Title: A Novel Change Detection Method for Multitemporal Hyperspectral Images Based on Binary Hyperspectral Change Vectors

Author(s): Daniele Marinelli, Francesca Bovolo, Lorenzo Bruzzone

A Novel Change Detection Method for Multitemporal Hyperspectral Images Based on Binary Hyperspectral Change Vectors

Daniele Marinelli, *Student Member, IEEE,* Francesca Bovolo, *Senior Member, IEEE,*
Lorenzo Bruzzone, *Fellow, IEEE*

Abstract—**Hyperspectral (HS) images provide dense sampling of target spectral signatures. Thus, they can be used in a multitemporal framework to detect and discriminate between different kinds of fine spectral change effectively. However, due to the complexity of the problem and the limited amount of multitemporal images and reference data, only few works in the literature addressed Change Detection (CD) in HS images. In this paper we present a novel method for unsupervised multiple CD in multitemporal HS images based on a discrete representation of the change information. Differently from the state of the art methods, that address the high dimensionality of the data using band reduction or selection techniques, in this work we focus our attention on the representation and exploitation of the change information present in each band. After a band-by-band pixel-based subtraction of the multitemporal images, we define the Hyperspectral Change Vectors (HCVs). The change information in the HCVs is then simplified. To this end, the radiometric information of each band is separately analyzed to generate a quantized discrete representation of the HCVs. This discrete representation is explored by considering the hierarchical nature of the changes in HS images. A tree representation is defined and used to discriminate between different kinds of change. The proposed method has been tested on a simulated dataset and two real multitemporal datasets acquired by the Hyperion sensor over agricultural areas. Experimental results confirm that the discrete representation of the change information is effective when used for unsupervised CD in multitemporal HS data.**

Index Terms—**Change Detection, Hyperspectral Images, Multitemporal Images, Binary Codewords.**

I. INTRODUCTION

The analysis of the natural or anthropogenic land-cover dynamics is fundamental for environmental monitoring. In this context, Earth Observation satellites provide a powerful tool for the analysis of changes on the Earth surface due to their point of view and their capability of regularly acquiring multiple images over the same area at different times. Thus, multitemporal remote sensing images can be effectively used to perform Change Detection (CD), i.e., to identify the land cover changes occurred between two or more images acquired over the same area at different times [1].

In the literature, CD methods exist for multitemporal images acquired both by passive optical and active Synthetic Aperture Radar (SAR) systems. Here we focus on the former one. CD methods for optical passive data can be split into two main categories: i) binary CD [2]–[11], and ii) multiple CD [12]–[23]. Binary CD consists in the discrimination between changed and unchanged samples only. It can be performed either in the spectral band domain using all the spectral channels or by manually selecting the relevant ones [9], or by working in a transformed feature space. The most common approach is to apply a threshold to the magnitude of the difference image (obtained by pixel-by-pixel subtraction of each spectral channel of the multitemporal images). The statistical distribution of the magnitude variable can be modelled either in parametric [2] or non parametric ways [24]. The decision strategy can be designed in the Bayesian framework according to minimum error [2] or minimum cost [11] rules. Other methods are based on the use of genetic algorithms [3], fuzzy clustering [4], data fusion techniques [5] or feature level fusion techniques [6]. Pixel based results can be improved by taking into account the spatial correlation. Examples of this include Hopfield-Type neural networks [7], self-organizing feature map neural network [8] or Markov Random Fields [10].

Multiple CD methods allow for the discrimination between multiple kinds of change. Existing methods can be either supervised [12]–[19] or unsupervised [20]–[23]. Supervised methods use training data to achieve accurate CD results. Among them we recall: i) post classification comparison [12]–[14]; ii) direct multidate classification [15], [16]; and iii) compound classification [17]–[19]. A critical analysis of these kinds of approach can be found in [25], [26]. The key issue is that multitemporal labelled reference data are seldom available thus making unsupervised methods potentially more useful also in the more complex multiple CD case. Some of the existing unsupervised multiple CD methods are based on transformations such as Independent Component Analysis (ICA) [20] or multivariate analysis techniques (e.g., Multivariate Alteration Detection [21]). Others are based on image differencing and thus Change Vector Analysis (CVA) [22], [23]. For example in [22] the authors use a 2-D polar representation of the Hyperspectral Change Vectors (HCVs) assuming the use of 2 bands only. In [23] the authors propose a compressed version of the CVA (C^2VA) to obtain a 2-D compressed representation of the multidimensional HCVs (i.e., they compress the information of more bands into a 2-D space).

Those methods have been designed for multispectral (MS) images where only changes inducing large spectral variations can be observed. However, changes may induce variations in small portions of the spectrum and there may be groups of changes that differ significantly only in a short range of

wavelengths [27]. Their detection and discrimination requires an accurate representation of the change signature (i.e., the HCV). To this end, a dense sampling of the electromagnetic spectrum is required. HS sensors show narrow spectral interval between adjacent bands and thus give a near-continuous representation of the spectral signature for each pixel. Since the availability of multitemporal HS data is limited, poor attention has been devoted to CD in HS images. This is even more true after the Hyperspectral Imager for the Coastal Ocean (HICO) failure due to a solar storm in September 2014 and the NASA EO-1 satellite (which mounted on board the first spaceborne Hyperion HS sensor) deactivation in March 2017. However, in the next years new satellite missions with HS sensors will be launched thus making multitemporal HS data a promising tool for CD applications. Among them we mention: 1) "PRecursore IperSpettrale della Missione Applicativa" (PRISMA) which will be launched in 2019 [28]; 2) "Hyperspectral Imager Suite" (HISUI) planned to be launched in 2019 [29]; 3) "Environmental Mapping and Analysis Programme" (EnMAP) planned for launch in 2020 [30]; 4) "Hyperspectral Infrared Imager" (HyspIRI) scheduled for launch after 2022 [31]; and 5) "Spaceborne Hyperspectral Applicative Land and Ocean Mission" (SHALOM) scheduled for launch after 2022 [32].

The detailed spectral information in HS data comes at a cost of an increase complexity when performing CD on HS images. The main challenge is related to the high dimensionality of HS data. When dealing with a high dimensional space, the effectiveness of methods developed for a low dimensional space decreases due to problems such as reduced efficiency of distance metrics [33] and the Hughes phenomenon [34]. These problems are related to the sparsity of the feature space when the number of dimensions increases. Dimensionality reduction approaches such as Principal Component Analysis (PCA) [35] or C^2VA [23] designed for MS data can be applied to HS images. However, while reducing the dimensionality they fail in preserving relevant change information. Another issue of HS data is related to redundancy. This is due to the fact that adjacent bands (in terms of wavelength) may contain similar information due to the narrow sampling of the spectrum. This makes the relevant change information more sparse [27].

To this date, the number of methods for CD in HS images is still quite limited. CD is carried out comparing the spectral signature of the same pixel at the two dates using spectral comparison measures such as the Spectral Angle Mapper (SAM) [36], the Spectral Information Divergence (SID) [37] or a composition of the two measures [38]. Other methods are based on transformations such as covariance equalization [39] to perform detection of changes that typically do not occur naturally (e.g., moving of targets). Anomaly CD has been performed also using Cluster Kernel Reed–Xiaoli (CKRX) algorithm [40] or by statistically modeling the problem using a multivariate Gaussian model [41]. Another type of transformation methods are those that perform the CD in transformed spaces. In [42] the authors use the ICA to separate the changes in different components. In [43] the PCA is applied to the two stacked multitemporal images to perform a Temporal PCA (TPCA). In [44] the author proposes an iterative Multivariate Alteration Detection (MAD) algorithm where at each iteration

the method focuses more on the samples for which the change status is not clear. C^2VA has been also used in HS images both in unsupervised [27] and supervised [45] manner. In [27] the authors mitigate the loss of information resulting from the compression of the high dimensional data in a 2-D space by applying the C^2VA iteratively and hierarchically. In [45] the authors follows an hierarchical approach where the changes are manually selected at each iteration in the compressed 2-D feature space. Other works are based on unmixing, which considers the possibility of having different materials inside a single pixel (due to the relatively low spatial resolution of HS sensors) [46]–[48]. The idea is to extract the different end-members of change to guide the change identification and detection process. Preliminary studies preparatory to this work already addressed unsupervised CD [49], [50]. Both provide only a partial solution to the problem and work on a simplified band-by-band representation of the change information. Further they showed to be unable to: i) adapt to the different information content of spectral channels [49] and ii) deal with the different levels of intensity variations among types of change (i.e., they do not allow for different levels of sensitivity) [49], [50].

The analysis of the literature points out that most of the existing methods address the high complexity of CD in HS images by applying transformations or feature reduction techniques. This is done to reduce: i) the dimension of the features space and thus the computational burden; and ii) the noise impact and thus the number of errors. This often results in a loss of change information. Moreover, most of the existing feature reduction techniques are developed for single date analysis and are often supervised thus limiting their use in a multitemporal framework. Indeed, in the multitemporal case change information showing low prior probability should be preserved and unsupervised method are preferred being multitemporal ground truth seldom available. Little work has been done regarding the definition of alternative effective representations of change that focus on the relevant change information of each band of the Hyperspectral Change Vectors (HCVs). Accordingly, we work on the HCVs to highlight the change information by representing the HCVs with binary codewords. For a given HCV, each spectral channel is represented using Q bits (where in HS images typically $Q \geq 10$) to code all the possible radiometric values. In the new representation we can reduce the number of bits used for each band thus moving from 2^Q (e.g., $2^{12} = 4096$) to few values that represent the essential change information in each channel. To this end, the method first performs a binary CD to discriminate between changed and unchanged HCVs. It then focuses only on the changed HCVs (the ones with a high magnitude value induced by change impact on one or more spectral channels) in order to convert them into binary codewords working on each band separately. The method adaptively quantizes the values of the HCVs at the considered band so that different kinds of change fit to different quantization intervals. In such a way, the behaviour of each change (or group of changes) is modeled accurately in each band and coded using binary strings. The fusion of the band-by-band binary representations provides a codeword that models the change information content of

the full HCV. Thus, despite binarization is conducted on each band separately, the correlation of the change information in adjacent spectral bands is preserved. The binary codewords are then compressed across bands to reduce the remaining redundant information. Finally, the method organizes the codewords in a dendrogram (i.e., a tree structure) by means of an agglomerative hierarchical clustering. The hierarchical structure allows us to obtain different results in terms of level of detail of the CD by cutting the dendrogram at different heights. The entire process is unsupervised.

The main novel contributions of the proposed method are:

- An adaptive band-dependent model to represent the information of multiple kinds of change in multitemporal HS images.
- A compressed representation of the multitemporal information associated to changes that is simple, efficient and reduces data volume.
- An adaptive mechanism to discriminate between the kinds of change depending on the selected sensitivity.

The method has been tested on a simulated HS dataset and on two real multitemporal HS datasets acquired by the Hyperion sensor over two agricultural areas in Washington state (USA) and in the Albacete province (Spain). Experimental results confirmed the effectiveness of the proposed method. The rest of the paper is divided in 3 Sections. Section II describes in detail the proposed method. Section III presents and analyzes the experimental results. Finally Section IV draws the conclusion.

