A Neural-Statistical Approach to Multitemporal and Multisource Remote-Sensing Image Classification

Lorenzo Bruzzone, *Member, IEEE,* Diego Fernández Prieto, *Student Member, IEEE,* and Sebastiano B. Serpico, *Member, IEEE*

***Abstract*—A data fusion approach to the classification of multisource and multitemporal remote-sensing images is proposed. The method is based on the application of the Bayes rule for minimum error to the "compound" classification of pairs of multisource images acquired at two different dates. In particular, the fusion of multisource data is obtained by using multilayer perceptron neural networks for a nonparametric estimation of posterior class probabilities. The temporal correlation between images is taken into account by the prior joint probabilities of classes at the two dates. As a novel contribution of this paper, such joint probabilities are automatically estimated by applying a specific formulation of the expectation-maximization (EM) algorithm to the data to be classified. Experiments carried out on a multisource and multitemporal data set confirmed the effectiveness of the proposed approach.**

***Index Terms*— Data fusion, expectation maximization, image classification, multisource multitemporal images, neural networks, remote sensing.**

I. Introduction

In the last decades, remote sensing has proved a powerful technology for monitoring the earth's surface and atmosphere at a global, regional, and even local scale. This is made possible by the large amount of data acquired by different types of sensors, which provide repeated coverage of the planet on a regular basis. As a consequence, an increasing quantity of multisource and multitemporal remote-sensing data acquired in many geographical areas is available. For a proper exploitation of these data, it is mandatory to develop effective data fusion techniques able to take advantage of such multisource and multitemporal characteristics. In particular, in the context of classification problems, data fusion may provide an improvement in accuracy (as compared with standard techniques applied to single-sensor/single-date images), which may be of primary importance in real applications.

Data-fusion techniques for the classification of remote-sensing images have been extensively investigated in the past years. Many papers that address the development of methodologies for the classification of multisensor (or multisource) images have been published [1]–[28]. Some studies have also been carried out on the integration of the information contained in multitemporal images in order to improve classification accuracy [29]–[33]. However, only a few papers have dealt with the integration of both multisensor and multitemporal remote-sensing images by the data-fusion process [34]–[36].

In this paper, we face the problem of the fusion of multisource and multitemporal information for the classification of remote-sensing images. The approach we propose is based on the application of the Bayes rule for minimum error to the "compound" classification [37] of pairs of multisource images acquired at different dates. The temporal correlation between the multisource images acquired at two dates is captured by using the joint class probabilities related to possible land-cover changes. A technique based on a specific formulation of the expectation-maximization (EM) algorithm [38] is applied to estimate such joint probabilities. As it represents the main innovative aspect of this paper, special attention will be given to this technique. Single-date multisource data are then used to compute nonparametric estimates of posterior class probabilities; to this end, a technique based on multilayer perceptron neural networks is applied. Experimental results, obtained by using a multisource and multitemporal data set related to an agro-forestry area located in the north of Italy, confirmed the effectiveness of the proposed approach.

The paper is organized into six sections. The next section provides an overview of the main approaches to remote-sensing data fusion reported in the literature. Section III gives a general description of the proposed approach. In Section IV, the method (based on the EM algorithm) we have derived for the estimation of the joint probabilities of classes is detailed. The data sets used in the experiments and the results obtained are reported in Section V. Finally, after discussion, conclusions are drawn in Section VI.

II. Previous Work

In this section, we briefly discuss the main methods reported in the remote-sensing literature related to the classification of multisource images and of multitemporal images.

Several methodologies have been proposed for the classification of multisensor/multisource images; they are mainly based on statistical, symbolic (evidential reasoning), and neural-network approaches.

Among statistical methods, the "stacked-vector" approach is the simplest one [1]. According to this approach, each pixel is represented by a vector that contains components from different sources (e.g., measures obtained by different sensors, texture features, ancillary data). Classification is usually performed by using a parametric statistical technique. However, this approach is not suitable when a common distribution

Manuscript received June 22, 1998; revised November 1, 1998. This work was supported by the Italian Space Agency (ASI).

