International Journal of Applied Earth Observation and Geoinformation 20 (2013) 77–85

Contents lists available at SciVerse ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Supervised change detection in VHR images using contextual information and support vector machines

Michele Volpi [a,*], Devis Tuia [b], Francesca Bovolo [c], Mikhail Kanevski [a], Lorenzo Bruzzone [c]

[a] Institute of Geomatics and Analysis of Risk, University of Lausanne, Switzerland
[b] Laboratory of Geographic Information Systems, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[c] Department of Information Engineering and Computer Science, University of Trento, Italy

ARTICLE INFO

Article history:
Received 1 December 2010
Accepted 31 October 2011

Keywords:
Change detection
Support vector machines
Graylevel co-occurrence matrix
Mathematical morphology
Very high resolution

ABSTRACT

In this paper we study an effective solution to deal with supervised change detection in very high geometrical resolution (VHR) images. High within-class variance as well as low between-class variance that characterize this kind of imagery make the detection and classification of ground cover transitions a difficult task. In order to achieve high detection accuracy, we propose the inclusion of spatial and contextual information issued from local textural statistics and mathematical morphology. To perform change detection, two architectures, initially developed for medium resolution images, are adapted for VHR: Direct Multi-date Classification and Difference Image Analysis. To cope with the high intra-class variability, we adopted a nonlinear classifier: the Support Vector Machines (SVM). The proposed approaches are successfully evaluated on two series of pansharpened QuickBird images.

1. Introduction

One of the most challenging Earth observation task is the identification of land cover transitions and changes occurred on a given region. Land cover evolutions can be identified by the analysis of two or more coregistered remote sensing images of the same geographical area at different times (Singh, 1989; Coppin et al., 2004).

Nowadays, many commercial and governmental instruments provide images within small temporal intervals with high to very high spatial resolutions. This type of imagery is appropriate for the study and the analysis of localized ground cover changes. In the literature, several methods have been developed for this purpose and efforts were put in considering low and medium resolution imagery. In the last decade, many studies aimed at transferring this knowledge to high and very high geometrical resolution (VHR) images.

This paper focuses on VHR images and on the adaptation of existing automatic classification techniques to discover changes. Change detection is considered as a supervised multi-temporal classification problem, which aims at obtaining a complete description of the transitions occurred between the acquisitions. Moving to VHR imagery comes with the price of increased within-class variances, that prevent the successful application of traditional classification methods such as the Maximum Likelihood classifier. In VHR the use of a robust and nonlinear classifier is mandatory since noise and generally higher spread in class distributions makes the classification problem very complex.

Support Vector Machines (SVM) classifiers (Vapnik, 1998; Schölkopf and Smola, 2002; Shawe-Taylor and Cristianini, 2004) have demonstrated their effectiveness in several remote sensing applications (Camps-Valls and Bruzzone, 2009). In particular, several researches addressed the problem of VHR ground cover classification using SVM (Bruzzone and Carlin, 2006; Inglada, 2007; Tuia et al., 2009). The success of such approaches is related to the intrinsic properties of this classifier: can handle ill-posed problems and to the curse of dimensionality (Hughes, 1968), provides robust sparse solutions and delineates nonlinear decision boundaries between the classes.

Recently, kernel methods started to be considered also for change detection and multi-temporal classification. Despite the promising results in many remote sensing tasks, only few studies deal with change detection. In Nemmour and Chibani (2006) supervised multi-temporal classification is implemented using SVM. In their setting, two coregistered images are stacked and the bi-temporal dataset is classified with a multiple SVM approach. The comparison with a Neural Networks classifier proved that SVM are less prone to overfit the data and training issues related to non-convex error functions are avoided. Bovolo et al. (2008) perform transductive SVM for change detection initialized with a Bayesian selective thresholding method (Bruzzone and Fernández-Prieto, 2000) that allows the unsupervised application of this classifier.

* Corresponding author. Tel.: +41 21 6923546; fax: +41 21 6923535.
E-mail address: michele.volpi@unil.ch (M. Volpi).
URL: http://www.kernelcd.org (M. Volpi).

0303-2434/$ – see front matter
doi:10.1016/j.jag.2011.10.013

The final performance obtained outperformed classical change vector analysis. Bovolo et al. (2010) reformulated the change detection task as an outlier detection problem, modeling the target (changed patterns between the two times) via Support Vector Domain Description and detecting unchanged pixels as outliers. The superiority of the nonlinear approach was proven by their experiments.

As mentioned, in VHR images the underlying class distributions are often strongly overlapped, resulting in hardly classifiable pixels even using robust methods as SVM. The high within-class variance as well as the low between-class distance, due to the low spectral information, increase the need for approaches that enhance separability between the different classes. To solve this issue, contextual features providing information on the spatial relationships of pixels have been extensively studied for standard classification.

Spatial context features are often considered to ease the classification process of VHR images. Murray et al. (2010) proved that the joint use of spectral and textural features ameliorates the classification accuracy of VHR images considerably. On the opposite, classification performed using only spectral or textural features results in lower performance. In Tuia et al. (2009), different multi-scale morphological features are extracted and studied to classify QuickBird panchromatic images (thus with poor spectral resolution) using SVM. In Pacifici et al. (2009) local textural measures based on the Gray Level Co-Occurrence Matrix (GLCM) are studied for classifying VHR panchromatic images with a Neural Networks classifier. In Tuia et al. (2010b), specific kernel functions are designed to find optimal combinations of contextual information at relevant spatial scales. Summing up, these studies verify that the lack of spectral information is successfully balanced by the inclusion of contextual information.

The exploitation of spatial information is poorly documented in change detection literature, although the benefits of considering such variables are clearly demonstrated in classification tasks. In Dalla Mura et al. (2008) the advantages of including morphological reconstruction operators in the change vector analysis framework (Bovolo and Bruzzone, 2007) has been illustrated. By filtering the magnitude of the difference image (as an intermediate step), errors due to radiometric differences and noise are greatly reduced. In Bovolo (2009) a contextual parcel-based multi-scale approach to unsupervised change detection is presented. The usefulness of contextual information in VHR unsupervised change detection is clearly pointed out by these studies.

In this paper, we propose an effective way to deal with supervised change detection in VHR images by integrating spatial information in SVM multi-temporal classification. As introduced, it is already proven that the pixel context characteristics can provide accurate and coherent classification maps by filling the lack of spectral information. On the other hand, SVM are suitable tools for many remote sensing applications, thanks to their intrinsic properties. The rationale of this paper is to combine the advantages of both SVM and contextual information and to prove their benefits for supervised change detection in VHR images. This aims at mitigating class separability problems by completing the feature vector, and discovering the optimal nonlinear classification boundaries with SVM. Two change detection architectures are considered: Direct Multi-date Classification (DMC) and Difference Image Analysis (DIA).