II. PROPOSED METHOD FOR UNSUPERVISED CHANGE DETECTION IN HYPERSPECTRAL DATA

Figure 1 shows the block scheme of the proposed method that consists of three main steps: i) binary CD; ii) Hyperspectral Change Vectors Binary Coding; iii) Compressed Binary Hyperspectral Change Vectors analysis. The binary CD separates changed from unchanged samples. After that the method focuses on the changed samples only. The coding of the changed HCVs first adaptively quantizes the values of the HCVs in each band to simplify the representation of the change information. The codeword definition and compression steps allow us to obtain binary codewords coding the change information contained in all the spectral channels. Finally, the Compressed Binary Hyperspectral Change Vectors (CBHCVs) analysis separates the different kinds of change.

Let I_1 and I_2 be two hyperspectral images acquired at times t_1 and t_2, respectively with $I_1 = \{I_{1,b}\}_{b=1}^{B}$ and $I_2 = \{I_{2,b}\}_{b=1}^{B}$, where $I_{1,b}$ and $I_{2,b}$ are the single band images representing the b-th spectral channel. We define $\Omega = \{\omega_u, \Omega_C\}$ as the set of classes containing both unchanged (ω_u) and changed (Ω_C) pixels. $\Omega_C = \{\omega_1, \omega_2, \ldots, \omega_J\}$ is the macro-class of changed pixels that can be further divided into J classes, one for each type of change.

A. Binary Change Detection

The first step of the method is the binary CD that discriminates between changed and unchanged pixels. To this end, we follow a state of the art approach based on the statistical analysis of the magnitude of the difference image which has been extensively used in the literature [2], [11], [22], [51]. However, note that other binary CD strategies could be used. First we compute the HS difference image as $I_D = I_2 - I_1$ and its magnitude I_ρ as:

$$I_\rho = \sqrt{\sum_{b=1}^{B} (I_{D,b})^2}. \tag{1}$$

It is expected that unchanged pixels show a small magnitude value whereas changed one show large magnitude. Thus we analyze the Probability Density Function (PDF) of I_ρ in order to discriminate between ω_u and Ω_C. According to the literature, we approximate the PDF of the magnitude as a mixture of two parametric distributions. Unchanged pixels are approximated as Rayleigh distributed whereas the changed ones are approximated as Rice distributed [52]. After estimating the parameters of the two models using an Estimation Maximization algorithm, according to Bayes theory it is possible to determine a threshold T_ρ such that pixels below T_ρ belong to class ω_u, whereas pixels above T_ρ belong to class Ω_C. From here on we focus on the set $\{\mathbf{s}_n\}_{n=1}^{N}$ of N changed HCVs ($\mathbf{s}_n \in \Omega_c, \forall n = 1, \ldots, N$), where $\mathbf{s}_n = \{s_{n,b}\}_{b=1}^{B}$ is the n-th HCV and $s_{n,b}$ the b-th spectral channel of $\mathbf{s}_n$.

B. Hyperspectral Change Vectors Binary Coding

In the second step we move from the real valued representation of the HCVs usually employed in the literature to the novel binary codeword based one. The aim is to provide a simpler yet effective representation of the HCVs focused on the information associated with different kinds of change and thus able to improve their separability. Figure 2a shows that kinds of change in the original HCV space (i.e., the real valued space) can be discriminated more or less effectively depending on the considered band. As an example, in band b_y (Figure 2a) it is possible to separate between two groups of changes (ω_1, ω_2 and ω_3, ω_4) whereas in band b_z the four changes are indistinguishable from each other. This illustrates that bands may contain significantly different change information with some of them (e.g., bands b_x and b_y) being more relevant for the CD problem with respect to others (e.g., band b_z) where all the changes show similar values and cannot be separated. Therefore, we adaptively analyze each band to detect the information that can be useful for the change discrimination. At the same time we represent each HCV in a more efficient and compact way. In particular, we move from the common real valued representation to the proposed binary representation by assigning to each HCV $\mathbf{s}_n$ a binary codeword referred to as Compressed Binary Hyperspectral Change Vector (CBHCV) $\mathbf{cp}_n$. This is achieved by: 1) *Quantization*; 2) *Binary Codeword Definition*; 3) *Compression*.

1) Quantization: It identifies, by working separately on each band, the information that allows us to separate the different changes. We consider the content of a band relevant for solving the CD problem if we can identify more than one cluster of pixels, where each cluster contains the samples related to a given kind of change or a group of kinds of change.
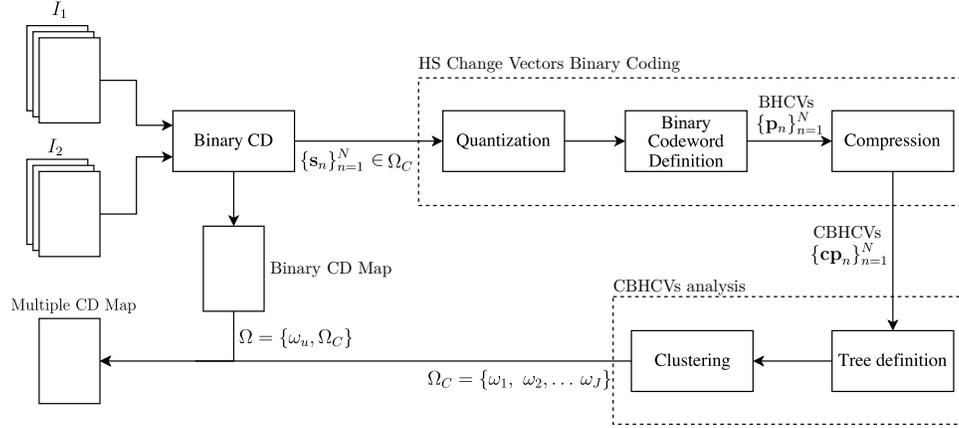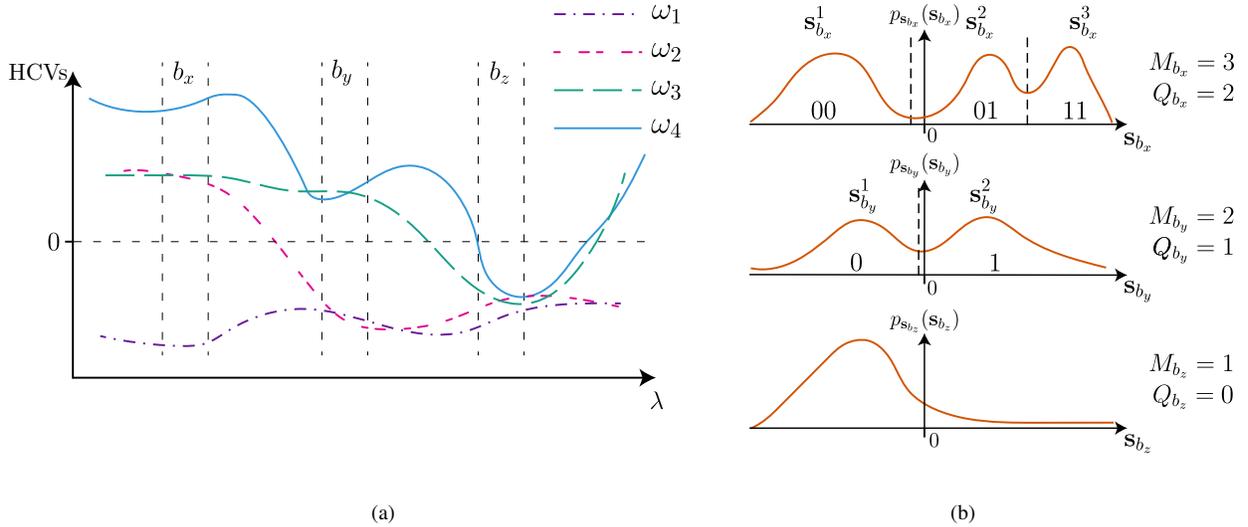
Fig. 1: Block scheme of the proposed method.

(a)   (b)

Fig. 2: Qualitative illustration of the quantization step of the proposed technique: (a) HCVs of 4 pixels belonging to 4 different kinds of change (ω_{1-4}). (b) Probability Density Functions of the values of the changed pixels for the 3 bands. The dashed lines identify the boundaries of the quantization intervals with the corresponding binary strings.

This is reflected in the PDF of the values of $\mathbf{s}_b = \{s_{n,b}\}_{n=1}^N$ (i.e., values of the b-th component of the HCVs which are associated with band b) since samples that fall inside the same mode are likely to belong either to the same kind of change or different changes that in turn are indistinguishable using the informative content of the considered band. In contrast, samples that fall inside different modes are likely to belong to different kinds of change. Thus, we are interested in identifying these modes. Figure 2b shows the PDFs of the values in $\mathbf{s}_b$ for the 3 different bands b_x, b_y, b_z. In band b_z only one mode exists whereas for bands b_x and b_y more modes can be identified. Therefore, we can simplify the radiometric information related to changes of each band by quantizing the $s_{n,b}$ in the same mode into one single quantization interval. Since different bands contain different change information, the proposed quantization step has to be performed adaptively in order to properly preserve and capture the change information. A uniform band-by-band quantization [49] would lead to split samples belonging to the same change class into different quantization levels and thus to a poor representation.

To apply the quantization, we identify and separate the modes so that each one corresponds to a quantization interval. To this end, first we estimate the PDF $p_{\mathbf{s}_b}(\mathbf{s}_b)$ of $\mathbf{s}_b$ using a Kernel Density Estimator (KDE) [24] (however, any other estimation methods can be used). Then we search for the M_b modes of $p_{\mathbf{s}_b}(\mathbf{s}_b)$ by identifying the local maxima and we define M_b quantization intervals. Note that M_b can vary for different spectral channels. If only one mode is identified (i.e., $M_b = 1$), the band is discarded (e.g., band b_z in Figure 2b). The boundaries of each quantization interval are positioned at the minimum between two modes with the corresponding interval as large as the distance between the two minima. Each $s_{n,b} \in \mathbf{s}_b$ is assigned to the corresponding quantization interval thus obtaining M_b sets $\{\mathbf{s}_b^m\}_{m=1}^{M_b}$ (see Figure 2b). The PDF estimation is intrinsically subject to errors. Moreover, typically the modes of the PDF are overlapped. Therefore, it is likely that the quantization leads to errors such as that pixels belonging to the same kind of change are split by a wrongly positioned boundary. Quantization errors mitigation will be addressed in the following. The proposed strategy results in

better performance than other approaches like those based on clustering.

2) Codeword Definition: After the quantization step, we move to the binary representation of the quantized intervals. For a given band, we assign a different binary string to each quantization interval and thus to all the pixels in the corresponding set $\mathbf{s}_b^m$. For each band b, the length Q_b of the corresponding binary string depends on the number of quantization intervals M_b. We compute the number of bits as $Q_b = \lceil \log_2 M_b \rceil$. Note that since $2^{Q_b-1} < M_b \leq 2^{Q_b}$ some combinations of binary strings may remain unused. The binary strings are coded according to the Gray code such that adjacent quantization intervals (i.e., modes) differ by only one bit at the considered band. This reduces the impact of the errors introduced in the quantization step in the binary codeword representation. Figure 2b shows an example of the binary string assigned to each quantization interval for the two bands bands b_x, b_y.

