Unsupervised Change Detection with Support Vector Domain Description

F. Bovolo, G. Camps-Valls

Abstract—In this paper the change detection problem is faced as a one-class classification problem where the minority class of changed pixels is considered as the target class. Among one-class classifiers present in the literature the support vector domain description (SVDD) one-class classifier is adopted thanks to its intrinsic regularization and robustness properties. The SVDD maps the data into a high dimensional feature space where the spherical support of the high dimensional distribution of changed pixels is computed. This results in the definition of a minimum enclosing ball (MEB) that encloses the highest number of target pixels. Boundary definition is performed involving both target (i.e., changed pixels) and outlier (i.e., unchanged pixels) samples. In order to identify nearly certain examples for the classes of both targets and outliers for training the SVDD an unsupervised procedure is adopted which is based on the thresholding of the magnitude of spectral change vectors. Experimental results obtained on two different multitemporal and multispectral remote sensing images pointed out the effectiveness of the proposed method.

Index Terms—Unsupervised Change Detection, Change Vector Analysis, Support Vector Domain Description, Kernel Methods, Bayesian Thresholding, Remote Sensing.

I. INTRODUCTION

BINARY change detection maps, where changed pixels are separated from unchanged pixels, are one of the products that can be produced when dealing with time series of remote sensing images. They can be generated according to either supervised or unsupervised approaches. The former ones require ground truth information for the set up of the system parameters, whereas the latter ones do not. Although the former result in higher change-detection accuracies, the latter are more appealing as ground truth information is not available in many real applications. Among unsupervised change-detection methods present in the literature [1]-[7], the change vector analysis (CVA) [1]-[3] technique represents one of the simplest (yet effective) and most widely used techniques for binary change detection. CVA involves multispectral images acquired by passive sensors, by using all the spectral channels that contain useful information with respect to the considered kind of change. CVA is based on three steps. The first step computes the vector difference of spectral feature vectors associated with couples of corresponding pixels in two images acquired on the same geographical area at two different times, and results in a multispectral difference image. Each pixel in this image is associated with a multidimensional vector named spectral change vector (SCV). In the second step the magnitude of each SCV is computed (sometimes also their direction is considered) [2]. This operation results in a 1-dimensional image usually called difference image. Finally, thresholding is applied to the difference image to obtain the desired change-detection map. Due to the statistical behaviors of multispectral images, and to the properties of the difference and magnitude operators, it is possible to assert that pixels with magnitude higher than a given threshold value are changed, while pixels with a magnitude lower than the threshold value are unchanged [1]-[3]. A major drawback of using the magnitude of SCVs is that the magnitude operator is not biunique and results in a decrease of information with respect to the SCVs feature space. Nonetheless, if no ground truth is available, the magnitude operator allows one to establish a relatively simple criterion (based on thresholding) for identifying nearly certain pixels belonging to either the class of changed or unchanged pixels [2],[4].

In this paper, in order to take advantage of the large amount of information present in the multispectral difference image, we formulate the change-detection problem in the higher dimensional SCVs feature space. Unsupervised analysis of SCVs requires the application of clustering algorithms in the context of an ill-posed problem. In order to address this problem, we reformulate the unsupervised change-detection problem in the multispectral difference image as a data domain description problem, also known as one-class classification. Among, the different methods for data domain description (or outlier detection) present in the literature [8],[9], the Support Vector Data Description (SVDD) method [10],[11] is adopted. This method aims at mapping the data into a high dimensional feature space where a hypersphere enclosing most of the patterns belonging to the class of interest (target class) and penalizing the other patterns (outlier class) can be defined. As all kernel methods, SVDD shows some interesting advantages over other techniques, like intrinsic regularization and robustness [15]-[18]. SVDD was recently introduced in the
remote sensing literature [12]-[14] and demonstrated to be effective in solving classification and change-detection problems when ground truth information is available. In the present paper, the SVDD is included in a system for unsupervised change detection, that aims at separating pixels belonging to the class of change (target class) from all unchanged pixels (outlier class) without any ground truth information. In order to properly constrain the learning process in absence of ground truth information, an unsupervised procedure for identifying examples is adopted, which is based on a selective thresholding of the magnitude of SCVs [4]. Thanks to the specific nature of the change-detection problem, this approach leads to the identification of both positive and negative examples. The outlier seeds are included in the training of the SVDD leading to a more effective description of the change-detection problem. The resulting one-class classifier (OCC) shows a higher capability in describing the target data [10],[11],[19].

The paper is organized into four sections. In the next section the architecture of the proposed change-detection approach is presented and each of its components is described in detail. In section 3 experimental results obtained by applying the proposed technique to two different remote sensing data sets are presented. Finally, Section 4 draws the conclusion of this work.

