Multiple support vector machines for land cover change detection: An application for mapping urban extensions

Hassiba Nemmour *, Youcef Chibani

Signal Processing Laboratory, Faculty of Electronic and Computer Science, University of Sciences and Technology HOUARI BOUMEDIENE, USTHB, P. B. 32, El-Alia 16111, Algiers, Algeria

Received 14 December 2005; received in revised form 13 September 2006; accepted 13 September 2006

Abstract

The reliability of support vector machines for classifying hyper-spectral images of remote sensing has been proven in various studies. In this paper, we investigate their applicability for land cover change detection. First, SVM-based change detection is presented and performed for mapping urban growth in the Algerian capital. Different performance indicators, as well as a comparison with artificial neural networks, are used to support our experimental analysis. In a second step, a combination framework is proposed to improve change detection accuracy. Two combination rules, namely, Fuzzy Integral and Attractor Dynamics, are implemented and evaluated with respect to individual SVMs. Recognition rates achieved by individual SVMs, compared to neural networks, confirm their efficiency for land cover change detection. Furthermore, the relevance of SVM combination is highlighted.

© 2006 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

Keywords: Change detection; Fuzzy Integral; Combination; Support vector machines; Attractor dynamics

1. Introduction

Based on the large number of available techniques, remote sensing images constitute the most feasible and efficient tool for land cover change detection. Usually, changes are detected by comparing multiple images of the same ground area acquired at different dates. Thereby, in recent years, advances made in change detection using various pattern discovery methods have shown its importance in many real world applications (Singh, 1989; Carlotto, 1997; Dai and Khorram, 1997). However, conventional change detection methods such as differencing and rationing do not exploit spectral characteristics of the different land cover changes since they employ only one spectral channel. Besides, the user is often interested in complete information about the change, including positions and spatial extent, as well as the precise nature of change, which is not available with conventional change detectors. Therefore, recently, attention has been focused on machine learning techniques such as Artificial Neural Networks (ANNs) since they use all spectral channels and provide complete information about land cover changes (Dai and Khorram, 1997; Nemmour and Chibani, 2006). In this framework, a straightforward and computationally attractive approach, called Support Vector Machines (SVMs), is presented. This method has been very successful in pattern recognition (Burges, 1998). The appeal of SVMs lies in their strong connection to the
underlying statistical learning theory, where they implement the structural risk minimization for solving two-class classification problems (Vapnik, 1995). This concept states that a better solution, in terms of generalization theory, can be found by minimizing an upper bound of the generalization error. This bound is formed by the sum of training errors and a term depending on the discrimination capacity of the learning machine given in its Vapnik–Chervonenkis (VC) dimension. Theoretically, the VC dimension is the largest number of data that can be separated in all possible ways using a given class of functions. Statistical learning theory shows that it is imperative to restrict the class of functions so that its capacity is suitable to the amount of available training data (Müller et al., 2001). Furthermore, SVMs include feature extraction in their own architecture. Users only need to define at most two parameters.

In remote sensing, SVMs are particularly used for classifying hyper-spectral images (Melgani and Bruzzone, 2002; Bruzzone and Melgani, 2003), as well as for modeling spectral mixtures (Brown et al., 2000). This paper introduces the use of SVMs for land cover change detection. Since we are interested in extracting changes, not in mapping the study area, our work is restricted to binary SVMs employed to highlight either all land cover transitions or some kinds of change. However, when using SVMs the user faces many possible choices of kernel functions commonly yielding different results. Hence, a second contribution of this paper is to propose a combination framework for SVMs in order to enhance change detection accuracy. Presently, two combination strategies based on Fuzzy Integral and Attractor Dynamics are used, respectively. The rest of the paper is organized as follows: Section 2 briefly reviews the basic theory of SVMs. Section 3 describes the SVM-based change detection system, as well as the combination strategies. Section 4 conducts several experiments to illustrate the relevance of SVMs. Finally, some conclusions are given in Section 5.