The authors are with the Department of Biophysical and Electronic Engineering, University of Genoa, I-16145, Genova, Italy.

Publisher Item Identifier S 0196-2892(99)03448-8.

model cannot describe the various sources considered. Other well known approaches are stratification [2]–[6]; relaxation labeling [7], [8] which also includes a spatial context in the classification process; and the extended statistical fusion model proposed in [9] and [10]. Benediktsson and Swain [11] modified the last-mentioned approach by including "reliability factors" to weigh the importance of sensors according to their reliability. Thanks to this modification, significant improvements in classification accuracy were obtained.

The Dempster–Shafer evidence theory [12]–[14] was applied to classify multisource data by taking into account the uncertainties related to the different data sources involved [8], [10], [15]–[17]. Lee *et al.* [10] compared an approach of this kind with a statistical approach and showed that the latter method performs better. Another approach to multisource classification based on the Dempster–Shafer evidence theory is reported in [17], where an unsupervised method is proposed.

Neural-network classifiers provide an effective integration of different types of data. The nonparametric approach they implement allows the aggregation of different data sources into a stacked vector without need for assuming a specific probabilistic distribution of the data to be fused. A detailed introduction to neural networks can be found in [18] and [19]. In the literature, several studies on the classification of multisource remote-sensing data by neural networks have been reported [10], [20]–[28]. Some of them investigated and compared the performances of neural-network classifiers with those of both parametric and nonparametric statistical methods. Results (obtained by using different kinds of multisource data) point out the effectiveness of neural-network approaches for the classification of multisource data.

The aforementioned methods do not take into account the multitemporal aspect, as they consider only single-date information. In particular, concerning the classification process in a time-varying environment, only a few papers can be found in the remote-sensing literature. An early work on this subject was presented by Swain [29]. The author proposed a "cascade classifier" based on the generalization of the Bayes optimal strategy to the case of multiple observations. Kalayeh and Landgrebe [30] described a classifier that considers types of ground cover as stochastic systems with nonstationary Gaussian processes as inputs and with temporal variations in spectral responses as outputs. Then, multitemporal information was used assuming that the behavior of these processes can be modeled by a first-order Markov model. Also, Khazenie and Crawford [31] proposed a method for contextual classification that accounts for both spatial and temporal correlations of data. According to this approach, the feature vectors are modeled as resulting from a class-dependent process plus a contaminating noise process; the noise process is considered autocorrelated in both the space and time domains. Middelkoop [32] described a knowledge-based classifier that integrates the data acquired at a given time with those acquired in previous years. Spatio-temporal contextual classification was also addressed by Jeon and Landgrebe [33]. Gibbs random fields were used to model the spatial context, whereas temporal class dependencies were taken into account by class transition probabilities. Experiments with multitemporal thematic mapper data showed a significant improvement in classification accuracy as compared with maximum likelihood classification.

All the papers on classification methods cited so far consider either the multisensor or the multitemporal aspect of remote-sensing data. A method that exploits both characteristics was proposed by Solberg *et al.* [34]. Starting from the multisource approach described in [11] and in order to make it suitable for multitemporal data, they incorporated *a priori* information about the likelihood of the changes that may occur between image acquisition dates. This early work was extended in [35] by using Markov random fields in the fusion model to allow a tight coupling of data sources (which may also include GIS ground cover data).

An issue emerging from the papers on classification in time-varying environments [29], [33]–[35] is the definition of a methodology for the estimation of the terms that take into account the probabilities of ground-cover changes between different image acquisition dates. The new methodology proposed in Section IV to estimate such terms (whose importance was already pointed out in [29]) represents the main aspect of our paper. The authors already faced a similar problem in [36], in the context of detection of land-cover changes. The solution suggested in that paper for the estimation of the aforementioned probabilistic terms was based on an iterative procedure for which a proof of convergence was not available. The approach presented in Section IV uses the EM algorithm, for which, on the contrary, convergence was proved (even though to a local maximum).

