Available online at www.sciencedirect.com

SCIENCE DIRECT®

Pattern Recognition Letters 25 (2004) 1491–1500

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Detection of land-cover transitions by combining multidate classifiers

L. Bruzzone [a,*], R. Cossu [a], G. Vernazza [b]

[a] *Department of Information and Communication Technology, University of Trento, Via Sommarive 14, I-38050 Povo, Trento, Italy*

[b] *Department of Biophysical and Electronic Engineering, University of Genoa, Via Opera Pia 11 A, I-16145 Genova, Italy*

Abstract

This paper addresses the problem of detecting land-cover transitions by analysing multitemporal remote-sensing images. In order to develop an effective system for the detection of land-cover transitions, an ensemble of non-parametric multitemporal classifiers is defined and integrated in the context of a multiple classifier system (MCS). Each multitemporal classifier is developed in the framework of the compound classification (CC) decision rule. To develop as uncorrelated as possible classification procedures, the estimates of statistical parameters of classifiers are carried out according to different approaches (i.e., multilayer perceptron neural networks, radial basis functions neural networks, and *k*-nearest neighbour technique). The outputs provided by different classifiers are combined according to three standard strategies extended to the compound classification case (i.e., Majority voting, Bayesian average, and Bayesian weighted average). Experiments, carried out on a multitemporal remote-sensing data set, confirm the effectiveness of the proposed system.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Detection of land-cover transitions; Change detection; Multitemporal classification; Multiple classifier systems; Multilayer perceptron neural networks; Radial basis function neural networks; *k-nn* technique; Expectation–maximization algorithm; Remote sensing images

1. Introduction

One of the most interesting applications of remote-sensing concerns the analysis of multitemporal images for detecting land-cover changes. This process involves the comparison of two co-registered images acquired in the same geographical area at two different times. Two main approaches to the change-detection problem can be adopted: the supervised approach and the unsupervised one (Richards, 1993; Bruzzone and Serpico, 1997). The former is based on supervised classification methods, which require the availability of a multitemporal ground-truth. The latter performs change detection by making a direct comparison of the two multispectral images

* Corresponding author. Tel.: +39-461882623; fax: +39-461882671.

E-mail address: lorenzo.bruzzone@ing.unitn.it (L. Bruzzone).

0167-8655/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2004.06.002

considered, without relying on any additional information. Despite the supervised approach requiring the availability of ground truth information for both considered dates, it exhibits some important advantages over the unsupervised one: (i) capability of explicitly recognising the kind of land-cover transitions which occurred in the investigated area; (ii) robustness to the different atmospheric and light conditions at the two acquisition times; (iii) ability to process multisensor/multisource images (Bruzzone and Serpico, 1997). For these reasons, in the rest of the paper we will focus our attention on the supervised approach.

In the supervised approach, the land-cover transitions are usually detected by comparing the thematic maps obtained by classifying independently the two considered images. Consequently, the accuracy yielded strongly depends on the errors present in the classification maps. For this reason, in the context of the detection of land-cover transitions, it is of great importance to develop effective classification approaches capable of achieving classification accuracies as high as possible. In spite of the importance of this issue, in the literature only few works addressed the abovementioned problem. A possible approach to increase the accuracy in classification of multitemporal images is the direct multidate classification (Bruzzone and Serpico, 1997). It allows one to obtain the land-cover transition map by jointly applying the classifier (directly trained for recognising land-cover transitions) to the two images acquired at different dates. This approach has the important advantage of fully exploiting the temporal correlation existing between the multitemporal images, thus increasing the accuracy of the classification process. However, it suffers from the serious drawback of requiring complex multitemporal ground-truth information. In fact, the training of the classifier requires the availability of: (i) ground-truth information for the same pixels (the same areas on the ground) at the two acquisition dates; (ii) significant statistics for all the possible land-cover transitions. In practical applications, it is often not possible to satisfy these constraints. In (Bruzzone et al., 1999), the authors proposed a method based on the compound classification of multitemporal data, which is able to detect land-cover transitions by exploiting the temporal correlation between images and removing the aforementioned critical constraints on the necessary ground truth. Such a method appeared to be effective on several remote-sensing data sets. However, in order to address ever more complex applications involving the detection of land-cover transitions, it is necessary to study approaches capable of further increasing the multitemporal classification accuracy.