The remainder of the paper is organized as follows: Section 2 introduces the reader to the extracted features, to the classifier and to the change detection architectures. Section 3 presents the datasets as well as the experimental setup. Section 4 presents results, Section 5 discusses the outcomes and Section 6 draws the conclusions of the paper.

2. Context-based supervised change detection

The contextual features are extracted for each scene and then combined in a specific multi-temporal classification scheme. This section presents the considered contextual features, the SVM classifier and the adopted change detection architectures.

Notation. Let $\mathbf{X}$ be a multi-temporal set representing a composition of the two multi-spectral images $\mathbf{X}^1$ and $\mathbf{X}^2$ acquired at different time instants $t=1$ and $t=2$. Classes are discriminated on the basis of a set of labeled multi-temporal pixels, composed by pairs $\{\mathbf{x}_i, \omega_i\}_{i=1}^N$, accounting for the D-dimensional multi-temporal spectral vectors $\mathbf{x}_i \in \mathbb{R}^D$ and $\Omega = \{\Omega_U, \Omega_C\}$, that is the set of L transitions associated to changes $\Omega_C = \{\omega_1, \ldots, \omega_L\}$ and S stable ground cover (no-change) at the two times $\Omega_U = \{\omega_{C+1}, \ldots, \omega_{L+S}\}$.

2.1. Textural features (TXT)

Occurrence and co-occurrence textural statistics (Haralick et al., 1973; Baraldi and Parmiggiani, 1995) are local indexes computed on the basis of moving windows of size $P \times Q$ (usually $P=Q$). The resulting variables emphasize the texture structure of the graylevel image. The considered image to retrieve such metrics can be of different forms: in the case of multi-spectral VHR scenes it is common to use the panchromatic band, the first principal component or a discriminative band.

Occurrence statistics. These measures are computed on the graylevel values contained in the sliding window centered on the pixel x_{ij}. They return a local texture value $\hat{x}_{ij}$. Two occurrence indicators are considered, local mean and variance:

$$\hat{x}_{ij}^{\mathrm{ME}} = \frac{1}{PQ} \sum_{p,q \in \mathcal{V}} x_{pq} \tag{1}$$

$$\hat{x}_{ij}^{\mathrm{VAR}} = \frac{1}{PQ} \sum_{p,q \in \mathcal{V}} (x_{pq} - \hat{x}_{ij}^{\mathrm{ME}})^2, \tag{2}$$

where $\mathcal{V}$ denotes the pixels contained in the window centered on x_{ij}. The ME feature returns the local average of the pixels in $\mathcal{V}$. Considering this variable reduces effects of noise and outliers (*e.g.*, saturated pixels), by smoothing extreme values. The local variance (VAR) indicator summarizes high differences in the graylevel values contained in the considered patch, emphasizing edges between objects at different scales. Other indicators such as skewness or kurtosis can be considered (Haralick et al., 1973).

Co-occurrence statistics. These indicators are based on the Graylevel Co-occurrence Matrix (GLCM), that represents the relative occurrence frequency $p(m, n)$ of two graylevel intensities m and n in the $P \times Q$ window at a given angular neighborhood. The lag is given by a connecting vector (δ_x, δ_y) in x and y spatial coordinates. On the basis of the GLCM many statistical texture descriptors can be extracted (Haralick et al., 1973; Petrou and Sevilla, 2006). In this paper three second moment descriptors are adopted: entropy (ENT), angular second moment (ASM) and homogeneity (HOM).

$$\hat{x}_{ij}^{\mathrm{ENT}} = -\sum_{m}\sum_{n} p(m, n) \log p(m, n) \tag{3}$$

$$\hat{x}_{ij}^{\mathrm{ASM}} = \sum_{m}\sum_{n} p(m, n)^2 \tag{4}$$

$$\hat{x}_{ij}^{\mathrm{HOM}} = \sum_{m}\sum_{n} \frac{p(m, n)}{1 + |m - n|}. \tag{5}$$

ENT is a measure of information content and can be interpreted as a measure of the randomness of the graylevel values. Regions with high variance will result in high entropy values, while smooth patches represent low entropy. ENT is a good indicator of the intensity of the texture in the considered patch. ASM indicates the local contrast. It provides an accurate estimate on the degree of uniformity of the values of the GLCM. A low ASM value indicates that no spatial coherence (texturing) characterizes the patch. HOM measures the variance around the diagonal of the GLCM. In homogeneous patches, the values are clustered around the diagonal resulting in high values. Other GLCM-based indicators can be used, such as correlation or contrast (Haralick et al., 1973).

2.2. Mathematical morphology

Since texture can be similar for different regions of the image, texture statistics can present similar ranges for different but same textured classes. To solve this issue, the joint use of texture indicators with multi-band morphological profiles (Benediktsson et al., 2005; Fauvel et al., 2008) is proposed. The mathematical morphology framework (see Soille and Pesaresi (2002), Soille (2004) for details) defines a family of operators that aim at emphasizing homogeneous spatial structures in a graylevel image. The resulting variables present higher autocorrelation for neighboring pixels in the same object, reducing noise, inner-class variance and, since a multi-band approach is adopted, increasing the between-class variance. These filters are based on a moving window of given shape and size called the structuring element S.

Basic operations are erosion and dilation, respectively denoted as $\epsilon_S(x_{ij})$ and $\delta_S(x_{ij})$. They are defined as follows:

$$\epsilon_S(x_{ij}) = \min\{x_{ij}, x_s\} \quad \forall x_s \in S_{ij} \tag{6}$$

$$\delta_S(x_{ij}) = \max\{x_{ij}, x_s\} \quad \forall x_s \in S_{ij}, \tag{7}$$

that return, respectively, the minimum and the maximum value between pixel x_{ij} and the ones contained in the structuring element S_{ij} centered on x_{ij}.

Opening and closing (OC). These two filters are the concatenation of erosion and dilation:

$$\gamma_S(x_{ij}) = \delta_S(\epsilon_S(x_{ij})) \tag{8}$$

$$\phi_S(x_{ij}) = \epsilon_S(\delta_S(x_{ij})). \tag{9}$$

The opening $\gamma_S(x_{ij})$ of the graylevel image filters out elements that are brighter than their surroundings (in the span of the structuring element S). Closing $\phi_S(x_{ij})$ filters out darker elements in the same range.

Opening and closing by reconstruction (OCR). Although emphasizing meaningful contextual information, opening and closing do not preserve the shape of objects represented in the image. To provide information at precise object level, recent studies propose the use of reconstruction filters (Soille, 2004; Fauvel et al., 2008).