II. PROPOSED METHODOLOGY

Let \( I_1 \) and \( I_2 \) be two co-registered multispectral images of size \( P \times Q \) acquired over the same geographical area at different times \( t_1 \) and \( t_2 \). Let \( N \) be the number of spectral channels of each considered image and \( \Omega = \{ \alpha_n, \alpha_k \} \) the set of classes of unchanged and changed pixels to be identified. The proposed technique is based on a two-step procedure: i) an initialization step that exploits a Bayesian thresholding of the magnitude of SCVs; and ii) a Support Vector Data Description (SVDD) method that analyzes the multispectral difference image \( I_3 = I_2 - I_1 \) (see Fig. 1). This choice permits to consider the high dimensional information present in the multispectral difference image, and not only the one present in the magnitude of SVCs. In the following, we analyze these steps in more detail.

A. Bayesian Initialization

The first step of the proposed unsupervised approach to change detection aims at identifying the sets \( S_T \) and \( S_O \) of target (unchanged pixels) and outlier (changed pixels) patterns to be used as seeds for initializing the support vector data description (SVDD) OCC. Following [4], these subsets should contain pixels that are associated with changed or unchanged areas. However, as in our problem no ground truth information is available, the ideal assumption is relaxed and replaced with the more realistic constraint that seeds included in the sets \( S_T \) and \( S_O \) are associated with a high probability to belong to changed or unchanged areas.

According to the procedure described in [4], pixels with a high probability to belong to the change and no-change classes are identified by applying the CVA technique to \( I_1 \) and \( I_2 \), and by selectively thresholding the statistical distribution \( p(i^\rho) \) of the magnitude of SCVs in \( I_p (i^\rho \) is the random variable associated with the magnitude of the spectral change vectors in \( I_p )\). In the literature, several threshold-selection methods (e.g., see [3]) have been proposed that can be used for identifying the threshold value \( T \), which separates changed from unchanged pixels. Among them, we recall threshold-selection approaches based on the Bayesian decision theory, which showed to be effective in many change-detection scenarios. The application of the Bayesian theory to threshold selection requires the estimation of the class-statistical parameters, i.e., class prior probabilities and class-conditional probabilities. As we are dealing with an unsupervised change-detection problem, these statistical quantities are estimated from the data (without any prior information) according to the Expectation-Maximization (EM) algorithm [3]. The estimated class-statistical parameters are then used with the Bayes decision rule for minimum error for identifying the decision threshold \( T \) that separates changed from unchanged patterns. However, if we apply the Bayesian threshold to \( I_p \), we obtain a change-detection map affected by the high uncertainty that characterizes patterns with a magnitude value close to the threshold. This problem arises from the loss of information associated with the magnitude operator. On the other hand, the threshold value \( T \) represents a reasonable reference point for identifying the subsets \( S_T \) and \( S_O \). According to this observation and following [4], the desired sets of patterns with a high probability to be correctly assigned to one of the two classes are obtained by defining a margin around the minimum-error threshold. This margin conceptually separates uncertain from certain patterns. Patterns that fall outside the margin and show a high magnitude have a high probability to be changed pixels and are labeled as targets, whereas patterns that fall outside the margin and show a low magnitude have a high probability to be unchanged and are labeled as outliers. Therefore, the resulting sets \( S_T = \{ x_n \in \mathbb{R}^N | i_n^\rho \geq T + \delta_1 \} \) and \( S_O = \{ x_n \in \mathbb{R}^N | i_n^\rho \leq T - \delta_2 \} \) (see Fig. 2) where \( i_n^\rho \) is a 1-dimensional pattern in \( I_p \) and \( x_n \) is a \( N \)-dimensional vector whose components are the spectral change vectors of the \( n \)-th pattern in \( I_3 \). According to the standard classification setup, the \( n \)-th target pattern in \( S_T \) is associated with label \( y_n = +1 \) whereas the \( n \)-th outlier pattern in \( S_O \) is associated with label \( y_n = -1 \). Constants \( \delta_1 \) and \( \delta_2 \) should be selected in order to obtain a high probability that
is a non-linear transformation that maps the data into a high dimensional Hilbert feature space $H$ where targets data description can be achieved with a hypersphere. Constraints in (2) come from the assumption that positive examples should fall inside the sphere, while outliers should fall outside (i.e., counter examples should be rejected) (Fig. 3).