2. Overview of SVMs

SVMs were originally formulated to construct binary classifiers from a set of training examples such that: \((x_j, y_j) \in \mathbb{R}^N \times \{ \pm 1 \}, j=1,..., n\) (Burgos, 1998). Data are mapped into a dot product space via a kernel function such that: \(k(x, y) = \langle \varphi(x), \varphi(y) \rangle\). Then, the decision function is expressed in terms of kernel expansion as:

\[
f(x) = \sum_{j=1}^{SV} \alpha_j y_j k(x, x_j) + b
\]

\(\alpha_j\) are Lagrange multipliers while \(SV\) is the number of support vectors which are training data for which \(0 \leq \alpha_j \leq C\). \(C\) is a user-defined parameter that controls the tradeoff between the machine complexity and the number of nonseparable points (Huang and Liu, 2002). The bias \(b\) is a scalar computed by using any support vector. Then, the optimal hyper-plane corresponds to \(f(x) = 0\). Hence, test data are classified according to:

\[
x \in \begin{cases} 
\text{positive class if } f(x) > 0 \\
\text{negative class if } f(x) < 0 
\end{cases}
\]

Furthermore, all mathematical functions, which satisfy Mercer’s conditions, are eligible to be a SVM-kernel (Mercier and Lennon, 2003). In the literature, kernels are grouped into two categories, global and local kernels, respectively. Global kernels use samples that are far away from each other, but still have an influence on the kernel value. All kernels based on the dot product are global. A typical example of this category is the polynomial kernel given in Eq. (3).

\[
K(x, x_i) = (x \cdot x_i + 1)^p
\]

\(p\) is the polynomial order. In contrast, local kernels are based on a distance measure such as the Radial Basis Function (RBF) kernel, which is defined as:

\[
K(x, x_i) = \exp \left( -\frac{D(x, x_i)^2}{2\sigma^2} \right)
\]

\(\sigma\) is the standard deviation of the Gaussian function. Commonly, \(D(x, x_i)\) is the Euclidian distance \((D_E)\) between 2 pixels:

\[
D_E(x, x_i) = ||x - x_i||
\]

Recently, new distances, which take into consideration spectral signature of pixels, are used to fit multi-spectral point of view into the kernel matrix. Presently, Spectral Angle (SA) and Spectral Information Divergence (SID) are used (Mercier and Lennon, 2003). The SA measures the spectral difference between \(x\) and \(x_i\) according to:

\[
D_{SA}(x, x_i) = \arccos \left( \frac{\langle x, x_i \rangle}{||x|| ||x_i||} \right)
\]
On the other hand, the SID is deduced from the Kullback–Leibler divergence of spectral signatures, which is defined as:

\[ D_{\text{SID}}(x, x_i) = d(x_i||x) + d(x||x_i) \]  

(7)

\[ d(x||x_i) \] is given by:

\[ d(x||x_i) = \sum_{l=1}^{L} p_x(l) \log \left( \frac{p_x(l)}{p_x_i(l)} \right) \] such as \[ p_x(l) \]

\[ = x(l) \left| \sum_{k=1}^{L} x(k) \right| \]  

(8)

\( l \) designates a spectral channel while \( L \) is their number. For simplicity, RBF\(_E\), RBF\(_S\) and RBF\(_\text{SID}\) designate the RBF kernel using Euclidian SA and SID distances, respectively.

3. Change detection method

3.1. SVM-based change detection

Fig. 1 shows the change detection system using SVMs. Two multi-spectral images of the same study area acquired at different times are spatially aligned and concatenated to form input data. SVMs handle gray level of pixels taken from both images without any feature extraction since it is incorporated throughout their own architecture. Pixels, which constitute the change class, are considered positive (first class), while pixels of the no change class are considered negative (second class). Specifically, the change class should contain changes of interest. These changes are either several or a single kind of land cover transitions according to the aim of the study. On the other hand, the no change class contains unchanged areas, as well as unimportant changes. Once trained, for each pixel SVM produces a single output through its decision function. If this function is positive, the pixel is assigned to the change class, otherwise it is assigned to the no change class.

3.2. SVM combination

One of the difficulties when using SVMs is the choice of the best kernel function. In remote sensing, it seems preferable to use kernels that take the spectral signature into consideration. Nevertheless, they do not guarantee higher performance compared to conventional kernels (Mercier and Lennon, 2003). This paper overcomes this issue by using multiple SVMs in a combination framework, which is namely Multiple SVM System (MSS). Presently, the MSS is performed using two different combination rules, which are based on Sugeno’s Fuzzy Integral and Attractor Dynamics, respectively. In the remaining part of this section, we briefly review these approaches and show how they can be used for combining multiple SVMs.

3.2.1. Fuzzy Integral

The Fuzzy Integral combines, non linearly, objective evidences in the form of expert decisions with respect to subjective evaluation of their performance expressed by a fuzzy measure. For a finite set of elements \( Z = \{z_1, \ldots, z_n\} \), a set function \( g : 2^Z \rightarrow [0, 1] \) is called fuzzy measure.