Recently, in the pattern-recognition literature several authors have proposed the use of multiple classifier systems (MCSs) to increase the accuracy and reliability of single-date classifiers (Kittler et al., 1998). In particular, it can be proven that combining classifiers that are accurate and different (i.e., classifiers that incur uncorrelated errors) results in a more accurate and robust classification system. For these reasons, the MCS seems to be a promising methodology also in the field of multidate classification and consequently can represent an effective approach to the detection of land-cover transitions.

In this paper, we propose to address the detection of land-cover transitions by means of an MCS composed of an ensemble of multitemporal classifiers. In this context, the main problems to be solved are related to the definition of an effective ensemble of multitemporal classifiers and to the use of combination strategies in the context of multitemporal techniques. Concerning the ensemble of multitemporal classifiers, it should be composed of classifiers that: (i) are characterised by high classification accuracy; (ii) do not require a too complex ground-truth information; (iii) incur approximately "uncorrelated" errors. In order to meet the aforementioned requirements, we propose to develop multidate classifiers in the context of the Bayesian rule for compound classification (CC) (Bruzzone et al., 1999; Swain, 1978). To define classifiers that incur approximately uncorrelated errors, different non-parametric algorithms are considered for estimating the statistical terms required by the Bayesian CC decision rule. The resulting classifiers are then integrated by using standard combination procedures. As compared to Bruzzone et al. (1999), the main novelties of the

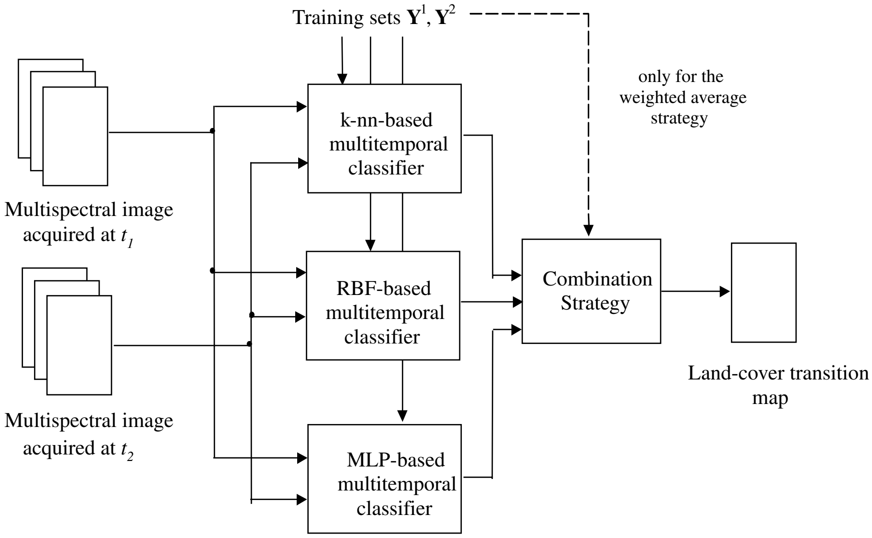

Fig. 1. Architecture of the proposed system.

proposed approach are: (a) the compound classification method (based on multilayer perceptron neural networks) presented in (Bruzzone et al., 1999) is extended to other non-parametric multitemporal classifiers; (b) the change detection problem is formulated within the framework of an MCS approach, thus increasing the reliability of the final map of land-cover transitions. Experiments, carried out on a multitemporal remote-sensing data set, confirm the effectiveness of the proposed MCS approach.

2. Definitions and problem formulation

Let us consider two co-registered remote-sensing images I_1 and I_2 acquired on the same geographical area at the times t_1 and t_2, respectively. Let us address the problem of detecting land-cover transitions which occurred in the considered geographical area by analysing I_1 and I_2. We characterise the pair of temporally correlated pixels (x_j^1, x_j^2) (made up of a pixel x_j^1 of image I_1 and of the spatially corresponding pixel x_j^2 of image I_2) by a pair of feature vectors (X_j^1, X_j^2). Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{M_1}\}$ be the set of possible land-cover classes at time t_1, and let $N = \{\upsilon_1, \upsilon_2, \ldots, \upsilon_{M_2}\}$ be the set of possible land-cover classes at time t_2. The two sets Ω and N may differ in both the number and the typology of land-cover classes. Let Y^i denote the set of pixels x_j^i for which ground truth information is available.