Concatenating the binary strings assigned to a pixel in the B bands, we obtain its codeword which codes the change information of the corresponding HCV. By applying the codeword definition for each band to each $\mathbf{s}_n \in \{\mathbf{s}_n\}_{n=1}^N$ we obtain a set of codewords $\{\mathbf{p}_n\}_{n=1}^N$ called Binary Hyperspectral Change Vectors (BHCVs). Each $\mathbf{p}_n = \{p_{n,k}\}_{k=1}^K, (p_{n,k} \in \{0,1\})$ is a binary string of length $K = \sum_{b=1}^B Q_b$. In the I_D the number Q of bits used for each band is constant and thus the total number of bits required to code a HCV is $Q \times B$. Since typically $Q \geq 10$ whereas Q_b is much smaller (we expect $Q_b \leq 3$ in most of the cases), then $K << Q \times B$.

3) Compression: The last step of the coding of the HCVs is the compression of the codewords in $\{\mathbf{p}_n\}_{n=1}^N$. Typically HS data are highly redundant due to the narrow sampling of the spectrum. This means having large number of samples showing very similar values at multiple bands. When moving to the codeword based representation, redundancy still exists. In this domain, there is redundancy when a large portion of the samples in set $\mathbf{s}_b$ is quantized identically (i.e., the same samples are split into the same groups) at multiple bands. In terms of binary representation, this translates in large number of samples that share the same value for more than one bit of the codeword. Thus, we propose to analyze the redundancy in terms of groups of bits (i.e., redundant bits) that have equal values for large number of samples.

In order to search for the redundant bits, first we sort them along the k index (i.e., we swap the position k of the bits) in such a way that groups of bits showing the same value for a large number of samples are positioned near each other. The sorting algorithm [53] first produces a tree structure with K leafs where the k-th leaf represents the binary vector of length N containing the values of the k-th bit of all the codewords in $\{\mathbf{p}_n\}_{n=1}^N$. It then applies an efficient leaf ordering method that aims at maximizing the sum of the similarities of the elements of adjacent leafs (note that others sorting algorithm could be used). We obtain a new set $\{\mathbf{p}_n'\}_{n=1}^N$ (with $\mathbf{p}_n' = \{p_{n,k}'\}_{k=1}^K$) of BHCVs that contains exactly the same change information of the set $\{\mathbf{p}_n\}_{n=1}^N$ with the bits sorted according to similarity. Tables Ia and Ib show an example with a set of 5 BHCVs before and after the ordering, respectively. The ordering allows us to simplify the identification of redundant bits since set $\{\mathbf{p}_n'\}_{n=1}^N$ is defined so that they are positioned near each other. We can now identify these groups by evaluating the redundancy of adjacent bits. To this end, we compute the sum of all the N pairwise Hamming distances of two bits k and $k+1$ as:

$$r(k, k+1) = \sum_{n=1}^N |p_{n,k+1}' - p_{n,k}'|. \tag{2}$$

Distance r will show small values when the pair of adjacent bits $(k, k+1)$ contains redundant information whereas it will show large values when the change information of the two bits is not redundant. The last rows of both Table Ia and Table Ib show the Hamming distances computed between each adjacent pair of bits. Note that in Table Ia the values of r are much larger and distributed with respect to Table Ib. This highlights how the bit ordering effectively positions near each other bits with similar information.

We analyze the vector r to find sequences of small values (i.e., groups of redundant bits) surrounded by large r values (i.e., adjacent pairs not redundant). We consider two adjacent bits k and $k+1$ redundant if $r(k, k+1) \leq T_r, (T_r > 0)$ and not redundant if $r(k, k+1) > T_r$. By applying the threshold to the distances we identify the groups of redundant bits (i.e., groups of adjacent bits for which $r \leq T_r$) and those that are likely to be not redundant an thus that contain relevant and unique change information. The dashed lines in Table Ib separates the different groups of bits that have been identified as redundant, the ones that can be compressed. Here we propose to work separately on each BHCV and compress each group of redundant bits into one bit using a majority rule. This allows us not only to reduce the redundant information, but also the number of errors thus reducing the number of outliers. Indeed, the majority rule can filter out outliers (e.g., a bit that in very few samples shows a different value with respect to the other bits of the redundant group). The new set $\{\mathbf{cp}_n\}_{n=1}^N$ is composed by binary strings defined as Compressed Binary Hyperspectral Change Vectors (CBHCVs) where each $\mathbf{cp}_n = \{cp_{n,i}\}_{i=1}^I, (I << K)$. Note that the compression rate K/I is strongly dependent on the set $\{\mathbf{p}_n'\}_{n=1}^N$ and the amount of redundant information it contains. Table Ic shows the CBHCVs of the corresponding sorted BHCVs of Table Ib. Note that unlike existing dimensionality reduction methods (e.g., [54]–[57]), the proposed compression of BHCVs is specifically designed to exploit the simplified binary representation of the change information.

C. Compressed Binary Hyperspectral Change Vectors Analysis

In the last phase we exploit the simplified binarized representation of the change information (i.e., the codewords) to discriminate between the different kinds of change. HS data allow us to discriminate between both major land cover changes and more subtle changes [27]. The sensitivity required in the discrimination between the kinds of change may vary depending on the considered application. Indeed, whereas in some cases one may want to discriminate between all the

TABLE I: Compression of the codewords of five pixels: (a) BHCVs before the bit sorting showing the Hamming distance r between adjacent pairs of bits. (b) BHCVs after the bit sorting. (c) Codewords after the compression (i.e., CBHCVs).

(a)

$\mathbf{p}_1$	1	1	1	0	1	1	1	1	1
$\mathbf{p}_2$	1	1	1	0	1	1	1	1	1
$\mathbf{p}_3$	0	1	1	0	0	0	1	0	0
$\mathbf{p}_4$	0	1	0	1	0	0	1	0	0
$\mathbf{p}_5$	0	1	0	1	0	0	1	0	0
r	3	2	5	4	0	3	3	0	

(b)

$\mathbf{p}'_1$	1	1	1	1	1	1	1	1	0
$\mathbf{p}'_2$	1	1	1	1	1	1	1	1	0
$\mathbf{p}'_3$	1	0	0	0	0	0	1	1	0
$\mathbf{p}'_4$	0	0	0	0	0	0	1	1	1
$\mathbf{p}'_5$	0	0	0	0	0	0	1	1	1
r	1	0	0	0	0	3	0	3	

(c)

$\mathbf{cp}_1$	1	1	0
$\mathbf{cp}_2$	1	1	0
$\mathbf{cp}_3$	0	1	0
$\mathbf{cp}_4$	0	1	1
$\mathbf{cp}_5$	0	1	1

possible kinds of change, this may not be the case for other applications in which the main aim is to detect specific land cover changes only. For this reason, we propose to represent the set of changed samples in a hierarchical tree structure defined using an agglomerative hierarchical clustering approach. This representation allows us to analyze the tree at different depths which correspond to different sensitivities in the discrimination of the kinds of change.

The HCVs coding phase resulted in a simplified representation of the change information. Thus, we expect to have a large number of changed pixels that share the same CBHCVs. First, we select only the unique CBHCVs $\{\mathbf{ucp}_u\}_{u=1}^{U}$ and we compute the corresponding prior probabilities as:

$$P_u = \frac{|\{n \in \{1, \ldots, N\} : \mathbf{cp}_n = \mathbf{ucp}_u\}|}{N}. \tag{3}$$

We can assume that, if the prior probability P_u of an unique CBHCVs is low, the corresponding samples are non interesting outliers or are related to errors in the HCVs coding step (which have been reduced by the compression step). Therefore, we discard the $\mathbf{ucp}_u$ for which $P_u \leq T_p$, where T_p is a threshold on the prior probability. However, it may happen that CBHCVs with very low prior probabilities actually correspond to a real change which occurred only in a very small area. Thus, the value of threshold T_p has to be selected considering the trade-off between the capability of removing noisy samples and outliers and the risk of losing true changes. The hierarchical clustering is then applied only to the $\mathbf{ucp}_u$ for which $P_u > T_p$, which is a set $\{\mathbf{ucp}'_u\}_{u=1}^{U'}$ where $U' \leq U$.

The agglomerative hierarchical clustering is based on the Unweighted Pair Group Method with Arithmetic Mean (UPGMA). The algorithm starts with U' clusters, each corresponding to an unique $\mathbf{ucp}'_u$ and at each iteration it merges the two most similar clusters. Merging is conducted according to the pairwise normalized Hamming distance. Since the compression involves groups of redundant bits with different sizes the Hamming distance is computed in a weighted manner, where the weight of each bit of the CBHCVs is directly proportional to the number of bits that were compressed to generate it. The pairwise distance computation is the most computationally demanding part of the clustering algorithm since it requires the computation of $U'(U' - 1)/2$ distances ($O(n^2)$ complexity). However, since $U' << N$ the computational time is significantly smaller with respect to computing the pairwise distances in the original domain which means computing $N(N-1)/2$ distances ($U'(U'-1)/2 << N(N-1)/2$). At each iteration, after merging the two most similar clusters, the algorithm recomputes the distance between the newly formed one and all the others. Let us consider the first iteration of the algorithm and three generic clusters $\mathbf{ucp}'_a, \mathbf{ucp}'_b, \mathbf{ucp}'_c$ with prior probabilities P_a, P_b, P_c, respectively. Let $d(\mathbf{ucp}'_a, \mathbf{ucp}'_b)$ be the pairwise distance between $\mathbf{ucp}'_a$ and $\mathbf{ucp}'_b$. Let us suppose that we merge clusters $\mathbf{ucp}'_a$ and $\mathbf{ucp}'_b$ (i.e., $[\mathbf{ucp}'_a, \mathbf{ucp}'_b] = \mathbf{ucp}'_a \cup \mathbf{ucp}'_b$ and $P_{a,b} = P_a + P_b$). The distance between the new cluster $[\mathbf{ucp}'_a, \mathbf{ucp}'_b]$ and $\mathbf{ucp}'_c$ can be computed using the linkage criterion defined as:

$$d([\mathbf{ucp}'_a, \mathbf{ucp}'_b], \mathbf{ucp}'_c) =$$
$$= \frac{P_a d(\mathbf{ucp}'_a, \mathbf{ucp}'_c) + P_b d(\mathbf{ucp}'_b, \mathbf{ucp}'_c)}{P_{a,b}} \tag{4}$$