Opening and closing by reconstruction are noted as $\rho_{\delta_S}(I_M)$ and $\rho_{\epsilon_S}(I_M)$ respectively. These operations reconstruct the original image by iterative cycles of erosions or dilations on a marker image I_M. If the initial marker image I_M is an erosion of the original image ($I_M = \epsilon_S(x_{ij})$), and the original image is reconstructed by iterative series of dilations of I_M as $I_M^k = \delta^1\delta^2\delta^3 \ldots \delta^k(I_M)$, the resulting filter is opening by reconstruction:

$$\rho^k_{\delta_S}(\epsilon_S(x_{ij})) = \min\{I_M^k, x_{ij}\} \tag{10}$$

and the process is iterated until $\rho^k = \rho^{k-1}$. Similarly, closing by reconstruction reconstructs the graylevel image starting from its dilated version $I_M = \delta_S(x_{ij})$ iteratively performing erosions of the marker image I_M as $I_M^k = \epsilon^1\epsilon^2\epsilon^3 \ldots \epsilon^k(I_M)$:

$$\rho^k_{\epsilon_S}(\delta_S(x_{ij})) = \max\{I_M^k, x_{ij}\}, \tag{11}$$

converging to the final filtering when $\rho^k = \rho^{k-1}$. As for the OC operators, opening and closing by reconstruction filter out brighter and darker elements smaller than S_{ij}, but preserving original shapes of spatial structures.

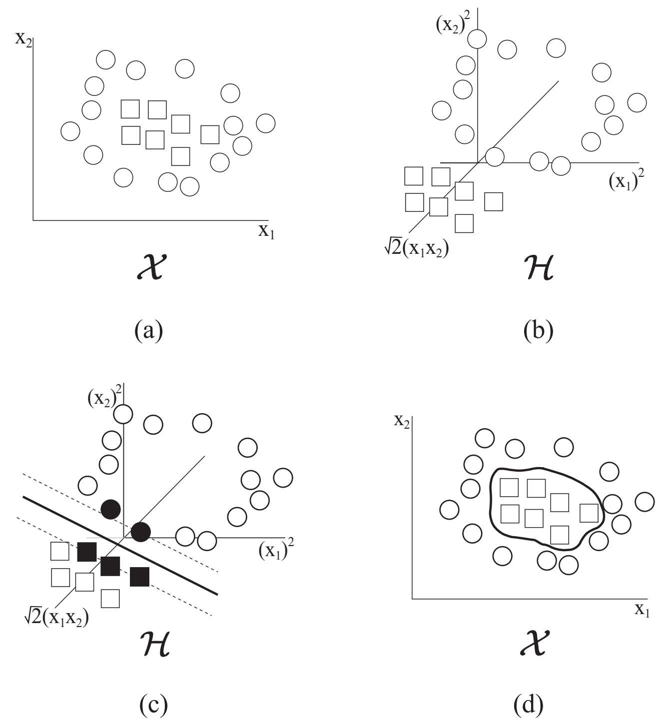

Fig. 1. Nonlinear classification by SVM. (a) A non-linearly separable dataset in $\mathcal{X}$ is implicitly projected in a higher dimensional space $\mathcal{H}$ (b). In $\mathcal{H}$, linear separation is possible (c), and corresponds to a nonlinear solution in $\mathcal{X}$ (d).

2.3. The support vector machines for classification

Once the set of features to be involved in the change detection problem has been defined, a robust classifier should be selected for the supervised classification step. SVM are chosen thanks to their intrinsic robustness to high dimensional datasets and to ill-posed problems.

SVM are a nonparametric supervised classifier relying on Vapnik's statistical learning theory (Vapnik, 1998). This classifier aims at building a linear separation rule between examples in a higher dimensional space induced by a mapping function $\boldsymbol{\varphi}(\cdot)$ on training samples. A linear separation in that space corresponds to a nonlinear separation in the original input space. An example is illustrated in Fig. 1(a)–(d). The core of such algorithm is given by the kernel trick: since in the SVM formulation mapped samples appear only in the form of dot products, these operations can be replaced by valid *kernel functions* $k(\cdot, \cdot)$ returning directly the inner product value in that space (dual formulation, Eq. (12)). The solution is given by the hyperplane with maximal margin width, that guarantees best generalization ability on previously unseen data. In the dual optimization formulation one has to optimize (Boser et al., 1992):

$$\max_{\boldsymbol{\alpha}} \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j\omega_i\omega_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{12}$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C \quad \text{and} \quad \sum_{i=1}^{N} \alpha_i\omega_i = 0,$$

where C is a user defined parameter controlling the trade-off between complexity and training error of the model, α_i are the

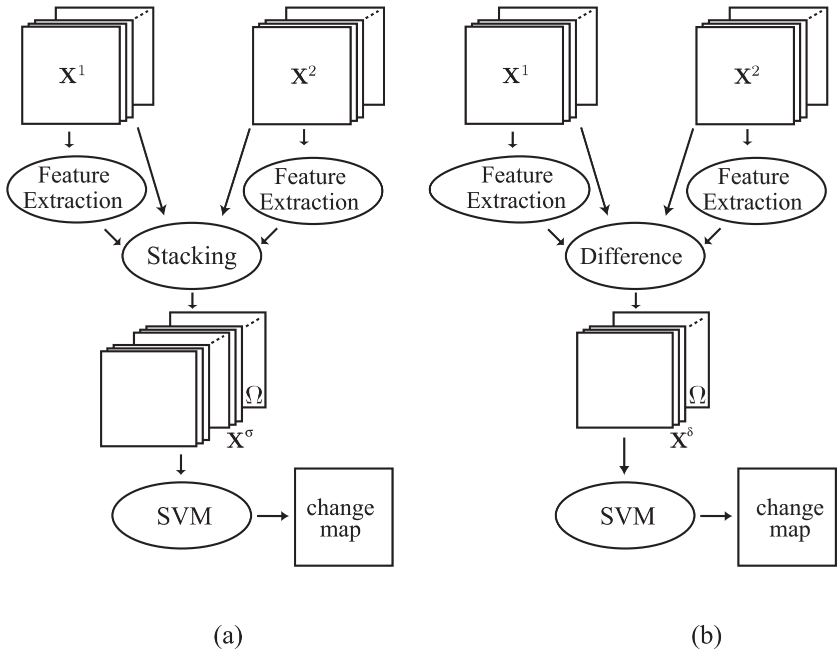

Fig. 2. Direct Multi-date Classification (a) and Difference Image Analysis (b) schemes.

coefficients determining the solution of the optimization and $\omega_i \in \{+1; -1\}$ (binary case) are the class labels associated to samples $\mathbf{x}_i$.