III. Classification of Multitemporal Multisource Data

A. The Bayes Rule for Compound Classification

Let us consider the problem of classifying a geographical area by analyzing two multisource image data sets I_1 and I_2 acquired at two times t_1 and t_2, respectively. Each data set may contain images derived from different sources (e.g., optical and SAR images). We assume that all the images of the two data sets refer to the same ground area, and that they are coregistered and appropriately transformed into the same spatial resolution.

In general, the spatial and temporal contextual information plays an important role in the classification process. One of the main purposes of this paper is to assess the potentialities of the technique we propose to estimate prior joint class probabilities, which are related to the temporal context of the two multisource data sets. Therefore, for simplicity, we focus on the temporal context only, and we do not explicitly consider the spatial context. Furthermore, we assume that, for each pixel of one data set, all the temporal contextual information is conveyed by the spatially corresponding pixel of the other data set. This seems a reasonable assumption for the current procedure by which we consider only two acquisition dates and disregard the spatial context.

We characterize the above pair of temporally correlated pixels (x_1, x_2), x_1 being a pixel of the image data set I_1 and x_2 the spatially corresponding pixel of the image data set I_2,

by a pair of feature vectors (X_1, X_2). Each feature vector X_i is obtained by stacking together the measures provided by the available sensors, as is done for the stacked vector approach. Let $\Omega = \{\omega_1, \omega_2, \cdots, \omega_{M1}\}$ be the set of possible land-cover classes at time t_1, and let $N = \{\upsilon_1, \upsilon_2, \cdots, \upsilon_{M2}\}$ be the set of possible land-cover classes at time t_2.

We can consider two different types of classification: 1) the "compound classification" of each pair of pixels (x_1, x_2), which involves finding the "best" pair of classes (ω_i, υ_j) to be assigned to each pair of pixels; or 2) the classification of the pixels of one of the two image data sets by utilizing the information contained in both image data sets I_1 and I_2.

As a classification strategy, we adopt the Bayes rule for minimum error and apply it to the first type of classification, i.e., the "compound classification" of (x_1, x_2) [37]

$$x_1 \in \omega_m \text{and} x_2 \in \nu_n \text{ so that } P(\omega_m, \nu_n / X_1, X_2) = \max_{\omega_i, \nu_j} \{P(\omega_i, \nu_j / X_1, X_2)\}. \quad (1)$$

For the second type of classification, i.e., if the image data set I_2 is to be classified, the Bayes rule becomes

$$x_2 \in \nu_n \text{ so that } P(\nu_n / X_1, X_2) = \max_{\nu_j} \{P(\nu_j / X_1, X_2)\}. \quad (2)$$

The two classification problems, as well as their solutions, are tightly linked. In this paper, we shall focus only on the former, i.e., the compound classification of the two image data sets (Swain in [29] addressed the latter problem).

Under the conventional assumption of class-conditional independence in the time domain, we can write [29], [33]

$$p(X_1, X_2 / \omega_i, \nu_j) = p(X_1 / \omega_i) p(X_2 / \nu_j). \quad (3)$$

Consequently, we can replace the search for the maximum in (1) with

$$\max_{\omega_i, \nu_j} \left\{ \frac{P(\omega_i / X_1) P(\nu_j / X_2) P(\omega_i, \nu_j)}{P(\omega_i) P(\nu_j)} \right\}. \quad (4)$$

In general, the above assumption may lead to suboptimal solutions. For example, when the ground cover associated with a given pixel does not change between t_1 and t_2 [i.e., $\omega_i \equiv \upsilon_j$ in (3)], it is likely that some properties of the ground, which contribute to determining the values of sensorial measures, may be saved between the two acquisition dates. This implies a correlation between X_1 and X_2 that is not taken into account on the right-hand side of (3). However, we adopt the above assumption as it allows a significant simplification of the problem.

The *a priori* class probabilities $P(\omega_i)$ and $P(\upsilon_j)$ in (4) are estimated from the two "training sets" (described in the next subsection) by computing the relative frequency of each class. The estimations of the remaining terms (i.e., the posterior class probabilities and the prior joint class probabilities) are worth being considered more deeply.

