf{X}^{t_j}$ ($j = 2$ or $j = 1$) has been blurred by

(a) (b) (c)

Fig. 3. Example of after-change HR-HS latent images $\mathbf{X}^{t_2}$ generated by each proposed change rule. (a) Zero abundance. (b) Same abundance. (c) Block abundance.

(a) (b) (c)

Fig. 4. Degraded versions of the before-change HR-HS latent image $\mathbf{X}^{t_1}$. (a) Spectrally degraded HR-PAN image. (b) Spectrally degraded HR-MS image. (c) Spatially degraded LR-HS image.

a 5×5 Gaussian kernel filter and downsampled equally in vertical and horizontal directions with a factor $d = 5$. This spatial degradation operator implicitly relates the generated HR change mask $\mathbf{D}_{\text{HR}}$ to its LR counterpart $\mathbf{D}_{\text{LR}}$. Each LR pixel contains $d \times d$ HR pixels. As $\mathbf{D}_{\text{HR}}$ is a binary mask, after the spatial degradation, if at least one of its HR parent pixels is considered as a change pixel, then a pixel in $\mathbf{D}_{\text{LR}}$ is also considered as a change pixel.

To illustrate the impact of these spectral and spatial degradations, Fig. 4 shows the corresponding HR-PAN [Fig. 4(a)] and HR-MS [Fig. 4(b)] images resulting from spectral degradations and an LR-HS image resulting from spatial degradation [Fig. 4(c)].

Note that, as mentioned in Section II, the modality-time order can be arbitrary fixed, and without loss of generality, one may state either $t_1 \leq t_2$ or $t_2 \leq t_1$. Thus, there are two distinct temporal configurations to generate the pair of observed HR and LR images.

1) *Configuration 1:* Generating the spectrally degraded observed image $\mathbf{Y}_{\text{HR}}^{t_1}$ from the before-change HR-HS latent image $\mathbf{X}^{t_1}$ and the spatially degraded observed image $\mathbf{Y}_{\text{LR}}^{t_2}$ from the after-change HR-HS latent image $\mathbf{X}^{t_2}$.
2) *Configuration 2:* Generating the spatially degraded observed image $\mathbf{Y}_{\text{LR}}^{t_1}$ from the before-change HR-HS latent image $\mathbf{X}^{t_1}$ and the spectrally degraded observed image $\mathbf{Y}_{\text{HR}}^{t_2}$ from the after-change HR-HS latent image $\mathbf{X}^{t_2}$.

VI. Experimental Results

This section assesses the performance of the proposed fusion-based CD framework. The considered figures of merit and compared methods are discussed in Sections VI-A and VI-B, respectively. Then, Section VI-C reports qualitative and quantitative results for three distinct scenarios associated with the experimental protocol introduced in Section V. Scenario 1 considers the CD problem between a pair of HR-MS and LR-HS images. Scenario 2 focuses on the CD problem between a pair of HR-PAN and LR-HS images. Scenario 3 considers a pair of HR-PAN and LR-MS images. Finally, additional illustrative results obtained on a pair of real HR-PAN and LR-MS images (akin to Scenario 3) are presented in Section VI-D. Note that complementary results are reported in [40].

A. Performance Criteria

The CD framework introduced in Section II has been evaluated following the simulation protocol described in the previous paragraph. As detailed in Section II, one HR CD map $\hat{\mathbf{D}}_{\text{HR}}$ and two LR CD maps $\hat{\mathbf{D}}_{\text{LR}}$ and $\hat{\mathbf{D}}_{\text{aLR}}$ are produced from the CD framework described in Fig. 1. These HR and LR CD maps are, respectively, compared with the actual HR $\mathbf{D}_{\text{HR}}$ and LR $\mathbf{D}_{\text{LR}}$ masks to derive the empirical probabilities of false alarm P_{FA} and detection P_D that are represented through empirical receiver operating characteristic (ROC) curves, i.e., $P_D = f(P_{\text{FA}})$. These ROC curves have been averaged over the 450 available pairs of observed images to mitigate the influence of the change region, the influence of the temporal configuration, and the influence of the change rule.