When the solution to Eq. (12) is found, the label of an unknown sample $\mathbf{x}'$ is given by the sign of the decision function, *i.e.*, its position with respect to the separating hyperplane:

$$\omega\prime = \text{sign}\left(\sum_{i=1}^{N} \alpha_i \omega_i k(\mathbf{x}_i, \mathbf{x}\prime) + b\right). \tag{13}$$

Experiments are performed using a Gaussian RBF kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-||\mathbf{x}_i - \mathbf{x}_j||^2/2\sigma^2)$, where σ is the user defined bandwidth of the Gaussian function. Many kernel functions exist, as the polynomial one, but in environmental applications it is common to use the Gaussian RBF thanks to its interpretability (cast as a local similarity) and to the positive performances already shown in many application fields (Kanevski et al., 2010). To solve multi-class problems the one-against-all scheme is adopted (Shawe-Taylor and Cristianini, 2004).

2.4. Considered change detection architectures

In order to effectively take advantage of the described features and classifier, proper approaches to change detection should be defined. Hereafter, two architectures are presented: (i) direct multi-date classification and (ii) difference image analysis.

Direct Multi-date Classification (DMC). In DMC, the two single time images $\mathbf{X}^1$ and $\mathbf{X}^2$ are stacked into a single multi-temporal set $\mathbf{X}^s = \mathbf{X}^1 \bigcup \mathbf{X}^2$ and classified on the basis of an exhaustive multi-temporal labeling. A flowchart illustrating the proposed approach is shown in Fig. 2(a). This approach produces a complete map reproducing all the occurred transitions represented in the training set. The main bottleneck of DMC is the creation of high dimensional datasets due to variable stacking. This may cause problems related to the curse of dimensionality (Hughes, 1968). On the other hand, all the available information is preserved, guaranteeing that no loss of information may harm the process.

Difference image analysis (DIA). In this case, the dimensionality of the problem is kept low by considering the multivariate difference of images $\mathbf{X}^d = \mathbf{X}^2 - \mathbf{X}^1$. Fig. 2(b) illustrates the DIA approach. As all unchanged pixels result in similar spectral differences (with $\mathbf{X}^d \approx 0$), the land cover class of such pixels cannot be modeled. In other words $\Omega_U = \omega_1$. On the contrary, those showing a difference vector far from 0 in at least one spectral band have a high probability to be associated to a transition in ground cover (Bruzzone and Fernández-Prieto, 2000; Bovolo and Bruzzone, 2007). When working with few spectral variables, this approach may present an ambiguity problem as same $\mathbf{X}^d$ values may correspond to different transitions. However, for our case studies relying on multi-spectral imagery and by further adding contextual variables, this issue does not harm the DIA-based change detection process.

In order to allow fair comparisons with the DIA, where unchanged pixels are treated as single class, a third approach referred to as *reduced DMC* is also considered: in this case, all the samples representing unchanged classes are assigned to the class 'no change' $\Omega_U = \omega_1$, and change detection is performed as for the complete DMC scheme.

3. Datasets and experimental setup

To validate the proposed architectures, two datasets are considered. Both scenes are subsets of two multi-spectral pansharpened QuickBird images of the city of Zurich, Switzerland, with a ground sample distance of roughly 0.7 m. The first is acquired in August 2002 and the second in October 2006.

3.1. Brüttisellen

The Brüttisellen multi-temporal images have size of 521 × 1188 pixels, accounting for NIR-R-G-B channels. By visual inspection, a total of 9 land cover classes and transitions has been detected, of which 3 are changes and 6 no change (see Fig. 3). The test set, used to estimate the generalization abilities of the proposed schemes, is composed by 76,185 spatially independent pixels. The test regions are spatially disjoint to avoid spatial autocorrelation with training samples and consequent overestimation in the generalization accuracy.

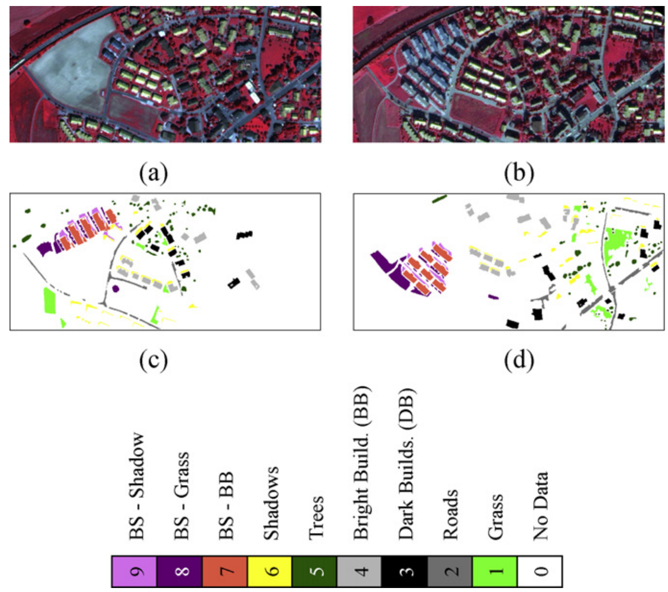

Fig. 3. The Brüttisellen dataset. In (a) and (b) respectively the acquisitions in 2002 and in 2006 in false color representation (NIR-R-G). In (c) and (d) respectively the regions used for extracting the training sets and for testing the generalization ability. In the legend BS refers to bare soil.

A large part of the images is unchanged, in which differences can only be observed at illumination and sun elevation levels. These regions present nonlinear characteristics typical of VHR images. The changed regions concern a group of houses in a bare soil region, which generates changes also in vegetation classes. The scene is challenging since bare soil can partially dissimulate radiometric changes related to newly constructed buildings, while other changes are related to transitions in grassland and shadow. The different acquisition times do not raise issues related to phenological differences (in grass and trees classes). In our classification setup, vegetation is considered in a wide sense and within-class changes are not modeled. Fig. 3 illustrates the datasets and the training/testing regions.

3.2. Steinacker

The second dataset, called "Steinacker" is composed of two pansharpened QuickBird images acquired in the same period as for the "Brüttisellen" dataset. The scenes account for 4 classes related to ground cover change and 6 no change classes, both discovered by visual inspection of the two 784×649 scenes (see Fig. 4). The spatially independent test set accounts for 58,293 samples.

Transitions related to cultivated crop (vegetated and not), to the construction of buildings over vegetated and bare soil regions characterize the scene. The rest of the image presents differences in reflectance due to the sun elevation level and small changes due to urban dynamics. Fig. 4 illustrates the datasets and the training/testing regions.