This equation is general and applies to each iteration considering as inputs the distance values computed at the previous iteration. By weighting the distance using as weights P_a and P_b we preserve the information regarding the prior probabilities of each unique CBHCVs. The algorithm iterates until all of the unique CBHCVs are merged into one single group and the tree structure (called dendrogram) is constructed. The length of the branches of the dendrogram is related to the distance between the clusters/$\mathbf{ucp}'_u$ (i.e., a long branch indicates a significant difference between the samples contained in the two clusters). To separate the samples into different kinds of change ω_j, the dendrogram can be cut at a given depth defined by a threshold T_h. Figure 3 shows two examples of dendrogram cut that highlights how by cutting the tree at different depths we separate the changed samples into different numbers of clusters (3 and 6 in the example). By varying T_h we vary the sensitivity of the method to the different kinds of change. Lowering T_h corresponds to increasing the sensitivity of the method to the different kinds of change. Indeed, it is equivalent to a decrease of the threshold on the Hamming distance above which two CBHCVs are considered to belong to different kinds of change. Instead, increasing T_h decreases the sensitivity of the method. Small values of T_h may allow for the discrimination of very similar (in terms of spectral signature) kinds of change, but at the cost of an increased vulnerability to noise and outliers. A large value of T_h may result in the grouping of different changes into the same cluster. Thus, the choice of T_h has to be done taking into
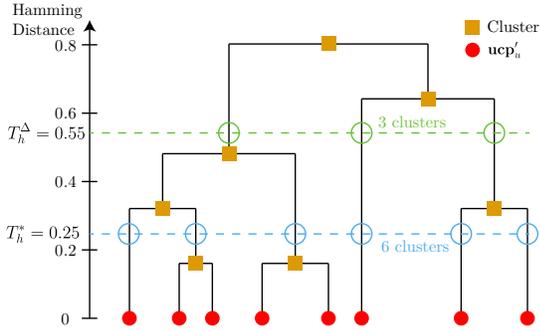
Fig. 3: Example of a dendrogram. Red dots represent leafs and correspond to unique $\mathbf{ucp}'_u$. Orange squares represent the clusters. The green (T_h^Δ) and blue (T_h^*) dashed lines show two examples of cut of the dendrogram.

account the trade-off between sensitivity to the changes and robustness to noise. Moreover, note that the selection of T_h can be easily done also in an iterative way by changing the threshold depending on the required level of detail of the CD map.

When T_h is selected and thus the kinds of change are identified, the samples for which the corresponding $\mathbf{ucp}_u$ shows a $P_u \leq T_p$ have to be assigned to one of the identified changes. To this end, we move back to the HCV representation and, for each sample that has to be assigned, we search for the nearest 50 samples (that have already been assigned to one change). We analyze the 50 nearest samples instead of considering only the nearest one in order to be more robust to outliers such as samples assigned to the wrong change. Finally, we assign the sample to the change that has the highest number of occurrences within the 50 samples.

III. EXPERIMENTAL RESULTS

A. Dataset Description

We tested the proposed method on one simulated dataset and two real multitemporal Hyperspectral datasets.

The simulated dataset is based on a real HS image available in the Real-World Hyperspectral Images Database [58]. The image (Figure 4a) represents a wall in an outdoor area and was acquired by a Nuance FX camera that has a spectral resolution of approximately 10 nm with 31 bands ranging from 420 to 720 nm. To simulate changes, 7 tiles (colored tiles in Figure 4a) were extracted from the original image and inserted in different areas of the same image (Figure 4b). The selected tiles represent different materials and illumination condition thus simulating different changes. Additive White Gaussian noise was added to the simulated images (SNR 15 dB). Figure 4c shows the reference map of the 7 simulated changes.

The two real multitemporal Hyperspectral datasets were acquired by the Hyperion sensor mounted on board the EO-1 satellite. The data used to generate these two datasets can be downloaded at [59]. The Hyperion data are characterized by 242 spectral bands ranging from 350 to 2580 nm, with a spectral resolution of 10 nm and a spatial resolution of 30 m. The second dataset represents an agricultural area in the Benton County, Washington, USA. The first image was acquired

on May 1st, 2004 while the second image was acquired on May 8th, 2007. Figures 5a and 5b show the true color composition of the two HS images characterized by circular agricultural fields (due to the pivot irrigation system). Almost all the changes in this area are related to variations of the crop, soil or water content (caused by different amount of irrigated water) in the fields. The false color compositions of the HS difference image in Figure 5c highlights how the study area is characterized by different kinds of change and that one single field can be affected by more than one type of change. Figure 5c gives only a partial representation of the change information present in the data since only 3 spectral channel are exploited. The blue dashed circle in Figure 5d shows that two changes can be distinguished at band 34 but not at band 103 (Figure 5e). Conversely, the violet circle highlights a field that has uniform behaviour at band 34 while it shows two uniform regions at band 103. This highlights how the different spectral channels can contain significantly different change information. For this dataset a reference map (Figure 6) was defined by photointerpretation from experts for a portion of the image (dataset box in Figure 5). We will refer to the portion of the Benton dataset with reference map as BentonRM.

The third dataset was acquired over an agricultural area in the Albacete province, Spain. The two images were acquired on May 22th 2012 (Figure 7a) and May 14th 2014 (Figure 7b). This area is characterized by agricultural fields with various shapes. The false color composition of the difference image (Figure 7c) shows that the area has been affected by different kinds of changes. Also for this dataset, most of the changes are related to variation of the crop, soil or water content and the presence of some clouds in the 2014 acquisition. The blue dashed circles and violet circles in Figures 7d and 7e confirm that for a given kind of change one or more bands may contain more relevant information with respect to other spectral channels.

Both the Hyperion datasets were first co-registered with a residual error of 0.5 pixels. Then, we applied a radiometric calibration followed by an atmospheric correction carried out using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm [60], [61]. We applied the FLAASH algorithm only to the calibrated non overlapping bands, i.e., channels 8-57 and 79-224 [62]. Bands related to water absorption, i.e., bands 121-126, 167-180 and 222-224 [63] were removed as well as noisy bands with wavelength close to the water vapor absorption ones thus obtaining the final set of channels 8-57, 82-119, 131-164, 182-184, and 187-220. In the second dataset we removed also bands 117, 141, 146, 188, 198 due to bad stripes (band 117) and noise (bands 141, 146, 188, 198). It is worth noting that the selected bands still contain also noisy bands thus preserving the complexity of the two datasets.

We used the simulated dataset and the BentonRM dataset to perform a quantitative evaluation and we compared the numerical accuracy with the one obtained by the S^2CVA [45], which is based on the manual identification of th different kinds of change. In order to have a fair comparison, we used the same binary CD map both for the proposed method and for the S^2CVA. For the BentonRM dataset, multiple CD results
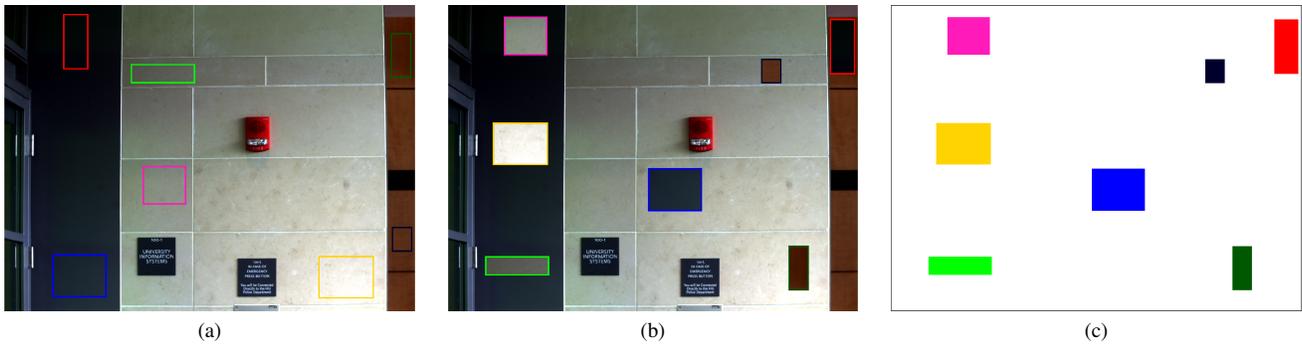
Fig. 4: Simulated dataset: (a-b) true color composites (R: 710 nm; G: 620 nm; B: 510 nm) of the original (a) and simulated image (b) . (c) Reference change map where each color represents one of the 7 changes and white represent the no changed class.
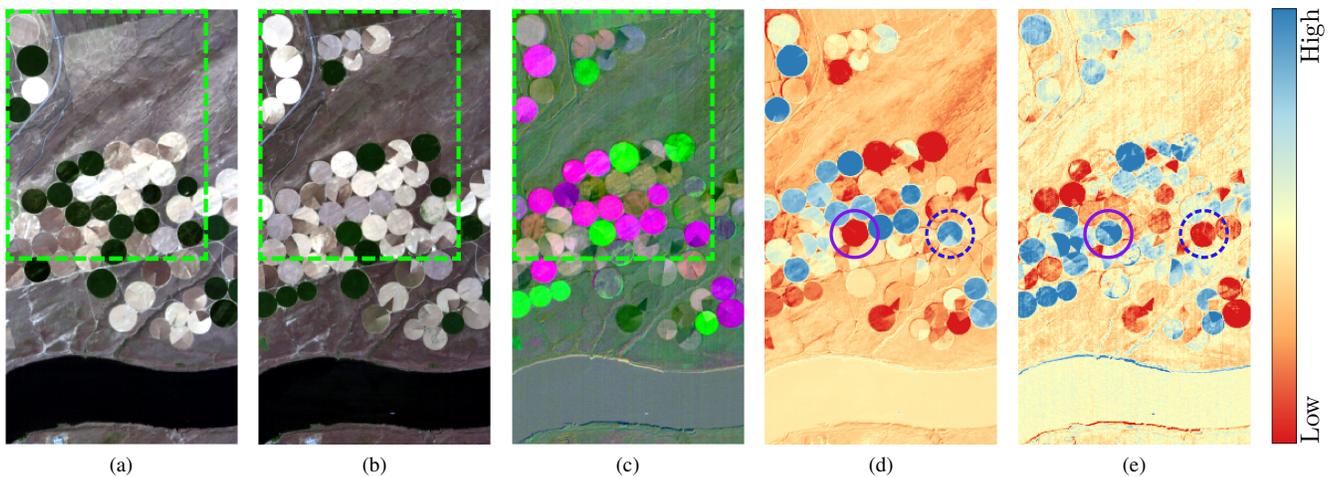

Fig. 5: Benton County agricultural area dataset: true color composites (R: 640.5 nm, G: 569.27 nm, B: 467.52 nm) of the HS data acquired at times (a) t_1 (2004) and (b) t_2 (2008), respectively. (c) False color composite (R: 1729.7 nm, G: 1023.4 nm, B: 447.17 nm) of the difference image (i.e., HCV image). Single bands (d) 34 (691.37 nm) and (e) 103 (1174.77 nm) of the HCV image, respectively, with the corresponding color bar. The green dashed lines in (a-c) identify the portion of the dataset for which a reference map is available (BentonRM).
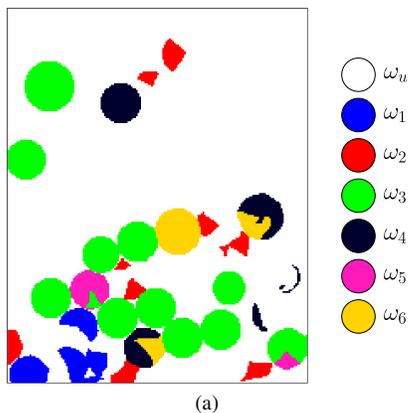