Moreover, as quantitative figures of merit, two metrics derived from these ROC curves have been considered: 1) the area under the curve (AUC), which is expected to be close to one for a good testing rule and 2) a normalized distance between the no-detection point (defined by $P_{\text{FA}} = 1$ and $P_D = 0$) and the intersect of the ROC curve with the diagonal line $P_{\text{FA}} = 1 - P_D$, which should be close to one for a good testing rule.

B. Compared Methods

While implementing the proposed CD framework, the fusion step in Section IV-A has been conducted following the method proposed in [28] with the Gaussian regularization because of its accuracy and computational efficiency. The corresponding regularization parameter has been chosen as $\lambda = 0.0001$ by cross validation. Regarding the detection step, when considering a pair of multiband images (i.e., MS or HS), the four CD techniques detailed in Section IV-A (i.e., CVA, sCVA, MAD, and IR-MAD) have been considered. Conversely, when one of the observed images is PAN, only CVA and sCVA have been considered, since MAD and IR-MAD require multiband images.

In the absence of state-of-the-art CD techniques able to simultaneously handle images with distinct spatial and spectral resolutions, the proposed CD framework has been compared with the crude approach that first consists in spatially (respectively, spectrally) degrading the observed HR (respectively, LR) image. The classical CD techniques described in Section IV-C can then be applied to the resulting pair of LR images, since they share the same, unfortunately low, spatial and spectral resolutions. The final result is a so-called worst case LR CD mask denoted by $\hat{\mathbf{D}}_{\mathrm{WC}}$ in the following.

C. Results

This paragraph provides the results associated with Scenario 1 (i.e., HR-MS and LR-HS images), Scenario 2 (i.e., HR-PAN and LR-HS images), and Scenario 3 (i.e., HR-PAN and LR-HS images). For each scenario, according to the protocol described in Section V, 75 regions have been randomly selected in the before-change HR-HS latent image $\mathbf{X}^{t_1}$ as those affected by changes. For each region, one of the three proposed change rules (zero abundance, same abundance, or block abundance) has been applied to build the after-change HR-HS latent image $\mathbf{X}^{t_2}$. The observed HR and LR images are generated according to one of the two temporal configurations discussed at the end of Section V-D. This leads to a total of 450 simulated pairs of HR and LR images corresponding to three sets of 150 pairs generated following each of the three distinct change rules described in Section V-C. To evaluate the robustness of proposed method against noise, both observed images for each simulated pair have been corrupted with a zero mean Gaussian noise leading to an SNR $= 30$ dB.

1) Scenario 1: Change Detection Between HR-MS and LR-HS Images: The first simulation scenario considers a set of HR-MS and LR-HS images. The ROC curves are plotted in Fig. 5 with the corresponding performance metrics reported in Table I. These results show that, for the majority of the implemented CD techniques (CVA, sCVA, MAD, or IR-MAD), the proposed framework generally offers high precision. In particular, the aLR change map $\hat{\mathbf{D}}_{\mathrm{aLR}}$ subsequently computed from the estimated HR change map $\hat{\mathbf{D}}_{\mathrm{HR}}$ provides better results than those obtained in the worst case and those obtained on the estimated LR change map $\hat{\mathbf{D}}_{\mathrm{LR}}$ directly. This can be explained by the intrinsic quality of the estimated HR change map $\hat{\mathbf{D}}_{\mathrm{HR}}$, which roughly provides similar detection performance to the aLR change map $\hat{\mathbf{D}}_{\mathrm{aLR}}$ with the great advantage to be available at a finer spatial resolution. Besides, all methods have their own advantages and disadvantages. The worst case method is based on systematic spectral and spatial degradations of the two images. These operations are performed through local weighted averaging, thus leading to not only resolution reduction but also noise reduction. Moreover, this method does not introduce estimation errors. Finally, the images, at the input of the change detector, are smoother than the original ones and the detection rates obtained with sCVA are relatively high. Nevertheless, the method detects changes with the minimum spatial and spectral resolutions of the two observed images. On the contrary, the other considered methods may introduce estimation errors, since, in particular, the predicted image has been smoothed by the fusion and the prediction steps. Thus, the final comparison between the observed and predicted images is slightly skewed due to the differences of noise levels between them. On the other hand, these methods detect changes with higher spectral and spatial resolutions than the worst case scenario. The HR CD allows for a more accurate exploration of the spatial domain. This results in higher detection rates when the associated CD method does not take into account spatial neighborhood. This is the case for (not spatially regularized) CVA, MAD, and IR-MAD methods.