3.3. Experimental setup

Textural features are computed on the corresponding panchromatic bands (one for each time instant). For each occurrence statistic, three window sizes are considered (3×3, 7×7 and 15×15), resulting in 6 variables. Regarding co-occurrence indicators, the average of the statistics computed in four directions (0°, 45°, 90° and 135°), with a shift in horizontal and vertical directions proportional to the moving window size, are considered. The reason of considering the average on four directions is that, since

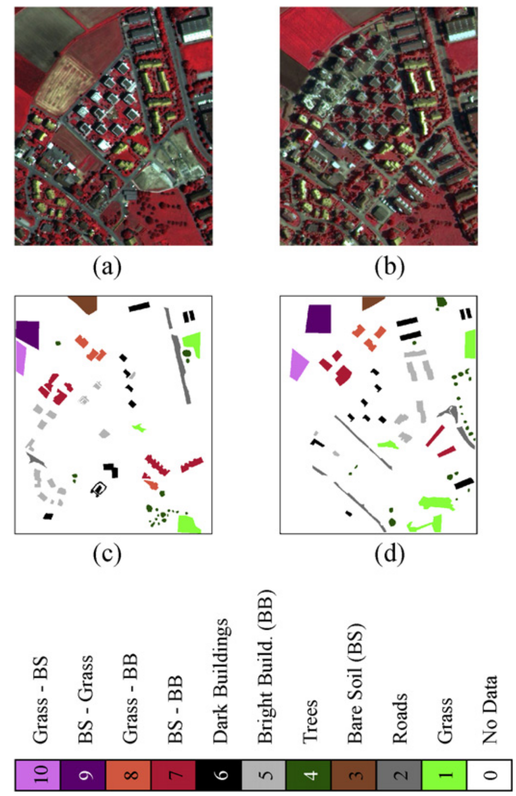

Fig. 4. The Steinacker dataset. In (a) and (b) respectively the acquisitions in 2002 and in 2006 in false color representation (NIR-R-G). In (c) and (d) respectively the regions used for extracting the training sets and for testing the generalization ability.

the GLCM-based indicators are symmetric (*e.g.*, $\hat{x}_{ij}(0^\circ) = \hat{x}_{ij}(180^\circ)$), their average is invariant under rotation. Three window sizes have been utilized for computing the GLCM (3×3 with a shift of 1 pixel, 7×7 with a shift of 2 pixels and 15×15 with a shift of 4 pixels) resulting in 9 co-occurrence variables. The choice of the window size is related to the resolution of the objects represented in the scene. To preserve the level of details, 3×3 pixels windows are considered (roughly corresponding to squares of 2 m of side), providing information on small patches as trees and small buildings, along with abrupt variations in object borders. The 7×7 window accounts for local structures in a range of 5 m, including information at building and road level, as well as smooth changes among different texture classes. Finally, the 15×15 window provides textural information for larger regions (around 10 m) accounting for large trends in fields and grasslands as well as commercial buildings. Larger windows are not considered since the scenes are mainly characterized by small and medium sized objects.

For each scene, morphological operations are considered with three different disk-shaped structuring elements, with radius 3, 7 and 9 pixels, independently for *all the spectral channels of the images*. The series of features with growing window sizes provide explicit multi-scale information to the change detection schemes. The size of the structuring element is again proportional to the size of the object of interest. The sets of features are summarized in Table 1.

To better understand the role of the spatial-contextual information within the process of supervised change detection, blocks of features and their combinations are tested independently and in growing order. For each feature block, eight experimental conditions are tested, accounting for different sizes of the training sets: 5, 10, 20, 50, 100 and 200 labeled examples per class, randomly extracted from the training ground truth. The size of the sets varies from very small to large, and for the smaller ones the dimensionality is often higher than the number of training

Table 1
Features blocks utilized in both experiments. The number of the features refers to a single date. For both dates, same features with same parameters are extracted. For each set of features, the pansharpened image is included (+4, the IMM set).

Set name	Dimensions	Description
IMM	4	Pansharpened bands
TXT	15 (+4)	6 occurrence and 9 co-occurrence
OC	24 (+4)	Opening and closing
OCR	24 (+4)	Opening and closing by reconstruction
OCOCR	48 (+4)	OC and OCR stacked
OCTXT	39 (+4)	OC and TXT stacked
OCRTXT	39 (+4)	OCR and TXT stacked
OCOCRTXT	63 (+4)	OC, OCR and TXT stacked

samples (*e.g.*, the Brüttisellen OC set accounts for 56 multi-temporal features and just 45 training samples for 9 classes in the smallest complete DMC setting). Classification results are very sensitive to the representativeness of training set, since in many cases the multi-temporal classification constitutes an ill-posed problem. To have robust statistical estimates, results are averaged on 10 independent experiments.

SVM hyper-parameters are selected by a 3-fold cross-validation. The C parameter is selected by exhaustive search in the range $C = [1, 10, 20, \ldots, 1000]$. To mitigate overfitting, in particular for small training sets, an initial guess σ_p on the Gaussian kernel bandwidth has been obtained by computing the median distance on 3000 randomly chosen pixels in the image. A refined search around this initial guess, in $\sigma = [0.5 \times \sigma_p, \sigma_p, 1.5 \times \sigma_p]$, has been performed and the parameters producing minimal error were retained. The free Torch 3 machine learning library is used to solve the SVM classification problem (Collobert et al., 2002).

Generalization accuracy is evaluated in terms of estimated Cohen's Kappa statistic (κ) (Foody, 2004) on the average of 10 independent trials. In order to assess significance of differences in accuracy, the McNemar test (Foody, 2004) is reported in Table 2. This table shows if the average accuracy is significantly higher (+), lower (−) or statistically similar (o) to the one obtained using the pure spectral baseline set (IMM).

4. Results

4.1. Brüttisellen results

The accuracies for the Brüttisellen experiments are reported in Fig. 5(a)–(c) as a function of the number of training samples per class.

The complete DMC on the IMM feature set shows an average estimated κ statistic of 0.77 when training the SVM with 5 samples per class. Then it increases to a κ of 0.89 points for the experiments using 200 training samples per class. Only the TXT set performs worse, in the ill-posed setting, and then equals the IMM results for larger training sets. Globally, the contextual sets of features, and in particular the composite textural-morphological, perform better with respect to the baseline set, with improvements in the range of 0.05 κ for small sized training sets and of 0.015 for the largest ones. The McNemar test reported in Table 2 indicates that, except for the TXT set, all the contextual features improve significantly the DMC results without contextual information.

The reduced DMC shows similar trends. It is worth mentioning that, since the number of classes is different (4 instead of 9), no direct comparisons on the absolute accuracies observed above can be made. Again, the baseline IMM set performs worse than the others. The contextual information improves significantly the change detection accuracies, except for the TXT set that on the average does not improve the process. The improvement in accuracy is roughly around a κ score of 0.1 points for small training sets, reducing to 0.03–0.01 for the large ones.

Regarding DIA, different observations can be made. As in the previous experiments, the spectral IMM feature set performs worse than the others. The bad performance is confirmed by the significantly lower IMM average accuracy. An interesting observation can be made by observing the performance of morphological sets. The three sets (OC, OCR and OCOCR) have similar κ scores and standard deviations. From a training set composed of 10 samples per class on, textural and morphological combined features perform better than the rest, improving in average the accuracy of about 0.15–0.18 κ points (IMM set vs OCOCRTXT). Thus, the texture seems an important information to mitigate the ambiguity of the spectral change vector representations, greatly reducing the false alarm rate.