B. Multisource Data Fusion

To estimate the single-date, multivariate, conditional probabilities in (4), we need to fuse multisource data (for example, to estimate $P(\omega_i / X_1)$ we have to fuse the multisource data in I_1). In general, the definition of a common statistical model of multisource data (e.g., optical and SAR images) may be quite complex. Therefore, we adopted a nonparametric technique. In particular, we utilized multilayer perceptron neural networks, which, if properly trained by the error backpropagation algorithm [18], provide estimates of posterior class probabilities; such estimates can be optimized according to a predefined criterion [39], [40]. In our case, we adopted the minimum mean square error (MSE) criterion.

To this end, two neural networks (one for the date t_1, the other for the date t_2) need to be trained separately on two "training sets," for which the ground truth must be available. The two training sets may be defined independently of each other (i.e., they may refer to different ground points).

As a result of the training phase, the values of the internal parameters of the neural networks (i.e., the so-called weights and biases) are obtained. At this point, each neural network can be used to compute the estimate of the posterior class probability at the corresponding date (output of the network), given the feature vector X_i (input of the network) (Fig. 1).

C. Multitemporal Data Fusion

The only term in (4) that takes into account the correlation between the images acquired at the two times is the joint probability $P(\omega_i, \upsilon_j)$. The value of such a probability may remarkably affect the classification result [29]. In particular, if one assumes $P(\omega_i, \upsilon_j) = P(\omega_i) P(\upsilon_j)$, then (4) is equivalent to a separate application of the Bayes rule to the two data sets I_1 and I_2. This involves the loss of all the advantages of the multitemporal analysis (such a situation will be used as a reference to evaluate the performances of the proposed method).

For the above reasons, an accurate estimation of prior joint class probabilities is very important. By contrast, a direct estimation from training data would require impractical constraints: training data should contain pairs of pixels (at spatially corresponding positions in the two images) representative of all possible combinations of classes at the two times.

In this paper, we propose the application of the EM algorithm [41] to estimate the joint probabilities $P(\omega_i, \upsilon_j)$, as described in the next section. Such a technique is fully automatic and adaptive to the data sets to be classified (even though it makes use of the prior class probabilities $P(\omega_i)$ and $P(\upsilon_j)$ and of the posterior class probabilities $P(\omega_i / X_1)$ and $P(\upsilon_j / X_2)$ computed from training data, as described above).

IV. Estimation of Prior Joint Probabilities by the EM Algorithm

A. The EM Algorithm

In several image-processing and pattern-recognition problems, it is necessary to estimate the parameters that characterize a probability distribution function. In many cases, this task may become rather complex because a direct access to the data necessary to estimate such parameters is impossible (or some of these data are lacking), and only some observations

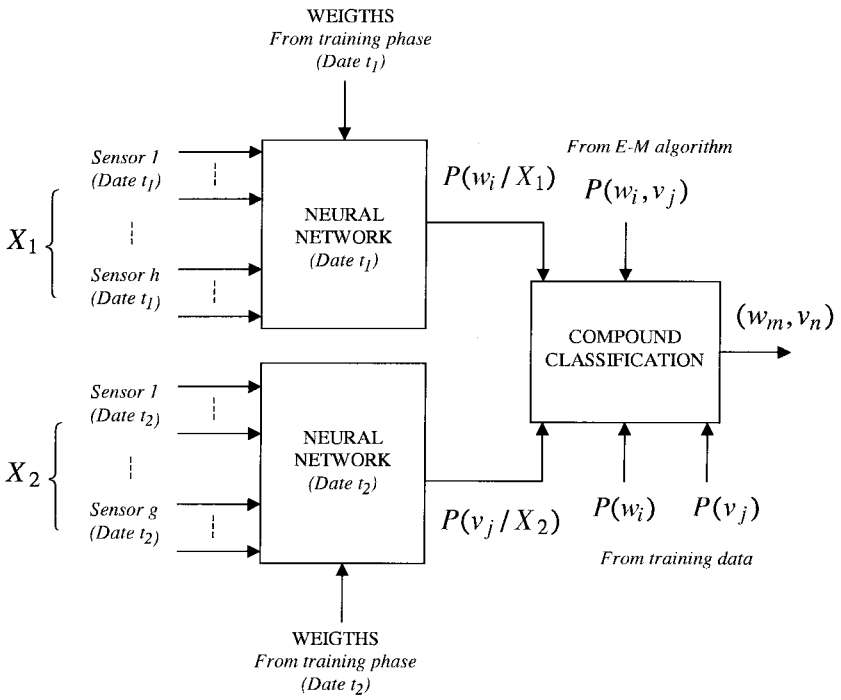


Fig. 1. Block diagram of the compound classification process. The pair (ω_m, υ_n) corresponds to the maximum in (4).