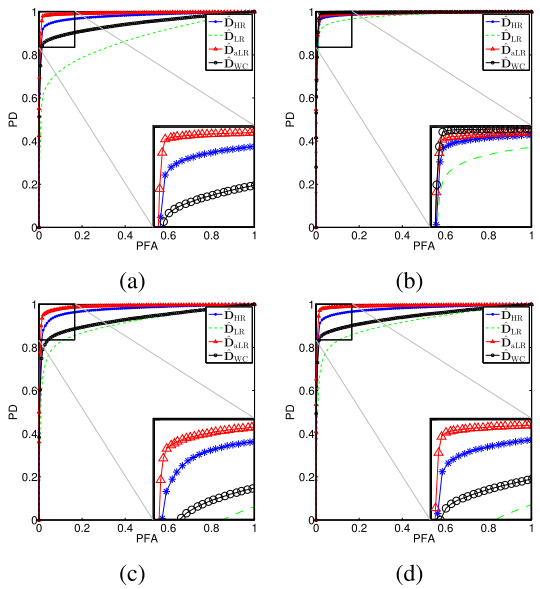


Fig. 5. Scenario 1 (SNR $= 30$ dB). ROC curves computed from (a) CVA, (b) sCVA(7), (c) MAD, and (d) IR-MAD.

TABLE I

SCENARIO 1 (SNR $= 30$ dB): DETECTION PERFORMANCE IN TERMS OF AUC AND NORMALIZED DISTANCE

		$\hat{\mathbf{D}}_{\mathrm{HR}}$	$\hat{\mathbf{D}}_{\mathrm{LR}}$	$\hat{\mathbf{D}}_{\mathrm{aLR}}$	$\hat{\mathbf{D}}_{\mathrm{WC}}$
CVA	AUC	0.977827	0.867478	**0.992242**	0.941408
	Dist.	0.944194	0.789379	**0.979298**	0.887789
sCVA(3)	AUC	0.994539	0.99104	0.995402	**0.995497**
	Dist.	0.982398	0.967797	**0.983498**	0.978398
sCVA(5)	AUC	0.992761	0.989886	0.993074	**0.996224**
	Dist.	0.980798	0.962896	0.980998	**0.982598**
sCVA(7)	AUC	0.990103	0.978382	0.993065	**0.995861**
	Dist.	0.973397	0.936794	0.980298	**0.983098**
MAD	AUC	0.974411	0.912307	**0.989774**	0.929109
	Dist.	0.934893	0.848785	**0.962896**	0.872587
IR-MAD	AUC	0.977063	0.914570	**0.992062**	0.940537
	Dist.	0.941594	0.851385	**0.976498**	0.886089

2) Scenario 2: Change Detection Between HR-PAN and LR-HS Images: In the second scenario, the same procedure as Scenario 1 has been considered while replacing the observed HR-MS image by an HR-PAN image. The ROC curves are depicted in Fig. 6 with the corresponding metrics in Table II. As for Scenario 1, the comparison of these curves shows that the HR CD map also leads to a high spatial accuracy, since it is sharper than the LR maps. In particular, when considering CVA, it provides a significantly more powerful test than