As for standard SVMs, the cost function can be reformulated in order to allow a certain amount of errors in both the positive and negative samples sets. Let us introduce slack variables $\xi_t (t = 1, \ldots, K_T)$ and $\xi_o (o = 1, \ldots, K_O)$ associated with the target and outlier patterns, respectively (see Fig. 3). Accordingly, the error function to be minimized becomes:

$$
\min_{\alpha, \xi_t, \xi_o} \left\{ R^2 + C_T \sum_{t=1}^{K_T} \xi_t^2 + C_O \sum_{o=1}^{K_O} \xi_o^2 \right\}
$$

subjected to

$$
\begin{align*}
\| \phi(x_t) - a \| &\leq R + \xi_t, \quad \forall t = 1, \ldots, K_T \\
\| \phi(x_o) - a \| &\geq R - \xi_o, \quad \forall o = 1, \ldots, K_O
\end{align*}
$$

where $\phi(\cdot)$ is a non-linear transformation that maps the data.

The second step of the proposed method aims at giving a description of the class of changes (target) in the SCVs feature space by exploiting the information in the target and outlier sets defined in the previous step. The higher dimensionality that characterizes the multispectral difference image allows integrating the incomplete information on targets and outliers extracted from the 1-dimensional magnitude of SCVs and achieving a better description of the target class. The problem of finding a description of the target class data is faced here by achieving a better description of the target class. The problem extracted from the 1-dimensional magnitude of SCVs and sets defined in the previous step. The higher dimensionality

In greater detail, the SVDD defines a minimum volume hypersphere in the kernel space that includes all (or most) of the target patterns in the training set by minimizing a cost function.

Let us assume that after subsampling, the set of target pixels is made up of $S_T$ whereas the set of outliers is made up of $S_O$ counterexamples (i.e., $S_T = \{x_i\}_{i=1}^{K_T}$) whereas the set of outliers is made up of $K_O$ counterexamples (i.e., $S_O = \{x_o\}_{o=1}^{K_O}$). Following the notation in [11], let us assume that $x_t$ and $x_o$ in $S_T$ and $S_O$, respectively, are column vectors. We can characterize the MEB with its center $a$ and radius $R (> 0)$, and define the problem of minimizing its volume as:

$$
\min_{R, a} R^2
$$

subjected to

$$
\begin{align*}
\| \phi(x_t) - a \| &\leq R, \quad \forall t = 1, \ldots, K_T \\
\| \phi(x_o) - a \| &\geq R, \quad \forall o = 1, \ldots, K_O
\end{align*}
$$

Patterns in $S_T$ and $S_O$ are changed and unchanged, respectively.

**B. Change Detection Based on SVDD with Outlier Information**

![Fig. 2 Example of distribution of the magnitude of SCVs $p(i^p)$ and of definition of the targets ($S_T$) and outliers subsets ($S_O$).](image)

![Fig. 3 An example of hyperspherical boundary defined by SVDD: light grey samples (inside the sphere) are targets; and dark grey samples (outside the sphere) are outliers; and samples in the boundary are support vectors. Both target and outlier samples on the wrong side of the boundary are associated with slack variables to deal with errors.](image)
The given data description always results in a closed boundary around the target class. The inner product of mapping functions \( \phi(\cdot) \) (which are in principle unknown) that appears in (5) can be replaced with a kernel function \( K(\cdot, \cdot) \):

\[
K(x_i, x_j) = \left\{ \phi(x_i), \phi(x_j) \right\} \quad \text{with} \quad i, j \in \{o, t, p, u\} \tag{7}
\]

Substituting (7) into (5) we obtain

\[
\begin{align*}
\max_{\alpha, \alpha_c} & \quad \sum_{i=1}^{K_o} \alpha_i K(x_i, x_i) - \sum_{o=1}^{K_o} \alpha_o K(x_o, x_o) \\
& \quad - \sum_{i=1}^{K_t} \alpha_i \alpha_o K(x_i, x_o) + 2 \sum_{o=1}^{K_o} \sum_{i=1}^{K_t} \alpha_o \alpha_i K(x_o, x_i) \\
& \quad - \sum_{o=1}^{K_o} \alpha_o \alpha_p K(x_o, x_p) \end{align*} \tag{8}
\]

This allows us to construct a non-linear SVDD by defining only the kernel function, without needing to know the mapping \( \phi(\cdot) \) explicitly.

After solving the dual problem, to decide whether any vector \( x_n \) in \( I_o \) belongs to the class of change (targets) or no-change (outliers) the distance to the center of the sphere should be evaluated. A pattern \( x_n \) is classified as changed if it falls inside the sphere (i.e., its distance from the center of the sphere is lower than the radius), it is marked as unchanged.