We formulate the problem of detecting land-cover transitions in the framework of an MCS approach. In particular, the MCS is composed of an ensemble of multitemporal classifiers, whose outputs are opportunely combined in order to obtain the final decision (i.e., the "best" pair of classes to be assigned to each couple of pixels (x_j^1, x_j^2), and hence the kind of land-cover transition occurred). The general architecture of the system is shown in Fig. 1.

3. Design of the ensemble of multitemporal classifiers

A major problem to be solved in the proposed system is the choice of an appropriate ensemble of multitemporal classifiers to be integrated in the multiple classifier system. As stated in Section 1, we propose to develop each classifier in the framework of the Bayes rule for the compound classification (Bruzzone et al., 1999, Swain, 1978). Accordingly, the following decision rule should be used:

$$x_j^1 \in \omega_s \text{ and } x_j^2 \in \upsilon_t \text{ such that } P(\omega_s, \upsilon_t | X_j^1, X_j^2) = \max_{\substack{\omega_m \in \Omega \\ \upsilon_n \in N}} \{P(\omega_m, \upsilon_n | X_j^1, X_j^2)\} \quad (1)$$

Under the reasonable simplifying assumption of class-conditional independence, the decision rule (1) is equivalent to (Bruzzone et al., 1999):

$x_j^1 \in \omega_s$ and $x_j^2 \in \upsilon_t$ such that

$$\frac{P(\omega_s|X_j^1)P(\upsilon_t|X_j^2)P(\omega_s,\upsilon_t)}{P(\omega_s)P(\upsilon_t)} = \max_{\substack{\omega_m \in \Omega \\ \upsilon_n \in N}} \left\{ \frac{P(\omega_m|X_j^1)P(\upsilon_n|X_j^2)P(\omega_m,\upsilon_n)}{P(\omega_m)P(\upsilon_n)} \right\} \quad (2)$$

where $P(\omega_m|X_j^1)$ is the conditional posterior probability of class ω_m, given pixel x_j^1, $P(\upsilon_n|X_j^2)$ is the conditional posterior probability of class υ_n, given pixel x_j^2, $P(\omega_m)$ and $P(\upsilon_n)$ are the prior probabilities of classes ω_m and υ_n, respectively, and $P(\omega_m, \upsilon_n)$ is the prior joint probability of the pair of classes (ω_m, υ_n).

In order to carry out the multitemporal classification, it is necessary to estimate the terms involved in (2) (in particular, in our case, we have to define different multitemporal classifiers to obtain an effective ensemble of algorithms). The same procedure is adopted for estimating the prior probabilities of classes $P(\omega_m)$ and $P(\upsilon_n)$ for all multitemporal classifiers considered. The conditional posterior probabilities $P(\omega_m|X_j^1)$ and $P(\upsilon_n|X_j^2)$ are estimated in a different way for each specific multitemporal approach. The strategy used to compute the prior joint probabilities $P(\omega_m, \upsilon_n)$ is the same for all the multitemporal classifiers composing the ensemble (however different estimates can be obtained for the considered classifiers since they are computed on the basis of the conditional posterior probabilities $P(\omega_m|X_j^1)$ and $P(\upsilon_n|X_j^2)$). The techniques used for estimating the aforementioned parameters are described in the following.

3.1. Estimation of the prior probabilities of classes

As it is usually done in the classification of remote-sensing images, the prior probabilities of classes $P(\omega_m)$ and $P(\upsilon_n)$ are estimated from the training sets by computing the relative frequencies of the patterns belonging to the different classes. For example, concerning the image t_1, we can write:

$$P(\omega_m) = \frac{\text{no. of pixels of the training set } t_1 \text{ that belong to } \omega_m}{\text{total no. of pixels in the training set } t_1} \quad (3)$$

It is worth noting that the accuracies of these estimates depend on the strategy adopted for the collection of the training samples.