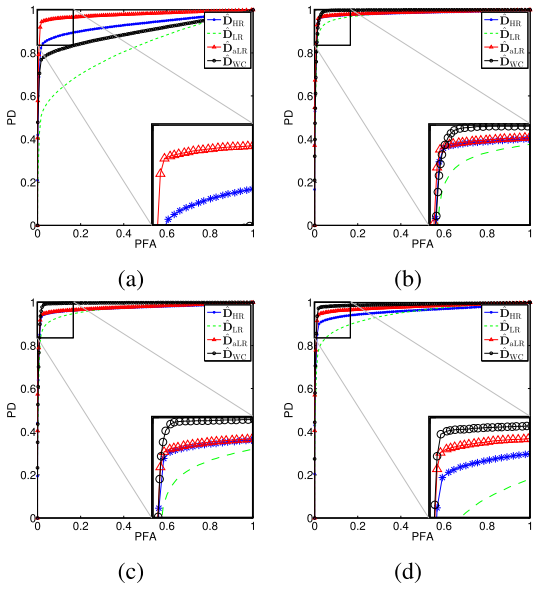


Fig. 6. Scenario 2 (SNR = 30 dB). ROC curves computed from (a) CVA, (b) sCVA(3), (c) sCVA(5), and (d) sCVA(7).

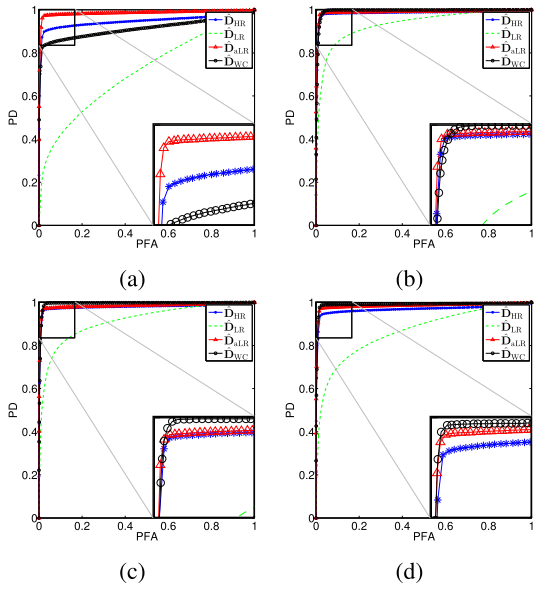


Fig. 7. Scenario 3 (SNR = 30 dB). ROC curves computed from (a) CVA, (b) sCVA(3), (c) sCVA(5), and (d) sCVA(7).

TABLE II

SCENARIO 2 (SNR = 30 dB): DETECTION PERFORMANCE IN TERMS OF AUC AND NORMALIZED DISTANCE

		$\hat{\mathbf{D}}_{\mathrm{HR}}$	$\hat{\mathbf{D}}_{\mathrm{LR}}$	$\hat{\mathbf{D}}_{\mathrm{aLR}}$	$\hat{\mathbf{D}}_{\mathrm{WC}}$
CVA	AUC	0.981039	0.872184	**0.99297**	0.94593
	Dist.	0.951995	0.79558	**0.980098**	0.89579
sCVA(3)	AUC	0.994627	0.990474	0.995458	**0.995545**
	Dist.	0.982198	0.966697	**0.983298**	0.978398
sCVA(5)	AUC	0.992861	0.98917	0.993024	**0.996278**
	Dist.	0.980898	0.961396	0.981598	**0.982298**
sCVA(7)	AUC	0.990192	0.977146	0.992978	**0.995843**
	Dist.	0.973397	0.934893	0.980598	**0.983098**