Fig. 6: Reference map of the 6 classes of change (BentonRM dataset delineated by the dashed line in Figure 5).

are available for the method proposed in [64]. This method is based on a supervised Band Selection (BS) followed by a supervised classification of the difference image using a Support Vector Machine (SVM). In the BS-SVM the binary CD was performed in the classification step and thus it was not possible to use the same binary CD map. However, note that the two binary maps are very similar. The datasets with reference maps were also used for a sensitivity analysis of the method parameters. To perform this quantitative analysis, for both datasets we set the sensitivity (i.e., threshold T_h) such that the number of identified change classes J is equal to the one of the reference map. Quantitative results are presented for the pixels correctly identified as changed in the binary CD map. This allows us to asses the effectiveness of the proposed method in distinguishing types of change without error propagation from the binary CD step (which in turn is common with the S^{2}CVA). Since no reference map is available for most of the second and the third datasets, we
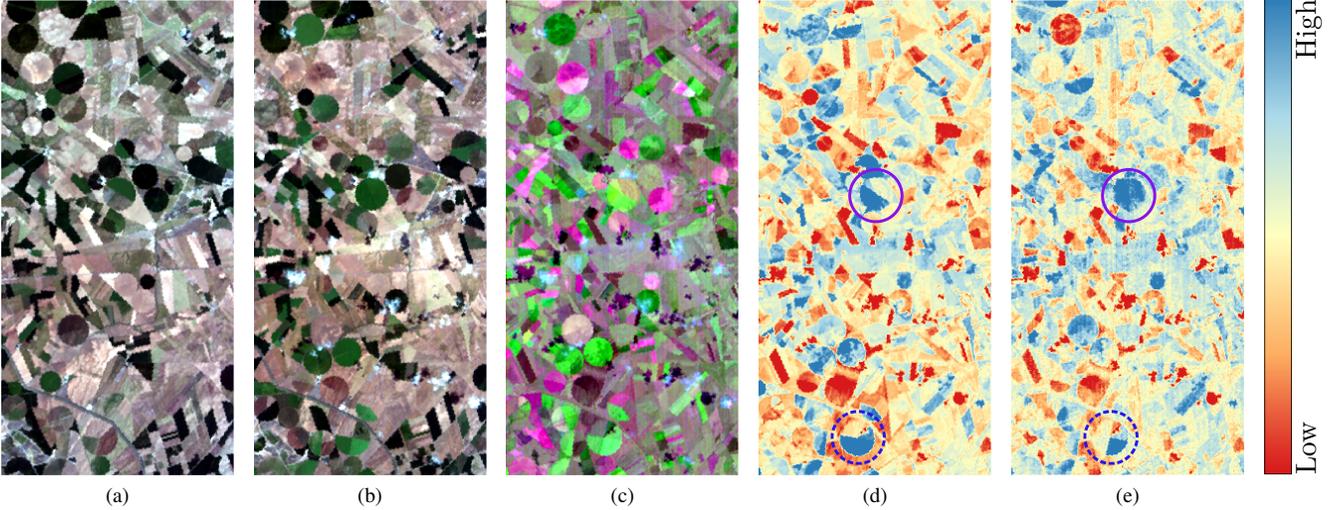
Fig. 7: Albacete Province agricultural area dataset: true color composites (R: 640.5 nm, G: 569.27 nm, B: 467.52 nm) of the HS data acquired at times (a) t_1 (2012) and (b) t_2 (2014), respectively. (c) False color composite (R: 1275.66 nm, G: 732.07 nm, B: 548.92 nm) of the difference image (i.e., HCV image). Single bands (d) 38 (732.07 nm) and (e) 100 (1144.48 nm) of the HCV image, respectively, with the corresponding color bar.

also validated the results by photointerpretation using different false color composites of the HCV images. For each of the two real datasets we analyzed the effectiveness of the method in detecting changes with different levels of sensitivity by setting the number of changes to $J = 4$ and $J = 15$. Moreover, we used the S^{2}CVA method to extract the changes and then we compared the resulting map with the one obtained by the proposed method setting J to the same number of changes as the S^{2}CVA map.

We set the threshold T_r on the Hamming distance between adjacent bits (of the BHCVs) to $0.1N$. The threshold T_p on the prior probability of each CBHCV was set to 0.001. Since our aim is to analyze the different CD maps obtained by varying the depth T_h at which the dendrogram is cut, we did not set a fixed threshold T_h, but we set a lower bound. In particular, we set the lower bound of T_h such that the maximum number of changes in which the samples can be separated is 15.

B. Experimental Results: Sensitivity Analysis

Sensitivity analysis was conducted on the simulated dataset and on the BentonRM dataset by running the proposed method with different values of T_r and T_p ($T_r \in [0.05N, 0.25N]$ and $T_p \in [0.001, 0.005]$). For each combination of the two parameters we computed the $\mathcal{K}_c$ (Cohen's Kappa) coefficient for the multiple CD result. Figure 8 shows a visual representation of the values of $\mathcal{K}_c$ coefficient for different settings for the simulated dataset (Figure 8a) and for the BentonRM dataset (Figure 8b). In the simulated dataset, the proposed method reached a $\mathcal{K}_c$ coefficient of 0.99 for a 44.4% of the combinations of the parameters, $\mathcal{K}_c = 0.81$ for another 44.4%. The remaining few cases are the ones corresponding to the combinations with $T_r = 0.25N$ which is at the very end of the considered interval. For the BentonRM dataset (Figure 8b), 44.4% of the combinations reached a $\mathcal{K}_c$ coefficient larger than
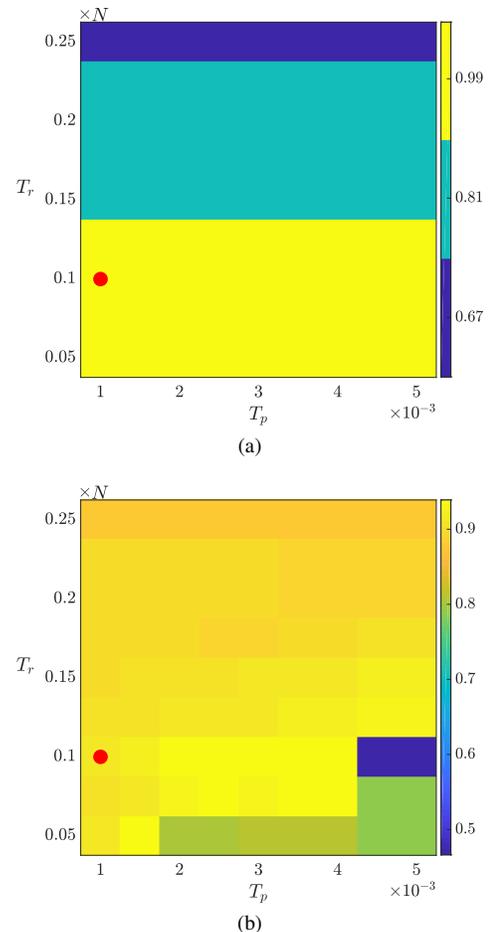

Fig. 8: Sensitivity analysis results showing the values of the $\mathcal{K}_c$ coefficient for different combinations of T_r and T_p: (a) simulated dataset and (b) BentonRM dataset. The red dot identifies the optimal settings.
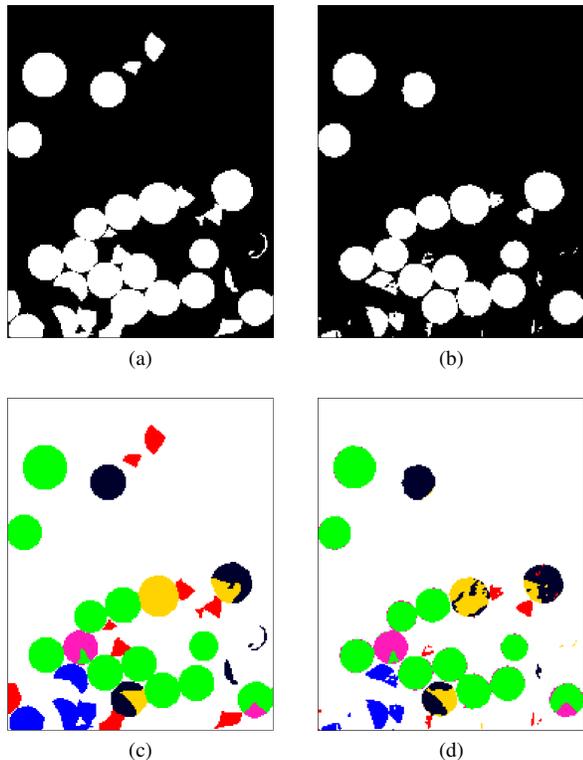
Fig. 9: CD results for the portion of the BentonRM dataset: (a) reference binary CD map, (b) proposed method binary CD map, (c) reference multiple CD map, (d) proposed method multiple CD map.

TABLE II: Quantitative results comparison among the proposed method, the S²CVA and the BS-SVM for the simulated and BentonRM datasets.

Dataset	OA[%]		$\mathcal{K}_c$	
	Simulated	BentonRM	Simulated	BentonRM
Proposed Method	99.88	94.55	0.99	0.91
S²CVA [45]	99.99	94.37	0.99	0.91
BS-SVM [64]	–	96.92	–	0.95

0.9, while 48.1% of the combinations reached a value between 0.8 and 0.9. By analyzing these results we can conclude that the algorithm is robust to the variation of the threshold T_r and T_p.

C. Experimental Results: Quantitative and Qualitative Comparison

For the simulated dataset (the reference map was used to eliminate unchanged samples), the quantization and codeword definition steps converted the $N = 113937$ changed HCVs into BHCVs composed by $K = 82$ bits. The compression step led to the CBHCVs obtaining codewords having a length of $I = 9$ bits. The CBHCVs analysis identified $U = 65$ unique CBHCVs with $U' = 28$ unique compressed codewords having a corresponding prior probability $P_u > T_p$. Therefore, the hierarchical clustering algorithm was applied only to

TABLE III: Confusion matrix for the multiple CD results of the portion of the BentonRM dataset. The dashed box identifies the samples correctly identified as changed by the binary CD.

		Ground Truth						
		ω_u	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
Proposed Method	ω_u	**30564**	502	829	455	178	29	55
	ω_1	0	**532**	0	0	0	0	0
	ω_2	1	0	**218**	0	11	0	1
	ω_3	5	0	0	**4509**	0	5	0
	ω_4	1	0	1	0	**1029**	0	221
	ω_5	4	0	0	147	0	**445**	0
	ω_6	4	0	0	0	43	0	**711**

28 codewords. The $U' = 28$ unique CBHCVs represent the 99.3% of the total number of changed samples, thus showing that the coding step effectively represents the change information. The analysis of the CD map showed that the method can discriminate between all the kinds of change in an unsupervised and automatic way with an Overall Accuracy (OA) of 99.98% and a $\mathcal{K}_c = 0.99$. The S²CVA provided similar performances but relying on a supervised and non-automatic analysis. Table II shows the numerical results for the simulated dataset. The very high accuracy is motivated by the fact that the simulation does not fully captures all the challenges that have to be addressed when dealing with real HS data.