3.2. Estimation of the conditional posterior probabilities of classes

In order to develop an effective system being able to classify multitemporal/multisensor data, we propose to use non-parametric algorithms to estimate the single-date conditional posterior probabilities $P(\omega_m|X_j^1)$ and $P(\upsilon_n|X_j^2)$. In particular, for defining different multitemporal classifiers, we propose to use three different estimation algorithms, i.e., multilayer perceptron (MLP) neural networks, radial basis function (RBF) neural networks, and the k-nearest neighbours (k-*nn*) technique.

3.2.1. Estimation procedure based on MLP neural networks

To obtain an MLP-based multitemporal classifier, the posterior probabilities of classes involved in (2) are estimated by means of MLP neural networks. The architecture of such kind of neural networks is composed of three or more layers (an input layer, an output layer, and one or more hidden layers). The input neurons are in number equal to the input features. The output layer is composed of as many neurons as the classes to be recognized. Each neuron is characterized by a sigmoid activation function. All the connections between neurons are associated with weights. The MLP neural networks, if properly trained by the error back-propagation (EBP) algorithm, provide estimates of the conditional posterior probabilities of classes $P^{\text{MLP}}(\omega_m|X_j^1)$ and $P^{\text{MLP}}(\upsilon_n|X_j^2)$ optimised according to a predefined criterion (Bishop, 1996).

In our case, we adopt the minimum square error (MSE) criterion (Serpico et al., 1996; Bruzzone and Serpico, 1997).

3.2.2. Estimation procedure based on RBF neural networks

RBF neural networks are composed of three layers (an input, a hidden, and an output layer). Input neurons (as many as input features) just propagate input features to the next layer. Each neuron in the hidden layer is associated with a kernel function φ_i (usually a Gaussian function) characterised by a centre π_i and a width σ_i. The output layer is composed of as many neurons as the classes to be recognised. Each output neuron computes a simple weighted average of the responses of the hidden neurons to a given pattern described by the input feature vector (Bishop, 1996; Bruzzone and Fernàndez Prieto, 1999). If properly trained (see Moody and Darken, 1989; Miller and Uyar, 1998), the RBF neural networks can be used for estimating the conditional posterior probabilities of classes $P^{\mathrm{RBF}}(\omega_m|X_j^1)$ and $P^{\mathrm{RBF}}(\upsilon_n|X_j^2)$. Consequently, an RBF-based multitemporal classifier can be obtained.

3.2.3. Estimation procedure based on the *k-nn* technique

The *k-nn* technique can be used for estimating in a supervised way the posterior probabilities of classes (Fukunaga, 1990). To simplify the notation, let us consider the case in which we are analysing the first image I_1 (the generalization to image I_2 is straightforward). To estimate the conditional probabilities $P^{k\text{-}nn}(\omega_m|X_j^1)$, the k prototypes nearest to pixel x_j^i are identified among all the patterns of the training set Y^i. The available ground-truth information gives us information about the land-cover class that each of these k patterns represents. It can be proven that a local and simple estimate of the posterior probabilities $P^{k\text{-}nn}(\omega_m|X_j^1)$ can be derived as follows:

$$P^{k\text{-}nn}(\omega_m|X_j^1) \cong \frac{k_m}{k} \tag{4}$$

where k_m is the number of prototypes belonging to class ω_m among the k-nearest neighbouring samples.

3.3. Estimation of the joint prior probabilities of classes

Once the estimates of the posterior probabilities of classes are obtained, the expectation–maximisation (EM) algorithm (Dempster et al., 1977; Moon, 1996) can be applied to estimate the joint prior probabilities of classes. In particular, on the basis of the estimates of the posterior probabilities provided by the three proposed techniques (i.e., MLP neural networks, RBF neural networks, *k-nn* technique), it is possible to derive different estimates of the joint prior probabilities of classes. It can be proven that the recursive equation to be used to estimate $P(\omega_m, \upsilon_n)$ is the following (Bruzzone et al., 1999):

$$\begin{aligned} P_{t+1}(\omega_m, \upsilon_n) &= \frac{1}{SP(\omega_m)P(\upsilon_n)} \\ &\quad \times \sum_{j=1}^{S} \frac{P_t(\omega_m, \upsilon_n)P(\omega_m|x_j^1)P(\upsilon_n|x_j^2)}{\sum_{\omega_p \in \Omega} \sum_{\upsilon_q \in N} \frac{P_t(\omega_p, \upsilon_q)}{P(\omega_p)P(\upsilon_q)} P(\omega_p|x_j^1)P(\upsilon_q|x_j^2)} \end{aligned} \tag{5}$$

where t and $t+1$ are the current and the next iterations, $P_t(\omega_m, \upsilon_n)$ is the joint probability of classes estimated at the tth iteration and S is the total number of pixels to be classified.