The decision rules becomes as follows:

\[
x_n \in \begin{cases} 
\omega_c & \| \phi(x_n) - a \|^2 \leq R^2 \\
\omega_o & \| \phi(x_n) - a \|^2 > R^2
\end{cases}
\tag{9}
\]

### III. EXPERIMENTAL RESULTS

The proposed method has been tested on a multitemporal dataset made up of two multispectral images acquired by the Thematic Mapper (TM) multispectral sensor of the Landsat 5 satellite on a Mexico area in April 2000 (\( I_1 \)) and May 2002 (\( I_2 \)). The area selected for the experiments is a section of 512\( \times \)360 pixels of the two full scenes (Figs. 4 (a) and (b)). Between the two acquisitions, two wildfires occurred in this area. A reference map concerning their location was available (Fig. 4 (c)). This map includes 29506 changed pixels and 154814 unchanged pixels. A preliminary analysis pointed out that spectral channels 4 and 5 are the most relevant for discriminating the burned area on this data set. Accordingly, we used these channels in our trials.

In all the experiments the initialization threshold value \( T \) was obtained according to a manual-trial-and-error procedure (MTEP). The use of this threshold leads to a change-detection map that shows the minimum overall error if compared to the reference map. This choice allowed us evaluating an "upper bound" of the performance of the proposed method without any bias due to human operator subjectivity or to the fact that the selection was made automatically. At an operational level, any automatic threshold-selection technique can be adopted (see [3] for more details about thresholding the magnitude image). Constants \( \delta_t \) and \( \delta_o \) were set equal to 10.

In the learning of the SVDD three parameters should be defined: i) the width \( \sigma \) of the employed Gaussian kernel function [8]; ii) the regularization parameters \( C_T \) for wrong pixels in the target class; and iii) the regularization parameters \( C_O \) for wrong pixels in the outlier class. In our experiments, for convenience, we assumed that \( C_T = C_O \). The model-selection was based on the available reference map according to a grid search strategy with free parameters in the following ranges: i) regularization parameters \( C_T = C_O = C \in [10^{-3}, 5] \); and ii) Gaussian kernel width \( \sigma \in [0.01, 1] \).

The results obtained with the proposed method were compared with the ones obtained with a standard CVA algorithm applied to \( I_o \) with manual trial-and-error thresholding procedure. The comparison is performed in terms of false alarms, missed alarms, overall errors and kappa coefficient of agreement evaluated on the reference map.

As can be seen from Fig. 5, this change-detection problem is a complex problem, as target and outlier classes are significantly overlapped. In this critical situation the proposed technique resulted in higher change-detection accuracy than the standard technique (see Table I). In greater detail, the proposed approach sharp increased the Kappa accuracy provided by the standard CVA from 0.844 to 0.9028. The improvement is associated to a sharp decrease of false alarms (from 3840 to 2031) and to a slight reduction of missed alarms (from 3879 to 3141).

Numerical results are confirmed also by a qualitative comparison of the two change-detection maps reported in Fig. 6. As can be seen, the map obtained with the proposed technique shows less isolated changed pixels with respect to the map yielded with the CVA. Pixels affected by registration noise result in a high magnitude as changed pixels do, but they show different SCVs components, therefore they are properly classified as outliers in the higher dimensional feature space of the multispectral difference image, whereas they are not in the magnitude domain. Therefore, burned areas are better...
identified confirming the effectiveness of the proposed approach.

IV. DISCUSSION

In this paper, a novel approach to unsupervised change detection based on CVA and SVDD has been proposed. The proposed method formulates the change-detection problem as a minimum enclosing ball problem with changed pixels as target objects. The MEB problem is solved after mapping spectral change vectors into a high dimensional Hilbert space. Once the minimum volume hypersphere is computed, it is mapped back into the original feature space where it results in a non-linear flexible boundary around target pixels. With respect to standard change vector analysis that considers only the 1-dimensional magnitude of SCVs, the proposed techniques take advantage from the higher amount of information present in the multidimensional SCVs feature space. This results in a better identification of changed areas, particularly in problems with overlapping classes. Furthermore, focusing on the changed pixels, it allows reducing the impact of residual registration noise in the final change-detection map.

As future developments of the proposed work we propose: i) to introduce in the system architecture an approach for performing the model selection of the SVDD in a complete unsupervised way; and ii) to define a semi-supervised strategy for the learning of the SVDD parameters in order involve in the SVDD learning phase also unlabeled pixels according to the cluster assumption.

ACKNOWLEDGMENT

This paper has been partially supported by the Italian and Spanish Ministries for Education and Science under grant “Integrated Action Programme Italy-Spain”.

REFERENCES


<table>
<thead>
<tr>
<th>Technique</th>
<th>Missed Alarms</th>
<th>False Alarms</th>
<th>Overall Error</th>
<th>Kappa Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVA</td>
<td>3879</td>
<td>3840</td>
<td>7719</td>
<td>0.8440</td>
</tr>
<tr>
<td>Proposed</td>
<td>3141</td>
<td>2031</td>
<td>5172</td>
<td>0.9028</td>
</tr>
</tbody>
</table>