![Fig. 1. Architecture of a two class-SVM change detector.](image-url)
if it verifies the following properties (Cho and Kim, 1995; Cho, 1995, 2002):

1. \( g(\phi) = 0 \)
2. \( g(Z) = 1 \)
3. \( g(z_i) \leq g(z_j) \) if \( z_j \subset z_i \)

In the present work, \( Z \) constitutes the set of change detectors (SVMs), while \( g(z_i) \) designate their performances. However, when combining change detectors one needs to know \( g \) of the obtained system, which is not possible using this basic form. In other words, this fuzzy measure does not follow the addition rule, that is if \( z_i, z_j \subset Z \) so that \( z_i \cap z_j = \phi \):

\[
g(z_i \cup z_j) \neq g(z_i) + g(z_j)
\]

Hence, Sugeno proposed a fuzzy measure depending on a \( \lambda \) parameter that expresses the degree of interaction between two elements.

\[
g(z_i \cup z_j) = g(z_i) + g(z_j) + \lambda g(z_i)g(z_j)
\]

Therefore, the Fuzzy Integral is calculated as follows. For each element \( z_i \) (SVM) to be combined, we associate a fuzzy measure \( g_k(z_i) \) to indicate its performance in the class \( k \). For a given pixel, let \( h_k(z_i) \) be the objective evidence of \( z_i \) in this class. The set of SVMs is then rearranged such that the following relation holds: \( h_k(z_1) \geq \ldots \geq h_k(z_n) \geq 0 \). Fuzzy measures of the new sequence of SVMs \( A_i = \{ z_1, \ldots, z_i \} \) are then constructed as:

\[
g_k(A_i) = g_k(z_1)
\]

\[
g_k(A_i) = g_k(A_{i-1} \cup z_i) = g_k(A_{i-1}) + g_k(z_i) + \lambda g_k(A_{i-1})g_k(z_i)
\]

\( \lambda \) is the unique root of a \( n-1 \) degree equation that is \( \lambda \in [-1, +\infty] \) and \( \lambda \neq 0 \). It is determined by solving the following equation:

\[
\lambda + 1 = \prod_{i=1}^{n} \left( 1 + \lambda g_k(z_i) \right)
\]

For each class \( k \in \{ +1, -1 \} \), the Fuzzy Integral (FI) is computed according to:

\[
\text{FI}_k = \max_{i=1}^{n} \left[ \min(h_k(z_i), g_k(A_i)) \right]
\]

### 3.2.2. Attractor dynamics

The basic idea behind this approach is that if several reliable SVMs agree on the assignment of a pixel, their outputs are averaged to construct a strong attractor (i.e. a strong change detector), while those the outputs of which strongly disagree are excluded and do not participate in the combination stage. This heuristic approach can be realized by mapping SVM outputs onto a dynamic system where each of them is considered as a local attractor. The formation of the set of decisions and their selections is dynamic and relies on models of human reasoning (Steinhage et al., 1999). Thus, \( h_k(z_i) \) are modeled by a Gaussian function such that (Steinhage et al., 1999; Bogdanov et al., 2003):

\[
\text{AD}_k = \sum_{i=1}^{n} \beta_k(z_i)(h_k(z_i) - m_k(z_i))e^{(h_k(z_i) - m_k(z_i))^2/\sigma_k^2}
\]

(15)

\( \beta_k(z_i) \) are called strengths (i.e. reliability indicators) of local attractors (SVMs). They are chosen according to a general knowledge about SVM reliabilities in both the change and no change classes. The standard deviation \( \sigma_k \) is the width of the basin of attraction. Theoretically, Eq. (15) describes dynamics of the state variable \( m_k(z_i) \), which is taken either as the linear average of \( h_k(z_i) \) or zero. SVMs compete with each other for possessing the “system’s state” and the strongest attractor wins.

#### 3.2.3. Fuzzy class membership model

On the basis of a change detector response obtained for a given pixel, objective evidences should express the membership degree of this pixel in both classes. Recall that there is no standard form or model for this function. The unique constraint is that it must be scaled in the interval \([0, 1]\) whereas SVMs produce a single output. Hence, we propose a fuzzy model, which assigns memberships for SVM output in both positive and negative classes. Let \( f_x(x) \) be the output of a SVM \( z_i \) obtained for a pixel \( x \) to be classified. The respective membership models \( h_k(z_i), k \in \{ +1, -1 \} \) associated to positive and negative classes are defined as follows.