Eq. (5) can be used with the posterior probabilities estimated according to the different proposed non-parametric algorithms. We refer the readers to (Bruzzone et al., 1999) for greater details on the procedure for the estimation of the joint probabilities of classes $P(\omega_m, \upsilon_n)$.

4. Combination strategies for multitemporal classifiers

Let us consider a set of C multitemporal classifiers capable of estimating the joint posterior probabilities $P(\omega_m, \upsilon_n|X_j^1, X_j^2)$, $\omega_m \in \Omega$, $\upsilon_n \in N$, for each pair of pixels (x_j^1, x_j^2) (see Section 3 for the design of each single multitemporal classifier with the simplifying assumption of conditional independence in the time domain). We propose to combine the different classifiers by applying three standard combination strategies modified for taking into account the multitemporal nature of our classifiers: the Majority voting (Lam and Suen, 1997), the combination by Bayesian average (Dietterich, 2000), and the Weighted average strategies. The first two strategies are unsupervised (they do not require the availability of a training set for defining the combination rule), whereas the third strategy is supervised (the combination strategy exploits the available training set).

4.1. Multitemporal majority-voting strategy

The Majority voting procedure faces the combination problem by considering the results of each single multitemporal classifier in terms of the pair of class labels assigned to the patterns. A given input pattern receives C classification labels from the MCS: each label corresponds to one of the $M_1 \times M_2$ pairs of classes (ω_m, υ_n) considered. The combination method is based on the interpretation of the classification label resulting from each classifier as a "vote" to one of the possible $M_1 \times M_2$ pairs of land-cover classes. The class pair that receives the largest number of votes is taken as the kind of land-cover transition related to the input pattern. When more than one class pair receives the maximum number of votes, the pair with the maximum posterior probability is selected.

4.2. Multitemporal Bayesian average strategy

When the combination by Bayesian average strategy is used, the final joint posterior probabilities of the MCS are obtained by averaging the outputs provided by the different multitemporal classifiers. In particular, given the estimates of the joint posterior class probabilities $P^i(\omega_m, \upsilon_n | X_j^1, X_j^2)$ ($\omega_m \in \Omega$, $\upsilon_n \in N$) provided by the ith classifier of the ensemble for the pair of pixels (x_j^1, x_j^2), the final values of the posterior class probabilities $P^{\text{ave}}(\omega_m, \upsilon_n | X_j^1, X_j^2)$ are computed as follows:

$$P^{\text{ave}}(\omega_m, \upsilon_n | X_j^1, X_j^2) = \frac{1}{C} \sum_{i=1}^{C} P^i(\omega_m, \upsilon_n | X_j^1, X_j^2) \qquad (6)$$

According to the Bayesian approach, each pair of pixels is assigned to the pair of classes for which the average posterior probability is maximised.

4.3. Multitemporal Bayesian Weighted average strategy

The rationale of this strategy consists in using the information present in the training set for evaluating and taking into account the effectiveness of each classifier on the different classes. Accordingly, a weighted average of the posterior probabilities of classes is considered instead of the simple average computed in the previous strategy. The adopted combination rule is the following:

$$P^{w\text{-ave}}(\omega_m, \upsilon_n | X_j^1, X_j^2) = \frac{\sum_{i=1}^{C} \alpha_{m,n}^i P^i(\omega_m, \upsilon_n | X_j^1, X_j^2)}{\sum_{m=1}^{M_1} \sum_{n=1}^{M_2} \sum_{i=1}^{C} \alpha_{m,n}^i P^i(\omega_m, \upsilon_n | X_j^1, X_j^2)} \qquad (7)$$

where $\alpha_{m,n}^i$ is the weight associated with the estimation of the posterior class probability related to the pair of classes (ω_m, υ_n) provided by the ith multitemporal classifier. As for the combination by Bayesian average strategy, the final decision is taken according to the maximisation of the estimated weighted posterior probability. The weights $\alpha_{m,n}^i$ are computed by minimising a cost function (i.e., the minimum square error) on the training samples with a procedure based on a pseudo-inverse matrix (it is worth noting that $\alpha_{m,n}^i$ must be non-negative so that $P^{w\text{-ave}}(\omega_m, \upsilon_n | X_j^1, X_j^2)$ represents a probability).