Additional observations can be made by comparing the reduced DMC and the DIA schemes. The difference in accuracy, around 0.03–0.07 κ points, can be related to the completeness of the multi-temporal signal contained in the stacked vectors of the reduced DMC approach. On the other hand, even if the accuracy provided by the DIA architecture is lower, the dimensionality of the dataset is exactly the half (given the same feature set), thus reducing the computational charge.

In Fig. 6 the improvements with respect to the basis spectral classification (IMM set) are illustrated for training sets composed of 50 samples per class. This size is chosen since a plateau effect on the accuracy is observed from this set size on. The details of the

Table 2
McNemar tests outcomes. The +indicates that the tested set of features is significantly better that the baseline IMM set with $z > 1.96$ at $\alpha = 0.05$ level, while − indicates that IMM is better than the compared approach $z < -1.96$. The o indicates no significant difference.

Method	Complete DMC						Reduced DMC						DIA					
Per class size	5	10	20	50	100	200	5	10	20	50	100	200	5	10	20	50	100	200
Brüttisellen																		
TXT	−	−	−	o	o	o	o	+	+	−	o	+	+	+	+	+	+	+
OC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
OCR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
OCOCR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
OCTXT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
OCRTXT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
OCOCRTXT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Steinacker																		
TXT	−	−	−	+	+	+	−	−	−	−	−	o	o	+	+	+	+	+
OC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
OCR	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+
OCOCR	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
OCTXT	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+
OCRTXT	+	+	+	+	+	+	o	+	+	+	+	+	+	+	+	+	+	+
OCOCRTXT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

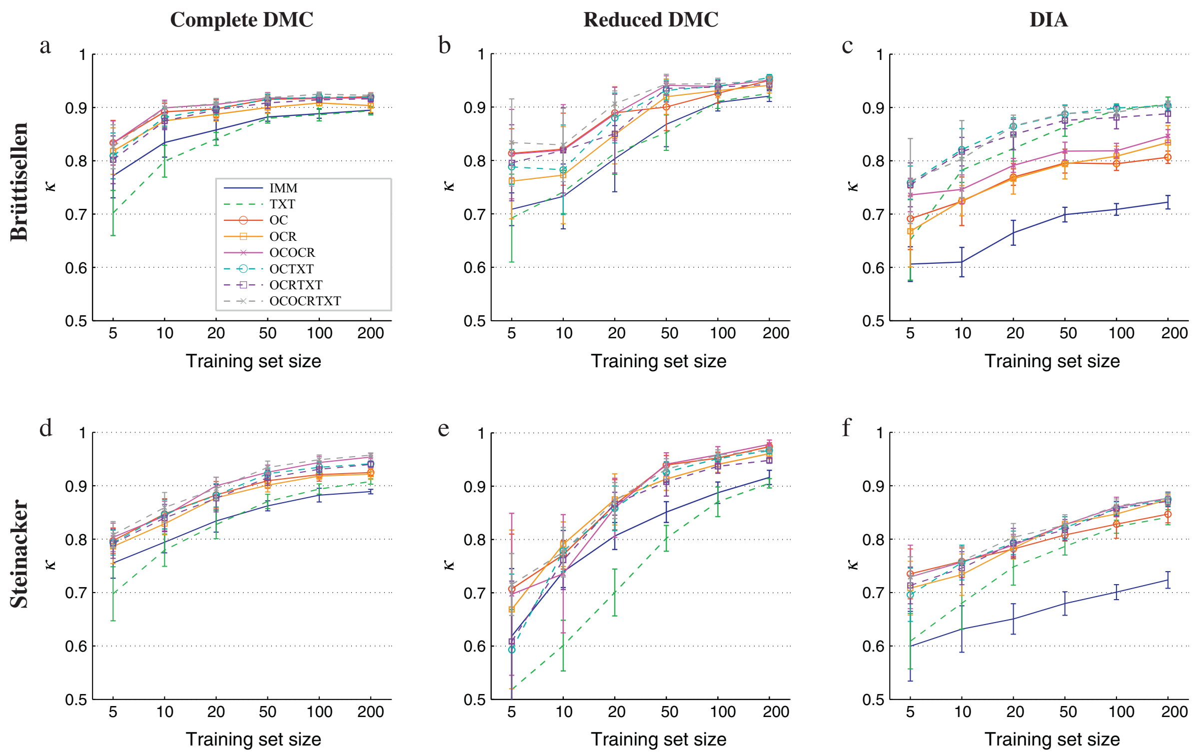

Fig. 5. Test accuracies for the considered datasets as a function of the per class training set size: Brüttisellen (a–c) and Steinacker (d–f).

change detection maps show an improved spatial coherence when adding spatial contextual features (OCOCRTXT).

4.2. Steinacker results

Experiments on the Steinacker dataset are conducted with the same setup as for the Brüttisellen images.

As observed for the previous dataset, the complete DMC performances of IMM and TXT sets are significantly lower than the other tested features. Table 2 reports the outcome of the McNemar test. Regarding the κ scores curves, it can be seen that, for training sets larger than 20 samples per class, standard deviations are very low, in the range of 0.01–0.001. This is an indicator of stable classification models. The morphological and the composite feature sets (in particular OCOCRTXT and OCOCR) outperform other sets, uniformly improving of about 0.04–0.08 the accuracy of the IMM set.

In the reduced DMC setting, the TXT feature set provides poor results (significantly worse than the IMM features) for each different training set size. The baseline IMM performs in the range of the other sets when considering 5 and 10 examples per class, then worse from 20 samples per class on. Better accuracies are obtained by models that include contextual information, improving the κ coefficient of 0.05–0.1 with respect to the baseline set. In general high standard deviations affect small training sets (5, 10 and partly 20 examples per class) indicating model instability. Morphological and textural-morphological composite sets show very similar behaviors, providing high change detection accuracies in the range of 0.95 κ points for large training sets.

Regarding DIA models, trends are similar to those observed on the previous dataset. The IMM set performs constantly worse than the rest and only pure texture information (TXT set), with the increase of the training sets size, shows a great improvement rate. All tested variables, except TXT with 5 training samples per class are significantly better than the IMM. Morphological and composite sets behave very similarly indicating again the appropriateness of this information for the DIA setting.

As for the previous experiments, the differences between reduced DMC and DIA are related to the loss in information that may harm the difference image. In Fig. 6 details of the change detection maps produced with training sets of 50 samples per class are reported. The spatial coherence of the basic spectral change detection map is greatly improved by the inclusion of contextual information. In this case, benefits of adding morphological features (OCOCR) to the basic set are shown.