related to them are available. In some of these cases, it is, however, possible to perform parameter estimations by using the EM algorithm [41].

The EM algorithm constitutes a general approach to an iterative computation of maximum-likelihood estimates of parameters when there is a many-to-one mapping from an underlying distribution to the distribution governing an observation. Such an algorithm is particularly useful in estimating the components of mixture distributions. At present, the EM algorithm is becoming increasingly popular for the analysis of remote-sensing images. In this context, an interesting example of the use of the EM algorithm for the estimation of class distributions has been given by Shahshahani and Landgrebe [42].

The general formulation of the EM algorithm consists of two major steps: an expectation step and a maximization step. The expectation is computed with respect to the unknown underlying variables, using the current estimates of the parameters and conditioned upon the observations. The maximization step provides new estimates of the parameters. These two steps are iterated until convergence.

An important aspect of the EM algorithm concerns its convergence properties. It is possible to prove that, at each iteration, the estimated parameters provide an increase in the likelihood function until a local maximum is reached. Despite the fact that convergence can be ensured, it is impossible to ensure that the algorithm will converge to the global maximum of the likelihood (only in specific cases is it possible to guarantee the convergence to the global maximum).

A detailed description of the EM algorithm and of the related theoretical aspects is beyond the scope of this paper. Refer to the literature for an in-depth analysis of such an algorithm [41], [43].

B. Estimation of Prior Joint Probabilities

To estimate the prior joint probabilities $P(\omega_i, \upsilon_j)$, we propose to use an iterative fixed-point EM-like algorithm [38], which is a specific version of the EM algorithm and is suited to evaluating only the proportions of a set of parameters. In our case, we estimate only the joint probabilities of classes, assuming no need to update the estimates of the posterior probabilities and of the *a priori* probabilities of classes (such estimations are performed as described in Section III) during the successive iterations. Thus, a more stable algorithm than the standard EM is obtained. However, it is worth noting that the standard EM would also allow both the posterior probabilities and *a priori* probabilities of classes to be estimated.

The probabilities $P(\omega_i, \upsilon_j)$ are regarded as the elements of the matrix JP (of size $M_1 \times M_2$), which is computed by maximizing the following pseudolikelihood:

$$L(JP) = \prod_{q=1}^{S} \left(\sum_{\omega_n \in \Omega} \sum_{\nu_m \in N} P(\omega_n, \nu_m) P\left(X_1^q, X_2^q / \omega_n, \nu_m\right) \right) \tag{5}$$

where S is the total number of pixels to be classified and X_k^q is the qth pixel of the image I_k. It is possible to prove that the recursive equation to be used to estimate $P(\omega_i, \upsilon_j)$ by maximizing (5) is

$$P_{k+1}(\omega_i, \nu_j) = \frac{1}{S} \sum_{q=1}^{S} \frac{P_k(\omega_i, \nu_j) P\left(X_1^q, X_2^q / \omega_i, \nu_j\right)}{\sum_{\omega_n \in \Omega} \sum_{\nu_m \in N} P_k(\omega_n, \nu_m) P\left(X_1^q, X_2^q / \omega_n, \nu_m\right)} \tag{6}$$

where $P_k(\omega_i, \upsilon_j)$ is the iterative joint probability estimate at the kth iteration. Such estimates are initialized by assigning