In the BentonRM dataset, the binary CD achieved an OA of 94.91% and a $\mathcal{K}_c$ coefficient of 0.85 with a commission error of 0.05% and an omission error of 5.06%. The method converted the $N = 7888$ HCVs into BHCVs composed by $K = 393$ bits, which were then compressed into CBHCVs having a length $I = 25$ bits. $U = 574$ unique CBHCVs were identified and the hierarchical clustering was applied to $U' = 107$ CBHCVs. The $U' = 107$ CBHCVs represent the 87.2% of the changed samples. This confirms that the coding step effectively represents in an efficient way the change information. Indeed, we are representing most of the changed samples with few binary codewords. The remaining samples (i.e., 12.8%) are not used in the construction of the dendrogram, since they are likely to be noisy samples. They are nevertheless assigned to one of the kinds of change after the change classes are identified.

Figure 9 shows the results in terms of binary and multiple CD obtained by the proposed method compared with the reference map. Table III shows the confusion matrix of the multiple CD for the proposed approach. Most of the samples correctly identified as changed have been assigned to the correct change class. The multiple CD achieved an OA of 94.55% and a $\mathcal{K}_c$ coefficient of 0.91. Table II shows a quantitative comparison between the proposed method, the S²CVA [45] and the BS-SVM [64]. The proposed method achieved a slightly higher accuracy compared to the S²CVA despite being completely automatic. The BS-SVM reached a modestly higher accuracy (+2.4%). However, note that this
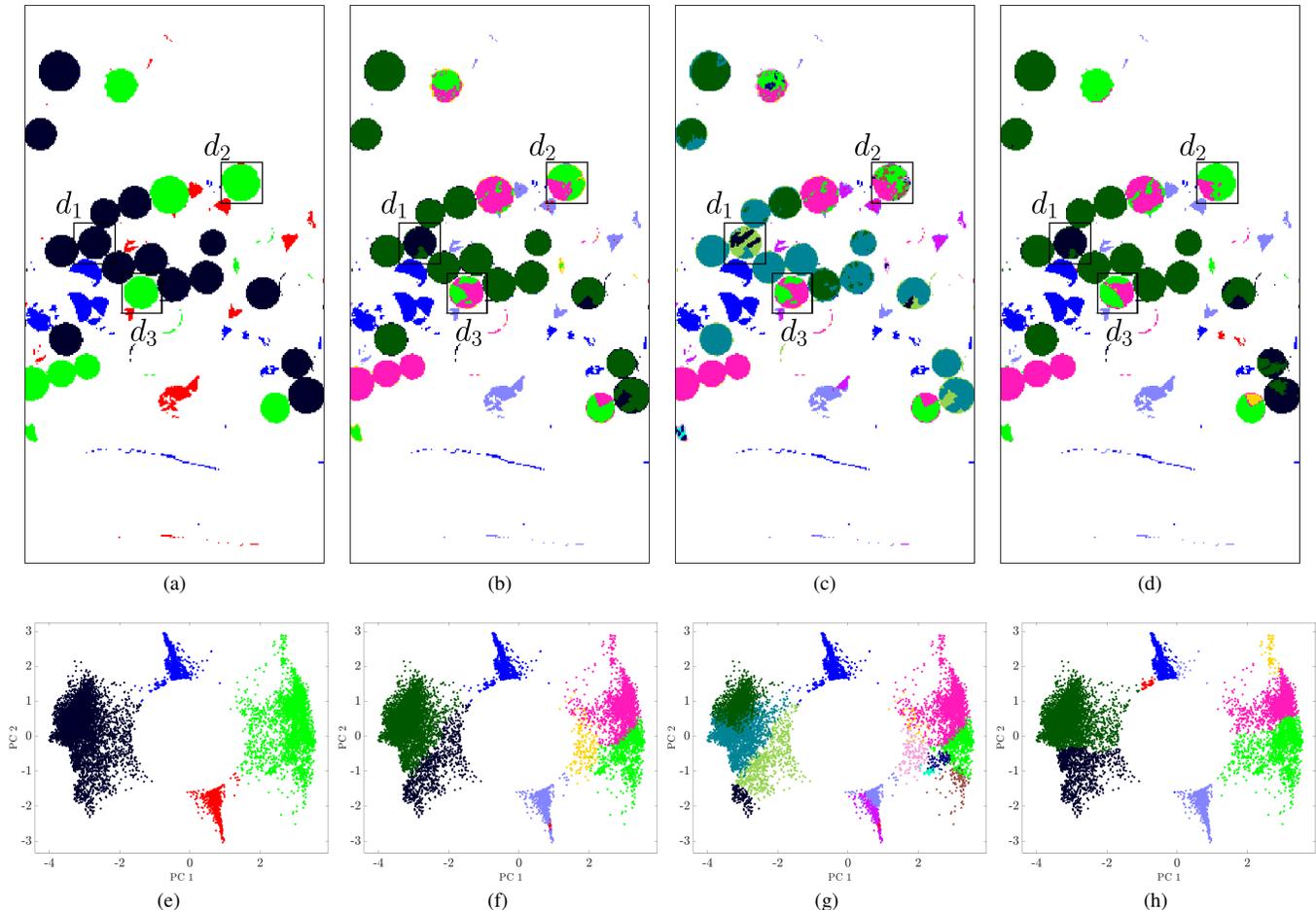
Fig. 10: Change Detection results for the Benton County dataset: CD maps obtained by cutting the dendrogram at a depth to obtain (a) $J = 4$, (b) $J = 8$, (c) $J = 15$. (d) S^{2}CVA map with 8 identified changes. 2-D scatter plots representing the first two components of the Principal Component Analysis of the HCVs (in the original representation) of the changed samples: the plots represent the (e) 4, (f) 8, (g) 15 (g), (h) 8 (S^{2}CVA) clusters. Each color represents a kind of change ω_j, whereas white identifies the no changed areas.

method is supervised and requires multitemporal reference samples to train the SVM. Thus the proposed unsupervised method is competitive with state of the art supervised methods.

Let us consider the entire Benton County dataset. The visual analysis of the binary Change Detection map shows that changes are identified with high accuracy with only few false alarms due to outliers in the HCV and residual registration errors. Also some missed alarms are present but they correspond to changes with low magnitude. The quantization and codeword definition steps converted the $N = 11965$ changed HCVs into BHCVs composed by $K = 375$ bits and the compression generated CBHCVs with $I = 27$ bits. $U = 681$ unique CBHCVs were identified with only $U' = 106$ compressed codewords with a prior probability $P_u > T_p$. The selected 106 unique CBHCVs represent 10544 of the 11965 samples (i.e., 88%).

Figure 10 shows the CD results for the second dataset. Figures 10a-10c show the CD maps obtained by cutting the dendrogram at different depths. The figures show that by

decreasing (moving from Figure 10a to 10c) the depth value T_h, it is possible to discriminate among an increasing number of changes. The scatter plots (Figures 10e-10g) point out that the discrimination into different kinds of change is made consistently for almost all the changes since most of the clusters show little overlapping between each other. Figure 10d shows the CD map obtained by the manual identification of the S^{2}CVA, which identified 8 changes. The change maps obtained by the proposed automatic method (Figure 10b) and the manual S^{2}CVA (Figure 10d) are very similar, with large portion of the data showing the same kinds of change in both images.

Figure 11 shows three details (d_1, d_2, d_3) of the CD maps of Figure 10 together with false color composition of the HS difference image. The images show that when the dendrogram is cut to obtain 4 clusters we identify one kind of change in each field. The false color compositions of the difference image in Figure 11 highlight that more than one kind of change is present in the selected fields. The Figures related to details d_1 and d_3 show that by increasing the sensitivity
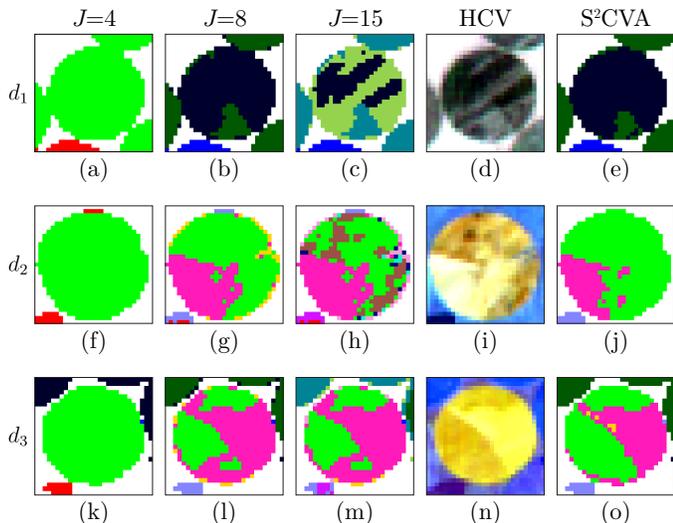
Fig. 11: Details of the CD maps of the Benton dataset (Figures 10a-10c): (a-c) detail d_1. (f-h) Detail d_2. (k-m) Detail d_3. False color composites of the HCV image representing detail (d) d_1 (R: 823.65 nm, G: 972.99 nm, B: 1053.69 nm), (i) d_2 (R: 844 nm, G: 1225.17 nm, B: 1689.3 nm), (n) d_3 (R: 1053.69 nm, G: 1336.15 nm, B: 548.92 nm), respectively. Details of the S²CVA map (Figure 10d): (e) detail d_1, (j) detail d_2, (o) detail d_3.

of the method we increase the sensitivity to the different kinds of change and we can separate subtle changes in uniform areas. In contrast, the Figures regarding detail d_2 show that when the sensitivity of the method goes above a given threshold, the noise in the CD map becomes significant ($J = 15$). However, note that the false color composition of the difference image of d_2 shows that in the considered field there is a high probability of having more than 2 changes. The S²CVA CD map show similar results to the proposed method for details d_2 and d_3. However, note that for detail d_1 the S²CVA identifies only two different changes, whereas the proposed method identifies all the changes of the considered field (as confirmed also by the false color composition). This analysis pointed out that while increasing the sensitivity allows us to discriminates between more kinds of change, this comes at the cost of an increase in the vulnerability of the method to noise and outliers. In other words, after a given value of T_h, as expected the method starts to identify the noisy samples as different changes.

In the Albacete dataset, a visual analysis of the binary CD map shows a slightly higher percentage of missed and false alarms with respect to the second dataset. This is due to the overlapping of changed and unchanged statistical distributions in the magnitude of the difference image. The $N = 33693$ samples were converted into BHCVs composed by $K = 372$ bits, which were then compressed into $I = 28$ bits to obtain the CBHCVs. The CBHCVs analysis selected $U = 2382$ unique CBHCVs and $U' = 150$ codewords with a prior probability grater than T_p. The U' represented 25150 of the 33693 samples (i.e., 74.6%). This confirms the results of the

Benton dataset regarding the efficient representation of the change information.