5. Experimental results

In order to assess the effectiveness of the proposed approach, different experiments were carried out on a data set made up of two multisensor images acquired by the Thematic Mapper (TM) multispectral sensor of the Landsat 5 satellite and by the Synthetic Aperture Radar (SAR) of the ERS-2 satellite. The selected test site was a section of a scene related to the delta of the Po river, Italy. The two images used in the experiments were acquired in April 1994 (t_1) and May 1994 (t_2). Four land-cover classes (i.e., urban area, bare soil, wet rice field, and woodland) were considered for the April image, whereas five land-cover classes (i.e., urban area, dry rice field, wet rice field, cereals, woodland) were considered for the May image.

Table 1
Training and test sets utilised for the experiments: (a) April 1994; (b) May 1994

	Number of pixels	
	Training set	Test set
(a) Land-cover classes (April 1994)		
Bare soil	9565	4319
Wet rice-fields	5867	2955
Wood	16,008	9269
Urban area	801	334
Total	32,241	16,877
(b) Land-cover classes (May 1994)		
Dry rice-fields	1919	866
Wet rice-fields	10,812	5062
Wood	16,008	9269
Cereals	2701	1346
Urban area	801	334
Total	32,241	16,877

The available ground truth was used to derive a training set and a test set for each image (see Table 1). (It is worth noting that the portion of the images and the distributions of classes considered in this paper are different from the ones considered in (Bruzzone et al., 1999).)

The three proposed multitemporal classifiers (i.e., *k-nn*, RBF, MLP based classifiers) were trained using all the spectral channels of the TM (but the band six) and the intensity of the SAR image. For each classifier, different architectures/ parameters (i.e., the value of the parameter k in the *k-nn* technique, the number of hidden nodes in the RBF neural networks, the number of layers and hidden nodes in the MLP neural networks) were investigated.

First of all, two *k-nn* classifiers were trained on the April and May images, respectively, in order to estimate the posterior probabilities $P^{k\text{-}nn}(\omega_m|X_j^1)$ and $P^{k\text{-}nn}(\upsilon_n|X_j^2)$. Trials with values of the parameter k of the *k-nn* classifiers ranging from 3 to 100 were carried out (12 different values were considered). Then the EM algorithm was applied to estimate the joint prior probabilities of classes $P^{k\text{-}nn}(\omega_m, \upsilon_n)$.

A similar procedure was applied to derive the RBF and MLP based multitemporal classifiers. In particular, the estimates of $P^{\mathrm{RBF}}(\omega_m|X_j^1)$, $P^{\mathrm{RBF}}(\upsilon_n|X_j^2)$, $P^{\mathrm{RBF}}(\omega_m, \upsilon_n)$, and $P^{\mathrm{MLP}}(\omega_m|X_j^1)$, $P^{\mathrm{MLP}}(\upsilon_n|X_j^2)$, $P^{\mathrm{MLP}}(\omega_m, \upsilon_n)$ were carried out. Concerning the RBF neural networks, 24 different architectures were analysed by varying the number of hidden neurons from 50 to 200. Concerning the MLP-based classifier, seven different architectures were investigated. Such architectures were obtained by varying (i) the number of hidden nodes (20 and 40 neurons); (ii) the initial conditions; and (iii) the numbers of iterations used for the training phase.

Table 2 gives the maximum, the minimum, the average and the standard deviation of the accuracies in detecting the land-cover transitions exhibited by the three aforementioned multitemporal classifiers in our experiments. As one can see, the accuracies yielded by all the proposed classifiers were very high (i.e., the *k-nn*-based, the RBF-based and the MLP-based classifiers exhibited an average accuracy equal to 95.40%, 94.69%, and 94.17%, respectively).