TABLE III

SCENARIO 3 (SNR = 30 dB): DETECTION PERFORMANCE IN TERMS OF AUC AND NORMALIZED DISTANCE

		$\hat{\mathbf{D}}_{\mathrm{HR}}$	$\hat{\mathbf{D}}_{\mathrm{LR}}$	$\hat{\mathbf{D}}_{\mathrm{aLR}}$	$\hat{\mathbf{D}}_{\mathrm{WC}}$
CVA	AUC	0.936336	0.828627	**0.981096**	0.900359
	Dist.	0.890289	0.750575	**0.958096**	0.839484
sCVA(3)	AUC	0.962478	0.926186	0.980694	**0.989641**
	Dist.	0.931093	0.856986	0.957696	**0.977098**
sCVA(5)	AUC	0.976484	0.950405	0.980059	**0.994151**
	Dist.	0.954695	0.891689	0.957396	**0.978398**
sCVA(7)	AUC	0.98491	0.960965	0.988116	**0.993732**
	Dist.	0.970097	0.909691	0.972697	**0.972797**

the crude approach that consists in degrading both observed HR-PAN and LR-HS images to reach the same spatial and spectral resolutions.

3) Scenario 3: Change Detection Between HR-PAN and LR-MS Images: In the third scenario, the same procedure as Scenario 2 has been considered while replacing the observed LR-HS image by an LR-MS image. This scenario is the most common for optical multiband change detection, since MS optical images are more readily available than HS optical images. The ROC curves are depicted in Fig. 7 with the corresponding metrics in Table III. As for Scenario 1, the comparison of these curves shows that the HR CD map also leads to a high spatial accuracy when resorting to CVA, since it is sharper than the LR maps. When resorting to the spatially regularized counterpart of CVA, the worst case method and the proposed aLR-based detection provide similar results, at a price of being both at a low spatial resolution.

D. Application to Real Multidate LANDSAT 8 Images

Finally, to illustrate the reliability of the proposed CD framework, a pair of real LR-MS and HR-PAN images acquired at different dates (thus complying with Scenario 3 considered above) has been analyzed. These images $\mathbf{Y}_{\mathrm{LR}}^{t_2}$ and $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ have been acquired by LANDSAT 8 over the Lake Tahoe region (CA, USA) on April 15 and September 22, 2015, respectively. The LR-MS image $\mathbf{Y}_{\mathrm{LR}}^{t_2}$ is of size 175×180 characterized by a spatial resolution of 30 m. According to the spectral response of the LANDSAT 8 sensor [41], the HR-PAN image $\mathbf{Y}_{\mathrm{HR}}^{t_1}$ is of size 350×360 with a spatial resolution of 15m and has a spectral range from 0.5 to 0.68 μm covering three bands of the LR-MS image. Fig. 8(a) and (b) shows the two multidate LR-MS and HR-PAN images that have been manually geographically aligned. The resulting CD binary masks recovered by the proposed fusion-based approach and the worst case approach both performing CVA CD are depicted in Fig. 8(c)–(e). For this pair of images, the ground-truth information (i.e., in terms of a binary map of actual changes) is not available. However, a visual inspection reveals that all methods succeed in recovering the most significant changes between the two images, namely, the pixels corresponding to the lake drought. Nevertheless, the proposed fusion approach at HR has the huge advantage of providing CD binary masks at HR, which helps to detect finer details than the worst case method, as illustrated by the zoomed-in-view regions in Fig. 8(f)–(j).