5. Discussion

The experiments on the VHR multi-temporal datasets provided interesting insights about the inclusion of spatial context information in the process of supervised change detection. Observing Table 2, it is clear that considering such information significantly improves the accuracy of the process. The complete DMC setting has the advantage of predicting a complete change detection map by shattering each stable class and transition separately. If the ground truth has been created carefully the different classes are unimodal and separability is further increased by including spatial information. The usefulness of the pixel context is also beneficial for obtaining smooth change detection maps, eliminating spurious changes and thus reducing the false alarm rate, as shown in change detection map details in Fig. 6.

Regarding the reduced DMC setting, performance is also high, but a problem arises when the training set is small, illustrated by the high variances of the outcomes. This is mainly due to the multimodal distribution of the no change class, that becomes sparse and clustered in the feature space. Thus, SVM need many training samples to discover correct separating hyperplanes for this class. Once this is ensured, this scheme provided the highest accuracies.

For the DIA approach it can be noticed that the inclusion of composite contextual information is always beneficial, reducing the effects of ambiguity and increased class overlapping. The comparisons with the reduced DMC scheme suggest that DIA can provide high accuracies by utilizing only textural information, thus

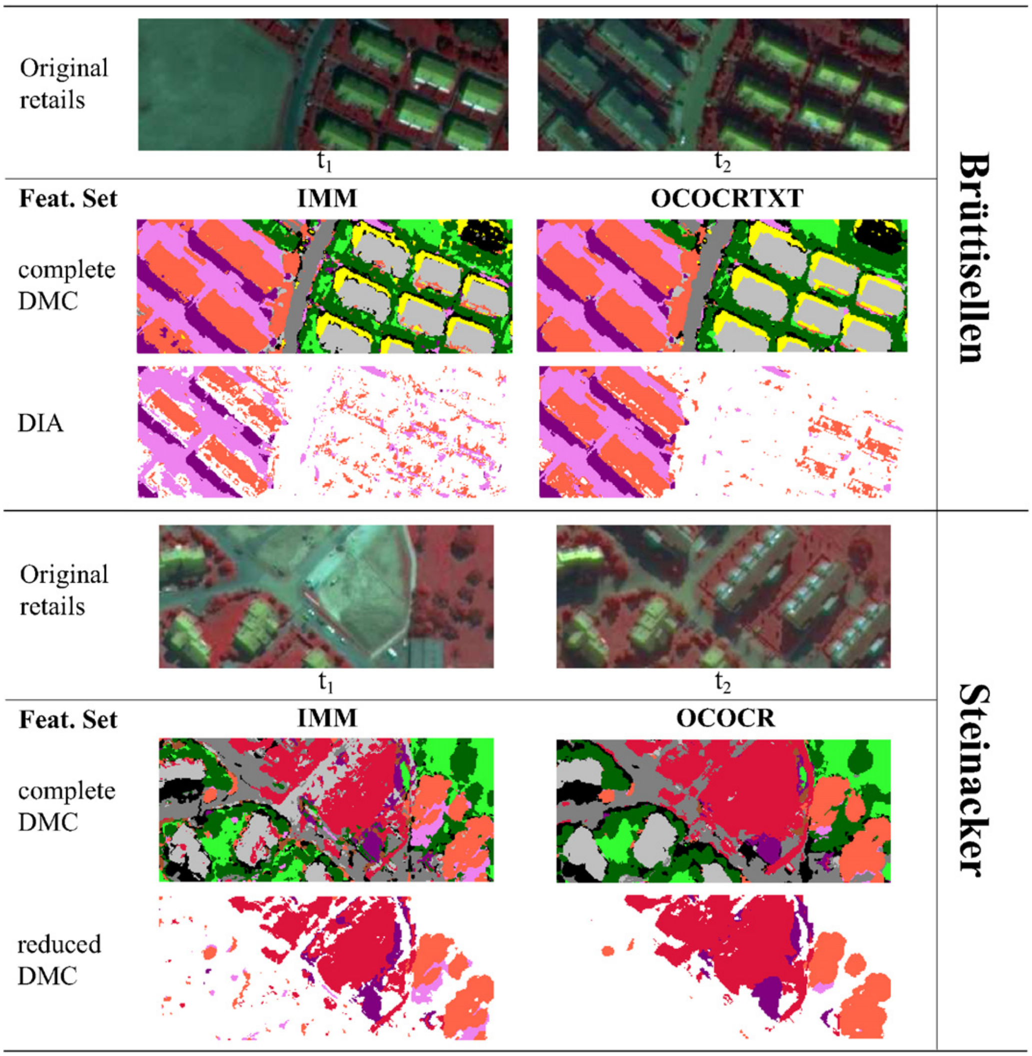

Fig. 6. Details of the Brüttisellen and Steinacker change detection maps. For the legend please refer to Fig. 3 (Brüttisellen) and Fig. 4 (Steinacker).

allowing the use of simpler classifiers due to the lower dimensionality of the dataset, assuming increased separability when considering pixel context.

When only few samples compose the training set, the dimensionality is often higher than the number of samples. Even if SVM are robust to the Hughes effect (Hughes, 1968), one has to control the N/D ratio (number of samples/dimensions) by providing enough samples to model correctly the class boundaries. In the experiments it is shown that in our case the N/D ratio should not be lower than 0.6–0.7 to have a stable solution. This fact is underlined by the decrease of the standard deviation with the increase of training samples, indicating stable models. However, note that the half of the considered training sets are too small for many classifiers. Hence SVM classifiers are strongly recommended due to their robustness against the curse of dimensionality. Nowadays, since SVM are standard tools in many classification tasks, many free packages become available for download. The use of other classifiers coupled to spatial information can be foreseen, provided an adequate number of training samples. For instance, the linear discriminant classifier needs at least $2 \times D$ training samples (N/D ratio of 2) to estimate unbiased class statistics and $N = D + 1$ samples per class to mitigate the singularity of the within-class scatter matrix.

Regarding computational complexity of SVM, it is dominated by the number of samples composing the training set, which controls training time. To keep a low computational burden, a careful extraction of an exhaustive training set as small as possible is suggested.

6. Conclusions

In this paper the usefulness of textural and morphological features has been demonstrated in the context of supervised change detection in VHR images. The use of nonlinear SVM provided an efficient nonparametric solution to the nonlinearity of the multi-temporal signals and relaxed the data requirements of the model. Experiments confirmed the gain in performances when including contextual information for the three SVM-based change detection schemes considered (complete DMC, reduced DMC and DIA). The spatial smoothing provided by this information eases the class separation by the SVM model by bringing useful discriminative information and by reducing noise affecting the VHR multi-temporal images (due to acquisition conditions and residual misalignments). The spatial coherence of the change detection maps is thus greatly improved.