For the sake of comparison, Table 3 shows the results obtained by applying a standard post-classification comparison (PCC) technique to the corresponding single-date classifiers (for each pair of single-date classifiers, the best-performing architecture was considered). By comparing Table 2 with Table 3, it can be observed that, as expected, the use of multitemporal classifiers allows one to significantly increase the change-detection

Table 2
Best, worst, and average accuracies in detecting the land-cover transitions exhibited by the three considered multitemporal classifiers. The standard deviation of the accuracy in the different trials is also given

	k-nn based classifier	RBF-based classifier	MLP-based classifier
Best accuracy (%)	95.58	95.44	95.93
Worst accuracy (%)	94.62	90.55	85.83
Average accuracy (%)	95.40	94.69	94.17
Standard deviation	0.32	0.91	3.68

Table 3
Accuracies in detecting the land-cover transitions exhibited by a standard PCC technique applied to the best-performing single-date classifiers considered

	k-nn based classifier	RBF-based classifier	MLP-based classifier
Accuracy in detecting land-cover transitions (%)	92.86	92.58	92.78

Table 4
Best and average accuracies in detecting the land-cover transitions exhibited by the proposed MCS approach with the three considered combination strategies. The standard deviation of the accuracy in the different trials is also given

	Majority voting	Bayesian average	Bayesian weighted average
Best accuracy (%)	96.04	96.53	96.69
Average accuracy (%)	95.86	96.22	96.37
Standard deviation	0.22	0.15	0.28

Table 5
Increase in the accuracies in detecting the land-cover transitions exhibited by the proposed combination strategies in the different trials versus (Panel a) the worst-performing and (Panel b) the best-performing classifiers composing the ensemble

Increase (%)	Majority voting	Bayesian average	Bayesian weighted average
(Panel a)			
Maximum	9.67	10.17	10.04
Minimum	1.13	1.30	1.57
Average	2.53	2.89	3.04
(Panel b)			
Maximum	0.55	0.91	1.16
Minimum	−0.15	0.35	0.29
Average	0.24	0.60	0.75

accuracy with respect to the accuracy achieved by the standard PCC technique. This is obtained by exploiting the temporal correlation between class-labels in the two considered images.

At this point, the three multitemporal classifiers (i.e., the *k-nn*-based, the RBF-based, and the MLP-based multitemporal classifiers) were integrated in the MCS by means of the considered combination strategies (i.e., Majority-voting, Bayesian average, and Bayesian weighted average). In order to statistically analyse the effectiveness and the stability of the three combination strategies, different ensembles of multitemporal classifiers were considered. In particular, each ensemble was defined by varying the architectures/parameters of the composing multitemporal classifiers (it is worth noting that the considered ensembles also included the worst-performing architectures of each multitemporal classifier). The obtained results are reported in Table 4. As one can see, the average change-detection accuracies obtained were very high for all the combination strategies proposed (i.e., 95.86%, 96.22%, and 96.37% for the Majority-voting, Bayesian average, and Bayesian weighted average strategy, respectively).

In order to analyse the behaviour of the MCS as compared with the one of the single multitemporal classifiers composing the ensemble, in Table 5 the increase in the change-detection accuracy led by the three proposed combination strategies with respect to both the worst-performing and best-performing multitemporal classifiers composing the ensembles are given. In all the trials carried out, all the combination strategies significantly increased the accuracy as compared with the worst-performing multitemporal classifier composing the ensemble (an average increase of 2.53%, 2.89%, and 3.04%, was observed with the Majority-voting, the Bayesian average, and Bayesian weighted average strategies, respectively). This is a very important aspect (especially when classifiers whose performances strongly depend on the architecture definition are considered) since it confirms that the MCS approach increases the robustness in detecting land-cover transitions, mitigating the effects introduced by not-optimally designed classifiers. Furthermore, in almost all the trials carried out, the considered combination strategies increased the accuracy with respect to the best-performing multitemporal classifier composing the ensemble (an average increase of 0.24%, 0.60% and 0.75% was achieved with the Majority-voting, the Bayesian average, and Bayesian weighted average strategies, respectively).