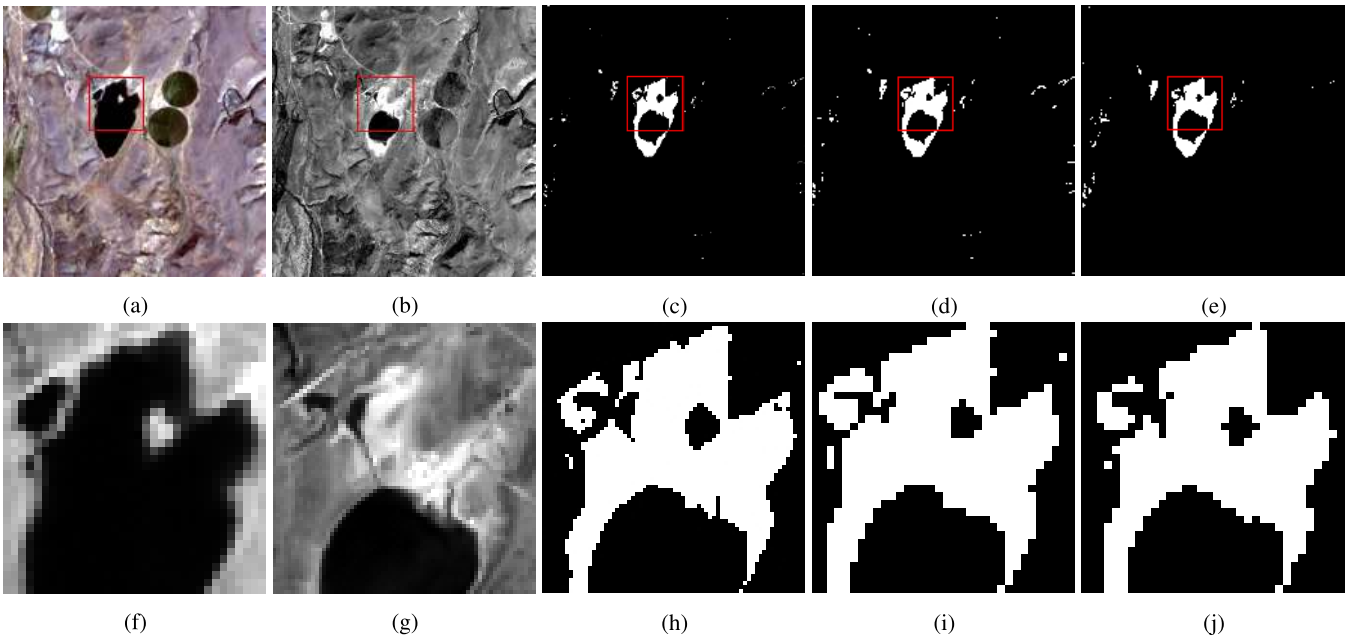


Fig. 8. Real scenario (LR-MS and HR-PAN). (a) LR-MS observed image $\mathbf{Y}_{\mathrm{LR}}^{t_2}$. (b) HR-PAN observed image $\mathbf{Y}_{\mathrm{HR}}^{t_1}$. (c) Change mask $\hat{\mathbf{D}}_{\mathrm{HR}}$. (d) Change mask $\hat{\mathbf{D}}_{\mathrm{aLR}}$. (e) Change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the worst case approach. (f)–(j) Zoomed-in-view versions of the regions delineated in red in (a)–(e).

VII. Conclusion

This paper introduced an unsupervised CD framework for handling multiband optical images of different modalities, i.e., with different spatial and spectral resolutions. The framework was based on a three-step procedure. The first step performed the fusion of the two different spatial/spectral resolution multiband optical images to recover a pseudolatent image of high spatial and spectral resolutions. From this fused image, the second step generated a pair of predicted images with the same resolutions as the observed multiband images. Finally, standard CD techniques were applied to each pair of observed and predicted images with the same spatial and spectral resolutions. The relevance of the proposed framework was assessed thanks to an experimental protocol. These experiments demonstrated the accuracy of the recovered HR CD map.

Future work will include the generalization of the proposed framework to deal with images of other modalities. Indeed, the newly proposed three-step procedure (*fusion*, *prediction*, and *detection*) is expected to be applicable, provided that a physically based direct model can be derived to relate the observed images with a pseudolatent image. A particular instance of the proposed general framework has been derived in [42] to specifically handle multiband optical images when an explicit image formation model is available.

References

[1] A. Singh, "Review Article Digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, Jun. 1989.

[2] M. K. Ridd and J. Liu, "A comparison of four algorithms for change detection in an urban environment," *Remote Sens. Environ.*, vol. 63, no. 2, pp. 95–100, 1998.

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

[4] F. Bovolo and L. Bruzzone, "The time variable in data fusion: A change detection perspective," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 8–26, Sep. 2015.

[5] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.

[6] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[7] J. B. Campbell and R. H. Wynne, *Introduction to Remote Sensing*, 5th ed. New York, NY, USA: Guilford Press, 2011.