If \( f_x(x) > 1 \) then \( \begin{cases} h_+(z_i) = 1 \\ h_-(z_i) = 0 \end{cases} \)

Else If \( f_x(x) < -1 \) then \( \begin{cases} h_+(z_i) = 0 \\ h_-(z_i) = 1 \end{cases} \)

Else \( \begin{cases} h_+(z_i) = 1 + f_x(x) \\ h_-(z_i) = 1 - f_x(x) \end{cases} \)

#### 3.2.4. Implementation of SVM combination

SVM combination using the Fuzzy Integral (FI) or Attractor Dynamics (AD) can be summarized throughout the following steps:

1. Train SVMs \( z_i, i = 1, \ldots, n \) on the same change detection task.
2. Calculate the reliability of each SVM ($g_k(z_i)$ or $\beta_k(z_i)$) in both classes $k = \{+1, -1\}$

3. For a pixel $x$ to be assigned, perform the following:
   - Calculate SVM outputs.
   - Generate $h_k(z_i), k = \{+1, -1\}$ using fuzzy class membership models.
   - For each class $k = \{+1, -1\}$ calculate $FI_k(x)$ for Fuzzy Integral and $AD_k(x)$ for Attractor Dynamics.

Then, $x$ is assigned to one of the two classes according to:

$$x = \begin{cases} +1 & \text{if } C_+ > C_- \\ -1 & \text{otherwise} \end{cases}$$

$C_+$ and $C_-$ designate the combined output associated to the respective positive and negative classes using either FI or AD rules. Recall that for both combination rules, $g_k(z_i)$ and $\beta_k(z_i)$ can be subjectively assigned by the expert or generated from training data (Cho and Kim, 1995; Steinhage et al., 1999).

4. Experimental results

The effectiveness of SVM-based change detection is investigated for detecting urban growth in Algerian capital. The experiment design includes both kernel selection and performance comparison with Artificial Neural Networks (ANNs). The most important criterion for performance evaluation is the overall accuracy, which is defined as the ratio between the number of correctly assigned test data and their total number. In addition, the Kappa coefficient, which is another measure of agreement between reference and classified data, is used. Presently, the Kappa coefficient is expressed by the Khat statistic calculated from both correct and incorrect assignments (Congalton, 2001). The Khat statistic is typically scaled between 0, which indicates complete disagreement between assignment and reference data, and 1, which indicates complete agreement. Furthermore, SVM data may be scaled in the range of $[-1, 1]$ or $[0,1]$ in order to avoid the saturation of kernel function, while for ANNs they should be normalized into $[0,1]$ (Paola and Schowengerdt, 1995). Therefore, to make an objective comparison between these change detectors, data were linearly scaled to the range $[0,1]$ by dividing radiometric values by 255.

4.1. Description of the study area

Two Landsat multi-spectral images covering Algerian capital, acquired in May 1985 (10:45 AM) and June 1996 (11:10 AM), are used. Atmospheric corrections were performed using the ATCOR2 software. Thereafter, the six spectral bands of both images, except thermal bands, were spatially aligned by a polynomial warp on ENVI software using 89 control points selected from linear and unchanged structures such as roads, coastal areas, as well as the old part of Algiers airport. The number of control points was controlled by the width of the overlapping area between images, which is about 1000 pixels. This region includes the center of Algerian capital surrounded by vacant areas among which areas of urban growth constitute the primary type of land cover change. The registration was achieved by a residual error less than 0.2 pixels to avoid the detection of spurious changes (Lambin and Strahler, 1994). Then, a section of $500 \times 500$ pixels was selected to perform the change detection task (Fig. 2). All new buildings and construction that appeared in the 1996 image constitute the positive class that represents urban growth. The rest of the image is assigned to the negative class.

Fig. 2. TM2 channel of two images of Algiers. (a) Image recorded in 1985. (b) Image recorded in 1996.
Once the nature and locations of training sites were decided for urban extensions, 200 pixels were manually selected from both classes to be used in the training stage. In addition, a set of 400 pixels from each class were selected to calculate reliability indicators (fuzzy measures and strengths) used for the SVM combination. For the test stage, 1000 and 1300 pixels were selected for urban extensions and the rest of the image, respectively.

### 4.2. Evaluation of SVM-based change detector

Experimental analysis of SVMs was conducted using different kernels with various parameter values. The $C$ value is fixed at 10, while the best parameters of polynomial and RBF kernels are tuned to $p = \{1,2,3,4,6\}$ and $\sigma = \{0.1,0.5,1,2,3\}$. For comparison, we use two hidden layer-neural networks trained by the standard back propagation rule. In order to allow the greatest possible accuracy, the step size, momentum and the number of iterations were fixed at 0.08, 0.99 and 400, respectively.