Figure 12 shows the CD maps and the corresponding scatter plots for the Albacete dataset. The maps (Figures 12a-12c) show that the proposed method discriminates consistently among the different kinds of change since most of the changed areas show homogeneous spatial behaviour. This is also confirmed by the scatter plots (12e-12h) with the identified cluster showing little overlapping between each other. Figure 12d shows the CD map obtained with the manual iterative S²CVA which identified 11 changes. By comparing it with the map of the proposed method with $J = 11$ (Figure 12b), it is possible to see that there is a large number of fields that show the same change pattern in the two maps. However, there are also some fields that show different change patterns. In some cases, this is due to the fact that the selected sensitivity does not allow to identify some changes. This is the case of detail d_3 in Figure 13. The false color composition of the difference image points out that the lower portion of the field shows two different changes. These changes are correctly identified by the S²CVA. Note that by increasing the sensitivity, the proposed method correctly identifies the considered changes (Figure 13m). A similar analysis can be done for the detail d_1. Another cause of these differences is related to the complexity of this multitemporal dataset. Compared to the Benton dataset, the Albacete dataset is a much more complex dataset since the changes are much less separable. This can lead to errors in the CD maps of the proposed method but also in that of the S²CVA since the manual selection of the changes can not be performed accurately due to the difficulties in identifying the boundaries of the different changes. As an example, let us consider detail d_2 shown in Figure 13. The false color composition of the HCV image shows a complex spatial pattern. The CD maps of the proposed method separates correctly the different changes when the sensitivity of the method is increased. In contrast, the S²CVA map shows only one kind of change.

IV. CONCLUSION

In this work we proposed a method for change detection in hyperspectral images. HS data are characterized by a rich information content due to the dense sampling of the spectrum. This comes at a cost of having to deal with much more complex data, due to the high dimensionality and redundancy. These factors make the extraction of the change information more complex with respect to MS data, thus increasing the difficulty in discriminating between different kinds of change. In this work we developed a technique that aims at extracting from the individual spectral channels the relevant information for the CD problem. In greater detail, after a binary CD, we focus on the changed samples to move from the real valued representation of the HCVs to a codeword based representation. After the coding of the HCVs, we select only the unique CBHCVs with a prior probability higher than a given threshold thus reducing the number of elements to be processed in the following steps. Finally, we apply the
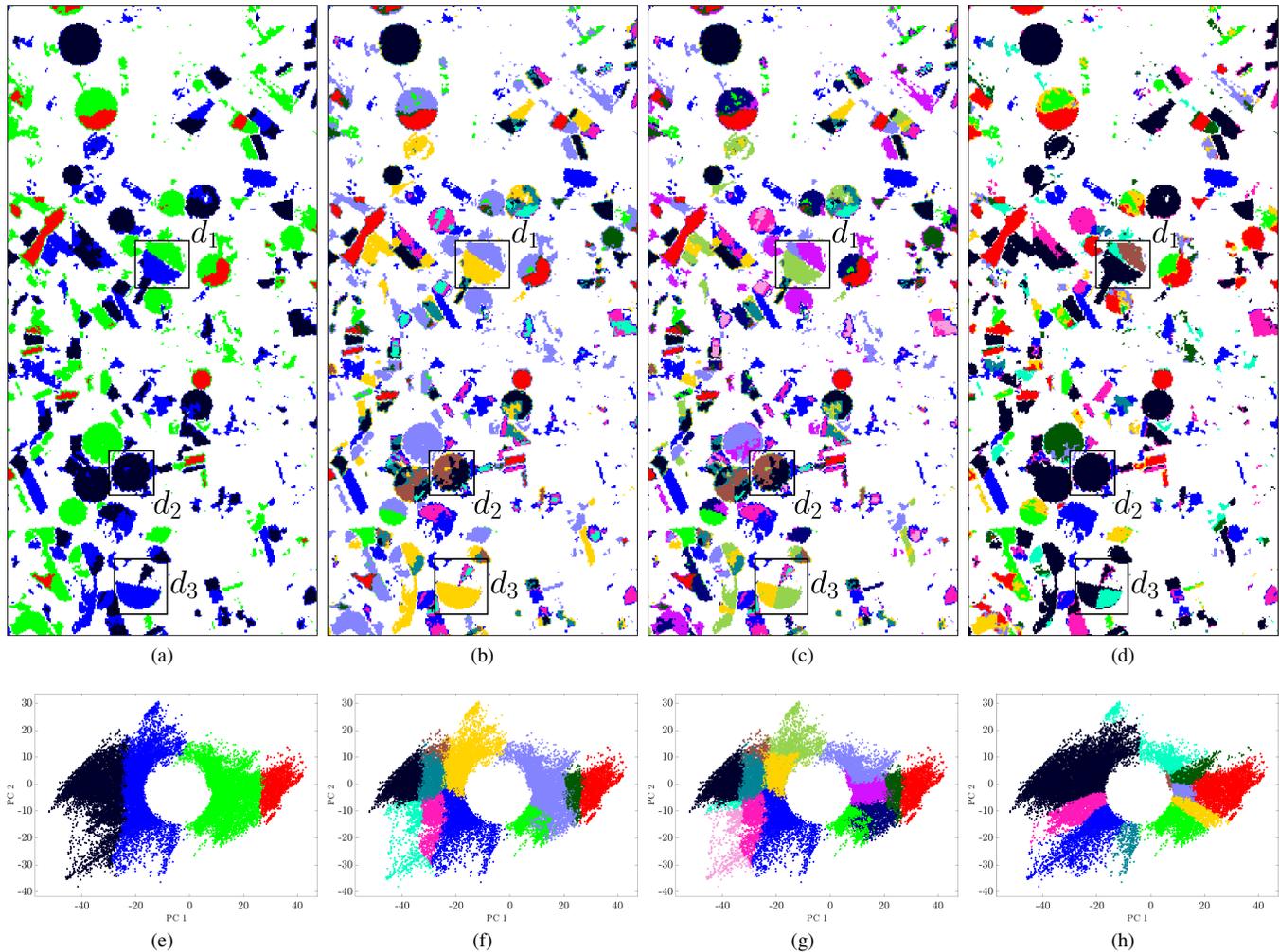
Fig. 12: Change Detection results for the Albacete province dataset: CD maps obtained by cutting the dendrogram at a depth to obtain (a) $J = 4$, (b) $J = 11$, (c) $J = 15$. (d) S²CVA map with 11 identified changes. 2-D scatter plots representing the first two components of the Principal Component Analysis of the HCVs (in the original representation) of the changed samples: the plots represent the (e) 4, (f) 11, (g) 15 (g), (h) 11 (S²CVA) clusters. Each color represents a kind of change ω_j, whereas white identifies the no changed areas.

agglomerative hierarchical clustering to the selected unique CBHCVs thus obtaining a dendrogram. The dendrogram can be cut to discriminate between the different kinds of change. The sensitivity analysis on the selection of the values of T_r and T_p showed that the method reaches similar accuracies for a wide range of combinations thus proving that is robust to the variation of these parameters. The HCVs coding proved to be an effective way of simplifying the CD problem by highlighting the relevant change information. This was confirmed by the fact that, for all the datasets, a small set of unique CBHCVs represented most of the changed samples. This allowed us to construct the dendrogram using a very small number of samples while preserving almost all the relevant change information. The numerical results obtained on the simulated dataset ($\mathcal{K}_c = 0.99$) and on the BentonRM dataset ($\mathcal{K}_c = 0.91$) showed that the proposed binary codewords representation can be used to discriminate between different changes. Compared to the S²CVA, the proposed method ob-

tained a slightly lower value of OA for the simulated dataset, whereas for the BentonRM it slightly increased the OA and $\mathcal{K}_c$. The BS-SVM yielded a higher accuracy compared to the proposed method. However, note that unlike the proposed approach, the S²CVA is not an automatic method and requires the manual identification of the changes while the BS-SVM requires the availability of multitemporal training samples which are seldom available. The numerical results proved that the proposed unsupervised and automatic approach is competitive with supervised methods and methods based on the manual identification of the changes. The qualitative analysis of the CD maps of the Benton and Albacete datasets confirmed that the proposed method identifies the different changes with a high accuracy. Moreover, the tree structure representation allows us to produce multiple CD maps with different levels of details. The possibility of changing the sensitivity to the different kinds of change showed that the proposed approach can produce different CD maps depending
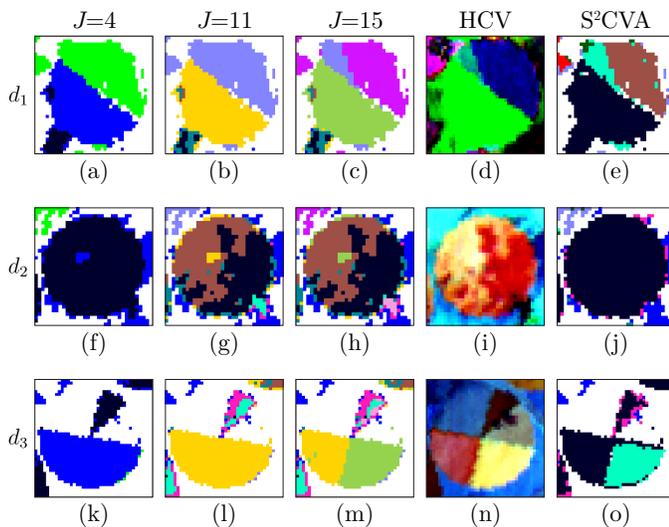
Fig. 13: Details of the CD maps of the Albacete dataset (Figures 12a-12c): (a-c) detail d_1. (f-h) Detail d_2. (k-m) Detail d_3. False color composites of the HCV image representing detail (d) d_1 (R: 660.85 nm, G: 833.83 nm, B: 2183.63 nm), (i) d_2 (R: 833.83 nm, G: 538.74 nm, B: 1669.1 nm), (n) d_3 (R: 833.83 nm, G: 1235.27 nm, B: 1669.1 nm), respectively. Details of the S²CVA map (Figure 12d): (e) detail d_1, (j) detail d_2, (o) detail d_3.

on the selected sensitivity. It is worth noting that a trade-off between sensitivity to changes and sensitivity to noise, outliers and errors generated in the coding step should be obtained. When the depth of the cut T_h is too small, outliers may be identified as a separate kind of change.

As future developments we plan to: 1) integrate the Binary CD in the coding process in order to reduce missed and false alarms; 2) use advanced information theory techniques to optimize the compression of the binary codewords to reduce the redundant information; and 3) test the proposed method on other multitemporal HS data and also on more real datasets for which a reliable multitemporal reference map is available (still difficult to obtain).

ACKNOWLEDGMENT

REFERENCES

[1] L. Bruzzone and F. Bovolo, "A novel framework for the design of change-detection systems for very-high-resolution remote sensing images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 609–630, March 2013.

[2] L. Bruzzone and D. F. Prieto, "Automatic analysis of the difference image for unsupervised change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1171–1182, May 2000.

[3] T. Celik, "Change detection in satellite images using a genetic algorithm approach," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 2, pp. 386–390, April 2010.