It is worth noting that the robustness of the proposed MCS is also confirmed by the analysis of the standard deviations computed on the change-detection accuracy, which reveals that the overall performances are not significantly dependent on the particular architecture of the classifiers composing the ensemble. In particular, the standard deviations of the accuracy obtained by the *k-nn*-based, the RBF-based, and the MLP-based multitemporal classifiers were equal to 0.32%, 0.91%, and 3.68% (see Table 2) whereas the standard deviations of the accuracies exhibited by the MCS with the Majority voting, Bayesian average, and Bayesian weighted average strategies were equal to 0.22%, 0.15% and 0.28%, respectively (see Table 4).

6. Conclusions

In this paper, a novel approach to the supervised detection of land-cover transitions has been proposed. Such an approach is based on a multiple classifier architecture composed of different multitemporal classifiers developed in the framework of the compound classification decision rule. In particular, MLP neural networks, RBF neural networks and the *k-nn* technique have been used to define three different multitemporal classifiers to be included in the MCS.

The results obtained on a real multitemporal remote sensing data set confirmed the effectiveness of the proposed approach. In particular, different ensembles of classifiers were considered in the MCS by varying the architectures/parameters of the single multitemporal classifiers. In all the trials carried out, the three combination strategies proposed resulted in a significant increase of the accuracy with respect to the worst-performing classifier composing the ensemble. As already observed, this is a very important aspect since it points out that the MCS approach increases the robustness of the process of detection of land-cover transitions, mitigating the effects introduced by non-optimally designed classifiers. The analysis of the standard deviations of the change-detection accuracies also reveals that the overall performance is not significantly dependent on the particular architecture of the classifiers composing the ensemble. Thus, the results obtained confirm the MCS approach as a well-founded way of increasing the robustness of the system for detecting land-cover transitions.

As a final remark, it is worth noting that in remote-sensing applications the training and the test sets used for the learning and the selection of the classifier architectures, respectively, often do not represent completely the variability of classes in the investigated scene; consequently, they do not allow one a reliable evaluation of the generalization capabilities of each classifier architecture. For this reason, it is important to define robust systems based on an ensemble of classifiers that can intrinsically overcome the possible unexpectedly poor generalization performance of a single classification algorithm.

References

Bishop, C., 1996. Neural Network for Pattern Recognition. Oxford University press.

Bruzzone, L., Fernàndez Prieto, D., 1999. A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images. IEEE Trans. Geosci. Remote Sensing 37 (2), 1179–1184.

Bruzzone, L., Serpico, S.B., 1997. An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images. IEEE Trans. Geosci. Remote Sensing 35 (4), 858–867.

Bruzzone, L., Férnandez Prieto, D., Serpico, S.B., 1999. A neural statistical approach to multitemporal and multisource remote-sensing image classification. IEEE Trans. Geosci. Remote Sensing 37 (3), 1350–1359.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. 39 (1), 1–38.

Dietterich, T.G., 2000. Ensemble methods in machine learning. In: Proceedings of the First International Workshop on Multiple Classifier Systems, Cagliari, Italy, pp. 1–15.

Fukunaga, K., 1990. Introduction Statistical Pattern Recognition, second ed. Academic Press.

Kittler, J., Hatef, M., Duin, R.P.W., Mates, J., 1998. On combining classifiers. IEEE Trans. Pattern Anal. Machine Intell. 20, 126–239.

Lam, L., Suen, C.Y., 1997. Application of majority voting to pattern recognition: An analysis of its behavior and performance. IEEE Trans. System man and Cybernet. 27, 553–568.

Miller, D.J., Uyar, H., 1998. Combined learning and use for a mixture model equivalent to the RBF classifier. Neural Comput. 10, 281–293.

Moody, J., Darken, C.J., 1989. Fast learning in networks of locally-tuned processing units. Neural Comput. 1, 281–294.

Moon, T.K., 1996. The expectation-maximization algorithm. Signal Proc. Mag. 13 (6), 47–60.

Richards, J.A., 1993. Remote Sensing Digital Image Analysis, second ed. Springer-Verlag, New York.

Serpico, S.B., Bruzzone, L., Roli, F., 1996. An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote-sensing images. Pattern Recognition Lett. 17 (13), 1331–1341.

Swain, P.H., 1978. Bayesian classification in a time-varying environment. IEEE Trans. System Man Cybernet. SMC-8, 880–883.