Tables 1 and 2 report the overall accuracy and Khat values obtained for SVMs and ANNs. As can be seen, all SVMs produce more precise results compared to the ANNs. More specifically, the polynomial kernel with $p=6$ gives the best performance with a gain more than 2% and 6.6% in the overall accuracy and Khat values, respectively, over the best ANN. Furthermore, two main remarks can be drawn for RBF kernels. On one hand, $\text{RBF}_{\text{E}}$ gives somewhat lower accuracies compared to $\text{RBF}_{\text{SA}}$ and $\text{RBF}_{\text{SID}}$. This outcome can be explained by the fact that SA and SID are more suitable to characterize the similarity between pixels than Euclidian distance. On the other hand, the $\text{RBF}_{\text{SID}}$ is significantly affected by the value of the standard deviation ($\sigma$) where its overall accuracy value decreases from 97.41% for $\sigma=0.1$ to 85.86% for $\sigma=3$. Except for this finding, SVM do not need a careful selection of the kernel function and its setup parameter to perform better than the ANNs.

**Table 1**

<table>
<thead>
<tr>
<th>Kernel parameter</th>
<th>$p=1$</th>
<th>$p=2$</th>
<th>$p=3$</th>
<th>$p=4$</th>
<th>$p=6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) SVMs: global kernel (polynomial)</td>
<td>94.13</td>
<td>97.58</td>
<td>97.06</td>
<td>97.06</td>
<td>97.58</td>
</tr>
<tr>
<td>(b) SVMs: local kernels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{RBF}_{\text{E}}$</td>
<td>96.89</td>
<td>95.50</td>
<td>95.86</td>
<td>95.86</td>
<td>94.82</td>
</tr>
<tr>
<td>$\text{RBF}_{\text{SA}}$</td>
<td>97.24</td>
<td>96.89</td>
<td>97.06</td>
<td>97.24</td>
<td>96.03</td>
</tr>
<tr>
<td>$\text{RBF}_{\text{SID}}$</td>
<td>97.41</td>
<td>96.72</td>
<td>95.51</td>
<td>91.55</td>
<td>85.86</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>(c) ANNs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93.67</td>
<td>94.24</td>
<td>94.92</td>
<td>87.71</td>
<td>86.82</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>Kernel parameter</th>
<th>$\sigma=0.1$</th>
<th>$\sigma=0.5$</th>
<th>$\sigma=1$</th>
<th>$\sigma=2$</th>
<th>$\sigma=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) SVMs: global kernel (polynomial)</td>
<td>88.26</td>
<td>94.81</td>
<td>94.12</td>
<td>94.12</td>
<td>95.15</td>
</tr>
<tr>
<td>(b) SVMs: local kernels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{RBF}_{\text{E}}$</td>
<td>93.77</td>
<td>91.02</td>
<td>91.71</td>
<td>91.72</td>
<td>89.64</td>
</tr>
<tr>
<td>$\text{RBF}_{\text{SA}}$</td>
<td>94.46</td>
<td>93.77</td>
<td>94.11</td>
<td>94.46</td>
<td>92.05</td>
</tr>
<tr>
<td>$\text{RBF}_{\text{SID}}$</td>
<td>94.81</td>
<td>93.43</td>
<td>91.02</td>
<td>83.15</td>
<td>71.91</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>(c) ANNs</td>
<td>87.39</td>
<td>87.73</td>
<td>88.53</td>
<td>75.39</td>
<td>74.46</td>
</tr>
</tbody>
</table>

Fig. 3 depicts change maps obtained using the best parametric selection for SVMs against the change map obtained for the most reliable ANN. In these maps, urban extensions in red are superimposed over the TM2 channel of the image acquired in 1985. As can be seen, the ANN produces a troublesome map that contains many regions of spurious changes (See circles in Fig. 3. (e)). This indicates that the change class is over-detected compared to the no change class. In other words, the ANN considers the regions surrounding urban extensions as changes, which means that it cannot precisely detect the boundaries between the two classes. This is in contrast to SVMs that produce much cleaner change maps in which urban extensions are well detected with fewer regions of false alarms.