[8] C. Collet, J. Chanussot, and K. Chehdi, *Multivariate Image Processing: Methods and Applications*. Hoboken, NJ, USA: Wiley, 2006.

[9] C. Elachi and J. J. van Zyl, *Introduction to the Physics and Techniques of Remote Sensing* (Wiley Series in Remote Sensing and Image Processing), 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[10] J. C. Price, "Spectral band selection for visible-near infrared remote sensing: Spectral-spatial resolution tradeoffs," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 5, pp. 1277–1285, Sep. 1997.

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

[12] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2196–2212, May 2012.

[13] 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, 1998.

[14] 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.

[15] M. J. Canty, A. A. Nielsen, and M. Schmidt, "Automatic radiometric normalization of multitemporal satellite imagery," *Remote Sens. Environ.*, vol. 91, nos. 3–4, pp. 441–451, Jun. 2004.

[16] J. Inglada and A. Giros, "On the possibility of automatic multisensor image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2104–2120, Oct. 2004.

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

[18] V. Alberga, M. Idrissa, V. Lacroix, and J. Inglada, "Performance estimation of similarity measures of multi-sensor images for change detection applications," in *Proc. IEEE Int. Workshop Anal. Multi-Temporal Remote Sens. Images (MultiTemp)*, Leuven, Belgium, Jul. 2007, pp. 1–5.

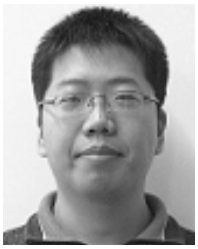
[19] G. Mercier, G. Moser, and S. Serpico, "Conditional copula for change detection on heterogeneous SAR data," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, Jul. 2007, pp. 2394–2397.
[20] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J. Y. Tourneret, "A new multivariate statistical model for change detection in images acquired by homogeneous and heterogeneous sensors," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 799–812, Mar. 2015.
[21] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "Performance assessment of a recent change detection method for homogeneous and heterogeneous images," in *Revue Française de Photogrammétrie et de Télédétection*, vol. 209. 2015, pp. 23–29.
[22] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A Bayesian nonparametric model coupled with a Markov random field for change detection in heterogeneous remote sensing images," *SIAM J. Imag. Sci.*, vol. 9, no. 4, pp. 1889–1921, 2016.
[23] V. Alberga, M. Idrissa, V. Lacroix, and J. Inglada, "Comparison of similarity measures of multi-sensor images for change detection applications," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2007, pp. 2358–2361.
[24] M. N. Klaric *et al.*, "GeoCDX: An automated change detection and exploitation system for high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2067–2086, Apr. 2013.
[25] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
[26] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
[27] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.
[28] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.
[29] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.
[30] N. Yokoya, N. Mayumi, and A. Iwasaki, "Cross-calibration for data fusion of EO-1/hyperion and Terra/ASTER," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 419–426, Apr. 2013.
[31] F. Heide *et al.*, "FlexISP: A flexible camera image processing framework," *ACM Trans. Graph.*, vol. 33, no. 6, Nov. 2014, Art. no. 231.
[32] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions* (Monographs and Surveys in Pure and Applied Mathematics), vol 4. London, U.K.: Chapman & Hall, 1999.
[33] J. Idier, Ed., *Bayesian Approach to Inverse Problems* (Digital Signal and Image Processing Series). London, U.K.: Wiley, 2008.
[34] 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.
[35] A. D'Addabbo, G. Satalino, G. Pasquariello, and P. Blonda, "Three different unsupervised methods for change detection: An application," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, vol. 3. Sep. 2004, pp. 1980–1983.
[36] D. Lu, P. Mausel, E. Brondízio, and E. Moran, "Change detection techniques," *Int. J. Remote Sens.*, vol. 25, no. 12, pp. 2365–2401, Jun. 2004.
[37] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
[38] J. M. P. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[39] D. C. Heinz and C.-I. Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 529–545, Mar. 2001.
[40] V. Ferraris, N. Dobigeon, Q. Wei, and M. Chabert, "Detecting changes between optical images of different spatial and spectral resolutions: A fusion-based approach—Complementary results," Dept. Signal Commun., Univ. Toulouse, Toulouse, France, Tech. Rep., Dec. 2016. [Online]. Available: http://www.enseeiht.fr/~dobigeon/papers/Ferraris_TechReport_2016.pdf
[41] USGS. (2017). *Landsat 8*. [Online]. Available: https://landsat.usgs.gov/landsat-8
[42] V. Ferraris, N. Dobigeon, Q. Wei, and M. Chabert, "Robust fusion of multiband images with different spatial and spectral resolutions for change detection," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 175–186, Jun. 2017.