[4] A. Ghosh, N. S. Mishra, and S. Ghosh, "Fuzzy clustering algorithms for unsupervised change detection in remote sensing images," *Information Sciences*, vol. 181, no. 4, pp. 699 – 715, 2011.

[5] P. Du, S. Liu, J. Xia, and Y. Zhao, "Information fusion techniques for change detection from multi-temporal remote sensing images," *Information Fusion*, vol. 14, no. 1, pp. 19 – 27, 2013.

[6] P. Du, S. Liu, P. Gamba, K. Tan, and J. Xia, "Fusion of difference images for change detection over urban areas," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 4, pp. 1076–1086, Aug 2012.

[7] S. Ghosh, L. Bruzzone, S. Patra, F. Bovolo, and A. Ghosh, "A context-sensitive technique for unsupervised change detection based on hopfield-type neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 778–789, March 2007.

[8] S. Ghosh, S. Patra, and A. Ghosh, "An unsupervised context-sensitive change detection technique based on modified self-organizing feature map neural network," *International Journal of Approximate Reasoning*, vol. 50, no. 1, pp. 37 – 50, 2009, special Section on Recent advances in soft computing in image processing and Special Section on Selected papers from NAFIPS 2006.

[9] L. Bruzzone and S. Serpico, "Detection of changes in remotely-sensed images by the selective use of multi-spectral information," *International Journal of Remote Sensing*, vol. 18, no. 18, pp. 3883–3888, 1997.

[10] A. Ghosh, B. N. Subudhi, and L. Bruzzone, "Integration of gibbs markov random field and hopfield-type neural networks for unsupervised change detection in remotely sensed multitemporal images," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3087–3096, Aug 2013.

[11] L. Bruzzone and D. F. Prieto, "A minimum-cost thresholding technique for unsupervised change detection," *International Journal of Remote Sensing*, vol. 21, no. 18, pp. 3539–3544, 2000.

[12] H. Liu and Q. Zhou, "Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison," *International Journal of Remote Sensing*, vol. 25, no. 5, pp. 1037–1050, 2004.

[13] F. Pacifici, F. D. Frate, C. Solimini, and W. J. Emery, "An innovative neural-net method to detect temporal changes in high-resolution optical satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 9, pp. 2940–2952, Sept 2007.

[14] H. Alphan, H. Doygun, and Y. I. Unlukaplan, "Post-classification comparison of land cover using multitemporal landsat and aster imagery: the case of kahramanmaraş, turkey," *Environmental Monitoring and Assessment*, vol. 151, no. 1, pp. 327–336, Apr 2009.

[15] B. Jeon and D. A. Landgrebe, "Classification with spatio-temporal interpixel class dependency contexts," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 4, pp. 663–672, Jul 1992.

[16] M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone, "Supervised change detection in vhr images using contextual information and support vector machines," *International Journal of Applied Earth Observation and Geoinformation*, vol. 20, pp. 77 – 85, 2013, earth Observation and Geoinformation for Environmental Monitoring.

[17] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 858–867, Jul 1997.

[18] L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1350–1359, May 1999.

[19] B. Demir, F. Bovolo, and L. Bruzzone, "Detection of land-cover transitions in multitemporal remote sensing images with active-learning-based compound classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1930–1941, May 2012.

[20] J. Zhong and R. Wang, "Multi-temporal remote sensing change detection based on independent component analysis," *International Journal of Remote Sensing*, vol. 27, no. 10, pp. 2055–2061, 2006.

[21] 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 Sensing of Environment*, vol. 64, no. 1, pp. 1 – 19, 1998.

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

[23] F. Bovolo, S. Marchesi, and L. Bruzzone, "A framework for automatic and unsupervised detection of multiple changes in multitemporal images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 6, pp. 2196–2212, June 2012.

[24] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Ann. Statist.*, vol. 38, no. 5, pp. 2916–2957, 10 2010. [Online]. Available: http://dx.doi.org/10.1214/10-AOS799

[25] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *International Journal of Remote Sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[26] F. Bovolo and L. Bruzzone, "The time variable in data fusion: A change detection perspective," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 8–26, Sept 2015.

[27] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Hierarchical unsupervised change detection in multitemporal hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 244–260, Jan 2015.

[28] L. Candela, R. Formaro, R. Guarini, R. Loizzo, F. Longo, and G. Varacalli, "The prisma mission," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2016, pp. 253–256.

[29] T. Matsunaga, A. Iwasaki, S. Tsuchida, K. Iwao, J. Tanii, O. Kashimura, R. Nakamura, H. Yamamoto, S. Kato, K. Obata, K. Mouri, and T. Tachikawa, "Current status of hyperspectral imager suite (hisui) onboard international space station (iss)," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2017, pp. 443–446.

[30] L. e. a. Guanter, "The enmap spaceborne imaging spectroscopy mission for earth observation," *Remote Sensing*, vol. 7, no. 7, pp. 8830–8857, 2015. [Online]. Available: http://www.mdpi.com/2072-4292/7/7/8830

[31] C. M. Lee, M. L. Cable, S. J. Hook, R. O. Green, S. L. Ustin, D. J. Mandl, and E. M. Middleton, "An introduction to the nasa hyperspectral infrared imager (hyspiri) mission and preparatory activities," *Remote Sensing of Environment*, vol. 167, pp. 6 – 19, 2015, special Issue on the Hyperspectral Infrared Imager (HyspIRI).

[32] T. Feingersh and E. B. Dor, "Shalom – a commercial hyperspectral space mission," in *Optical Payloads for Space Missions*. John Wiley & Sons, Ltd, 2015, ch. 11, pp. 247–263.

[33] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" in *International conference on database theory*. Springer, 1999, pp. 217–235.

[34] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, January 1968.

[35] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "Pca-based land-use change detection and analysis using multitemporal and multisensor satellite data," *International Journal of Remote Sensing*, vol. 29, no. 16, pp. 4823–4838, 2008. [Online]. Available: http://dx.doi.org/10.1080/01431160801950162

[36] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment*, vol. 44, no. 2, pp. 145 – 163, 1993, airbone Imaging Spectrometry. [Online]. Available: http://www.sciencedirect.com/science/article/pii/003442579390013N

[37] C.-I. Chang, "Spectral information divergence for hyperspectral image analysis," in *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. IEEE 1999 International*, vol. 1, 1999, pp. 509–511 vol.1.

[38] Y. Du, C.-I. Chang, H. Ren, C.-C. Chang, J. O. Jensen, and F. M. D'Amico, "New hyperspectral discrimination measure for spectral characterization," *Optical Engineering*, vol. 43, no. 8, pp. 1777–1786, 2004.

[39] A. P. Schaum and A. Stocker, "Hyperspectral change detection and supervised matched filtering based on covariance equalization," pp. 77–90, 2004.

[40] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel rx algorithm for anomaly and change detection using hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6497–6504, Nov 2016.

[41] N. Acito, M. Diani, G. Corsini, and S. Resta, "Introductory view of anomalous change detection in hyperspectral images within a theoretical gaussian framework," *IEEE Aerospace and Electronic Systems Magazine*, vol. 32, no. 7, pp. 2–27, July 2017.

[42] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Unsupervised hierarchical spectral analysis for change detection in hyperspectral images," in *2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, June 2012, pp. 1–4.

[43] V. Ortiz-Rivera, M. Vélez-Reyes, and B. Roysam, "Change detection in hyperspectral imagery using temporal principal components," pp. 623 312–623 312–10, 2006. [Online]. Available: http://dx.doi.org/10.1117/12.667961

[44] A. A. Nielsen, "The regularized iteratively reweighted mad method for change detection in multi- and hyperspectral data," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 463–478, Feb 2007.

[45] S. Liu, L. Bruzzone, F. Bovolo, M. Zanetti, and P. Du, "Sequential spectral change vector analysis for iteratively discovering and detecting multiple changes in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4363–4378, Aug 2015.

[46] A. Ertürk, M. D. Iordache, and A. Plaza, "Sparse unmixing-based change detection for multitemporal hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 2, pp. 708–719, Feb 2016.

[47] S. Liu, L. Bruzzone, F. Bovolo, and P. Du, "Unsupervised multitemporal spectral unmixing for detecting multiple changes in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2733–2748, May 2016.

[48] A. Ertürk, M. D. Iordache, and A. Plaza, "Sparse unmixing with dictionary pruning for hyperspectral change detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 321–330, Jan 2017.

[49] D. Marinelli, F. Bovolo, and L. Bruzzone, "A novel method for unsupervised multiple change detection in hyperspectral images based on binary spectral change vectors," in *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, June 2017, pp. 1–4.

[50] ——, "A novel change detection method for multitemporal hyperspectral images based on a discrete representation of the change information," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2017, pp. 161–164.

[51] F. Bovolo and L. Bruzzone, "An adaptive thresholding approach to multiple-change detection in multispectral images," in *2011 IEEE International Geoscience and Remote Sensing Symposium*, July 2011, pp. 233–236.

[52] M. Zanetti, F. Bovolo, and L. Bruzzone, "Rayleigh-rice mixture parameter estimation via em algorithm for change detection in multispectral images," *IEEE Transactions on Image Processing*, vol. 24, pp. 5004–5016, 2015.

[53] Z. Bar-Joseph, D. K. Gifford, and T. S. Jaakkola, "Fast optimal leaf ordering for hierarchical clustering," *Bioinformatics*, vol. 17, no. suppl 1, pp. S22–S29, 2001.

[54] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2631–2641, Nov 1999.

[55] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, Dec 2007.

[56] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1279–1289, June 2016.

[57] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 5910–5922, Oct 2018.

[58] A. Chakrabarti and T. Zickler, "Statistics of Real-World Hyperspectral Images," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 193–200.

[59] Earth science data archives of U.S. Geological Survey (USGS). [Online]. Available: https://earthexplorer.usgs.gov

[60] S. M. Adler-Golden, M. W. Matthew, L. S. Bernstein, R. Y. Levine, A. Berk, S. C. Richtsmeier, P. K. Acharya, G. P. Anderson, J. W. Felde, J. Gardner *et al.*, "Atmospheric correction for shortwave spectral imagery based on modtran4," pp. 3753 – 3753 – 9, 1999.

[61] G. W. Felde, G. P. Anderson, T. W. Cooley, M. W. Matthew, S. M. adler Golden, A. Berk, and J. Lee, "Analysis of hyperion data with the flaash atmospheric correction algorithm," in *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*, vol. 1, July 2003, pp. 90–92 vol.1.

[62] R. Beck. (2003) Eo-1 user guide v. 2.3. [Online]. Available: http://eo1.usgs.gov/

[63] B. Datt, T. R. McVicar, T. G. V. Niel, D. L. B. Jupp, and J. S. Pearlman, "Preprocessing eo-1 hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1246–1259, June 2003.

[64] S. Liu, Q. Du, X. Tong, A. Samat, H. Pan, and X. Ma, "Band selection-based dimensionality reduction for change detection in multi-temporal hyperspectral images," *Remote Sensing*, vol. 9, no. 10, 2017.