### 4.3. Evaluation of SVM combination

To evaluate the effectiveness of the SVM combination, a set of experiments with varying numbers and types of SVMs was performed using the Fuzzy Integral (FI) and Attractor Dynamics (AD) rules. In fact, there are no criteria to define an optimal number of combined SVMs needed in order to outperform an individual SVM. Furthermore, there is no rule that would be followed to assign the knowledge about the reliability of SVMs (i.e. fuzzy measures and strengths). Presently, the reliability index of a SVM is taken as the overall accuracy value calculated over the validation set. For the AD rule, classes are modeled by a normal Gaussian with a zero mean ($m_k=0$) and a unity variance ($\sigma_k=1$). The SVM combination was carried out by using the best...
parametric selection for each kernel. Tables 3 and 4 report the overall accuracy and Khat values derived from the two combination rules against those obtained from individual SVMs. As can be seen, the FI and AD produce similar results that outperform individual SVMs. Using two SVMs, the accuracy improvement is about 1% in both the overall accuracy and Khat values. Besides, the best results are obtained when combining three SVMs since the improvement of Khat statistic reaches 2%. Nevertheless, higher number of

![Fig. 3. Change maps obtained for the different change detectors. (a) Polynomial kernel. (b) RBF_E kernel. (c) RBF_SA kernel. (d) RBF_SID kernel. (e) ANN.](image)

<table>
<thead>
<tr>
<th>Table 3</th>
<th>OAR (expressed in %) obtained from some combinations in comparison with individual SVMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combination</td>
<td>$P_{(p=6)}$</td>
</tr>
<tr>
<td>1</td>
<td>97.58</td>
</tr>
<tr>
<td>2</td>
<td>97.58</td>
</tr>
<tr>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>97.58</td>
</tr>
<tr>
<td>7</td>
<td>97.58</td>
</tr>
<tr>
<td>8</td>
<td>97.58</td>
</tr>
<tr>
<td>9</td>
<td>97.58</td>
</tr>
<tr>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>11</td>
<td>97.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Khat (expressed in %) obtained from some combinations in comparison with individual SVMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combination</td>
<td>$P_{(p=6)}$</td>
</tr>
<tr>
<td>1</td>
<td>95.15</td>
</tr>
<tr>
<td>2</td>
<td>95.15</td>
</tr>
<tr>
<td>3</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>95.15</td>
</tr>
<tr>
<td>7</td>
<td>95.15</td>
</tr>
<tr>
<td>8</td>
<td>95.15</td>
</tr>
<tr>
<td>9</td>
<td>95.15</td>
</tr>
<tr>
<td>10</td>
<td>–</td>
</tr>
<tr>
<td>11</td>
<td>95.15</td>
</tr>
</tbody>
</table>
combined SVMs does not ensure more improvement since we obtain the same overall accuracy and Khat values for combinations of three and four SVMs.

5. Discussion and conclusion

Conventional change detection schemes such as differencing and principle component analysis suffer from two main problems. First, they do not provide information about the nature of change. Second, the decision about the change is commonly constructed by using a threshold for which the determination of the optimal value is a challenging task. Actually, this problem is mitigated by using classification-based schemes in order to allow an automatic and categorical extraction of the changes (Dai and Khorraram, 1997). Among these methods, ANNs were extensively and successfully used (Dai and Khorraram, 1997; Chibani and Nemmour, 2003; Nemmour and Chibani, 2006). This paper introduced a new change detection scheme based on SVMs that are theoretically able to achieve higher accuracy than ANNs. This is due to their optimization criterion, which is a maximization problem based on structural risk minimization. Experiments were conducted for mapping urban extensions in Algerian capital. A SVM-based change detector with different kernel functions was compared to ANN using both statistical and visual evaluations. Experimental results have shown that SVMs yield better performance regarding accuracy and generalization. Moreover, there are only two user-defined parameters (C and kernel parameter), which do not have an important effect on the SVM response. To the contrary, many experiments were conducted for ANN to find the appropriate values of the step size, momentum, as well as the weight initialization. Furthermore, since SVM kernels produced somewhat different results, this study investigated the usefulness of SVM combination. To reach this objective, two combination rules, namely, Fuzzy Integral and Attractor Dynamics were used since they take into consideration the reliability of individual SVMs, unlike conventional combination schemes such as voting rules. The results obtained showed the relevance of change detector combination. Although different, the two rules produced similar results, while the best accuracy improvement was obtained by combining three SVMs.

Acknowledgments

Authors would like to thank Professors A. Belhadj Aissa and Y. Smara of Remote Sensing Laboratory, Faculty of Electronic and Computer Sciences, USTHB, for providing the preprocessed TM images utilized in this work.

References