Vinicius Ferraris (S'16) was born in Rio de Janeiro, Brazil, in 1991. He received the Engineering degree in electronics engineering from ENSEEIHT, Toulouse, France, in 2013, the M.Sc. degree in signal processing from the National Polytechnic Institute of Toulouse, Toulouse, in 2013, and the Degree (Hons.) in electrical engineering from the Federal University of Rio Grande do Norte, Natal, Brazil, in 2014. He is currently pursuing the Ph.D. degree with the IRIT Laboratory, Signal and Communications Group, National Polytechnic Institute of Toulouse.

His research interests include statistical signal and image processing, with a particular focus on Bayesian inverse problems with applications to remote sensing imaging.

Nicolas Dobigeon (S'05–M'08–SM'13) was born in Angoulême, France, in 1981. He received the Engineering degree in electrical engineering from ENSEEIHT, Toulouse, France, in 2004, and the M.Sc. and Ph.D. degrees in signal processing from Toulouse INP, University of Toulouse, Toulouse, in 2004 and 2007, respectively.

From 2007 to 2008, he was a Post-Doctoral Research Associate with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA. Since 2008, he has been with Toulouse INP, where he is currently a Professor. Since 2017, he has been a Junior Member of the Institut Universitaire de France, Paris, France. He conducts his research within the Signal and Communications Group, IRIT Laboratory, University of Toulouse, where he is also an Affiliated Faculty Member with the TeSA Laboratory. His research interests include statistical signal and image processing, with a particular focus on Bayesian inverse problems with applications to remote sensing and biomedical imaging.



Qi Wei (S'13–M'15) was born in Shanxi, China, in 1989. He received the bachelor's degree in electrical engineering from Beihang University, Beijing, China, in 2010, and the Ph.D. degree in signal and image processing from the National Polytechnic Institute of Toulouse (INP-ENSEEIHT), University of Toulouse, Toulouse, France, in 2015, where his thesis Bayesian Fusion of Multi-Band Images: A Powerful Tool for Super-Resolution was rated as one of the best theses.

From 2015 to 2016, he was a Research Associate with the Signal Processing Laboratory, Department of Engineering, University of Cambridge, Cambridge, U.K., where he was involved in multiband image processing. Since 2016, he has been a Research Associate with Duke University, Durham, NC, USA, along with Prof. L. Carin. His research interests include machine/deep learning, computer vision/image processing, and Bayesian statistical inference.

Dr. Wei was a recipient of the Prix Léopold Escande.

Marie Chabert (M'10) received the Engineering degree in electronics and signal processing from ENSEEIHT, Toulouse, France, in 1994, and the M.Sc. and Ph.D. degrees in signal processing and the Habilitation a Diriger les Recherches from the National Polytechnic Institute of Toulouse, Toulouse, in 1994, 1997, and 2007, respectively.

She is currently a Professor of signal and image processing with INPT-ENSEEIHT, University of Toulouse, Toulouse. She is conducting research with the Signal and Communication Team, Institut de Recherche en Informatique de Toulouse, Toulouse. Her research interests include nonuniform sampling, time–frequency diagnosis and condition monitoring, and statistical modeling of heterogeneous data in remote sensing.