After the analysis of the outcomes, it remains difficult to draw strict conclusions about which set of features is appropriate for performing multi-temporal classification. As remarked by experimental results and discussion, composite textural and morphological sets have shown a constant, statistically significant and stable improvement in the κ coefficient for all the change detection schemes under all the tested conditions. However, it is worth mentioning that the relevance of the feature sets adopted here and their parameters (*e.g.* window size) are data dependent, and their choice must be addressed after careful visual inspection of the images.

As illustrated, inclusion of the spatial context information successfully filled the lack in spectral information for distinguishing the different transitions occurred in the images. However, prior or expert knowledge can be included in the process by choosing to combine features providing explicit information about specific ground covers. Moreover, to reduce the dimensionality and consequently apply a simpler classification routine (for instance the aforementioned LDA) and assuming an increased class separability by adding context information, further efforts must deal with dimensionality reduction techniques: Feature extraction (*e.g.*, principal component analysis, discriminant analysis feature extraction (Benediktsson et al., 2005)) and/or a feature selection (*e.g.*, ranking by independence criteria (Camps-Valls et al., 2010) or model-based (Tuia et al., 2010a)) can be utilized to extract or select features containing the most of the information of the data, easing the understanding of main patterns characterizing the change detection problem (*e.g.* geometrical structures, main texturing). This way, a classification of a lower dimensional set can be carried out without loosing significant accuracy.

Acknowledgement

This work has been partly supported by the Swiss National Science Foundation under the projects “KernelCD” 200021-126505, PBLAP2-127713 and PZ00P2-136827.

References

Baraldi, A., Parmiggiani, F., 1995. An investigation of the textural characteristics associated with gray level co-occurrence matrix statistical parameters. IEEE Trans. Geosci. Remote Sens. 33, 293–304.
Benediktsson, J., Palmason, J.A., Sveinsson, J.R., 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. IEEE Trans. Geosci. Remote Sens. 43, 480–490.
Boser, B., Guyon, I., Vapnik, V., 1992. A training algorithm for optimal margin classifiers. In: 5th ACM Workshop COLT, pp. 144–152.
Bovolo, F., 2009. A multilevel parcel-based approach to change detection in very high resolution multitemporal images. IEEE Geosci. Remote Sens. Lett. 6, 33–38.
Bovolo, F., Bruzzone, L., 2007. A theoretical framework for unsupervised change detection based on change vector analysis in polar domain. IEEE Trans. Geosci. Remote Sens. 45, 218–236.
Bovolo, F., Bruzzone, L., Marconcini, M., 2008. A novel approach to unsupervised change detection based on a semi-supervised SVM and a similarity measure. IEEE Trans. Geosci. Remote Sens. 46, 2070–2082.
Bovolo, F., Camps-Valls, G., Bruzzone, L., 2010. A support vector domain method for change detection in multitemporal images. Pattern Recognit. Lett. 31, 1148–1154.
Bruzzone, L., Carlin, L., 2006. A multilevel context-based system for classification of very high spatial resolution images. IEEE Trans. Geosci. Remote Sens. 44, 2587–2892.
Bruzzone, L., Fernández-Prieto, D., 2000. Automatic analysis of the difference image for unsupervised change detection. IEEE Trans. Geosci. Remote Sens. 38, 1171–1182.
Camps-Valls, G., Bruzzone, L., 2009. Kernel Methods for Remote Sensing Data Analysis. J. Wiley & Sons.
Camps-Valls, G., Mooij, J., Schölkopf, B., 2010. Remote sensing feature selection by kernel dependence measures. IEEE Geosci. Remote Sens. Lett. 7, 587–591.
Collobert, R., Bengio, S., Mariéthoz, J., 2002. Torch: A Modular Machine Learning Software Library. Technical Report IDIAP-RR 02-46. IDIAP.
Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., Lambin, E., 2004. Change detection methods in ecosystem monitoring: a review. Int. J. Remote Sens. 25, 1565–1596.
Dalla Mura, M., Benediktsson, J.A., Bovolo, F., Bruzzone, L., 2008. An unsupervised technique based on morphological filters for change detection in very high resolution images. IEEE Trans. Geosci. Remote Sens. 5, 433–437.
Fauvel, M., Benediktsson, J.A., Chanussot, J., Sveinsson, J.R., 2008. Spectral and spatial classification of hyperspectral data using svms and morphological profiles. IEEE Trans. Geosci. Remote Sens. 46, 3804–3814.
Foody, G.M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. Photogramm. Eng. Remote Sens. 70, 627–633.
Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. 3, 610–621.
Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theory 14, 55–63.
Inglada, J., 2007. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. ISPRS J. Photogramm. Remote Sens. 62, 236–248.
Kanevski, M., Pozdnoukhov, A., Timonin, V., 2010. Machine learning for spatial environmental data. EPFL Press, Lausanne.
Murray, H., Lucieer, A., Williams, R., 2010. Texture-based classification of sub-Antarctic vegetation communities on Heard Island. Int. J. Appl. Earth Obs. Geoinf. 12, 138–149.
Nemmour, H., Chibani, Y., 2006. Multiple support vector machines for land cover change detection: an application for mapping urban extensions. J. Photogramm. Remote Sens. 61, 125–133.
Pacifici, F., Chini, M., Emery, W., 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. Remote Sens. Environ. 113, 1276–1292.
Petrou, M., Sevilla, P., 2006. Dealing with Texture. Image Processing Series. John Wiley and Sons.
Schölkopf, B., Smola, A., 2002. Learning with Kernels. MIT Press, Cambridge (MA).
Shawe-Taylor, J., Cristianini, N., 2004. Kernel Methods for Pattern Analysis. Cambridge University Press.
Singh, A., 1989. Digital change detection techniques using remotely-sensed data. Int. J. Remote Sens. 10, 989–1003.
Soille, P., 2004. Morphological Image Analysis: Principles and Applications, vol. 391., 2nd edition Springer-Verlag, Berlin-Heidelberg.
Soille, P., Pesaresi, M., 2002. Advances in mathematical morphology applied to geoscience and remote sensing. IEEE Trans. Geosci. Remote Sens. 40, 2042–2055.
Tuia, D., Pacifici, F., Kanevski, M., Emery, W.J., 2009. Classification of very high spatial resolution imagery using mathematical morphology and support vector machines. IEEE Trans. Geosci. Remote Sens. 47, 3866–3879.
Tuia, D., Camps-Valls, G., Matasci, G., Kanevski, M., 2010a. Learning relevant image features with multiple kernel classification. IEEE Trans. Geosci. Remote Sens. 48, 3780–3791.
Tuia, D., Ratle, F., Pouzdnoukhov, A., Camps-Valls, G., 2010b. Multisource composite kernels for urban image classification. IEEE Geosci. Remote Sens. Lett. 7, 88–92.
Vapnik, V., 1998. Statistical Learning Theory. Wiley, New York.