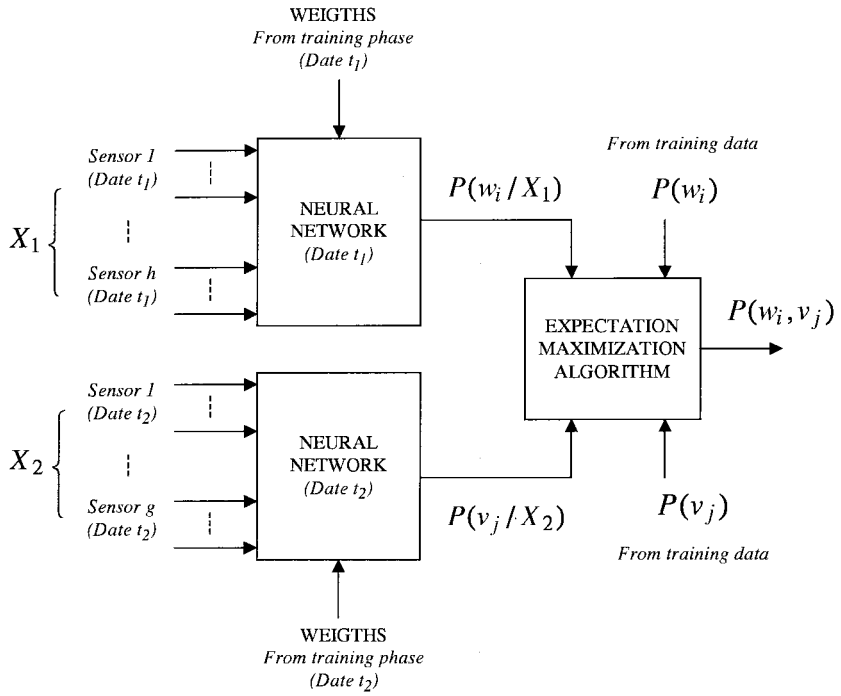


Fig. 2. Block diagram of the process used to estimate the prior joint probabilities $P(\omega_i, v_j)$.

equal probabilities to each pair of classes

$$P_0(\omega_i,\nu_j) = \frac{1}{M_1 \cdot M_2}, \quad \forall\omega_i \in \Omega, \quad \nu_j \in N. \tag{7}$$

Under the hypothesis made in our approach (see Section III), it is possible to prove that (6) can be rewritten as

$$P_{k+1}(\omega_i,\nu_j) = Aij \sum_{q=1}^{S} \frac{P_k(\omega_i,\nu_j)P(\omega_i/X_1^q)P(\nu_j/X_2^q)}{\sum_{\omega_n\in\Omega}\sum_{\nu_m\in N}\frac{P_k(\omega_n,\nu_m)}{P_k(\omega_n)P_k(\nu_m)}P(\omega_n/X_1^q)P(\nu_m/X_2^q)} \tag{8}$$

where

$$Aij = \frac{1}{SP(\omega_i)P(\nu_j)}. \tag{9}$$

The algorithm is iterated until convergence. Convergence is reached when the maximum difference between the estimates at two successive iterations is below a threshold. More precisely, the stop criterion is defined by the following inequality:

$$\max_{\omega_i,\nu_j}\{|P_{k+1}(\omega_i,\nu_j) - P_k(\omega_i,\nu_j)|\} < \varepsilon, \quad \omega_i \in \Omega, \quad \nu_j \in N \tag{10}$$

where $\varepsilon \in [0,1]$.

The estimates of $P(\omega_i, v_j)$ obtained at convergence (Fig. 2) are then applied to the compound classification rule (Fig. 1).

V. Experimental Results

A. Data Sets

The considered data sets are related to an agricultural area in the basin of the Po river (northern Italy). They consist of two multisensor image data sets acquired by the Landsat Thematic Mapper (TM) sensor and by the ERS-1 synthetic aperture radar (SAR) sensor in April and May 1994. As an example of the images used, Fig. 3 shows band 2 of the TM and ERS-1 images acquired in April 1994. Both the TM images and the ERS-1 May image were registered to the ERS-1 April image. The available ground truth was used to prepare the training sets (utilized to train neural networks and to estimate prior single-class single-date probabilities) and the "test sets" (utilized for performance evaluation and comparison). Table I provides the classes and the related numbers of pixels included in the training and test sets.

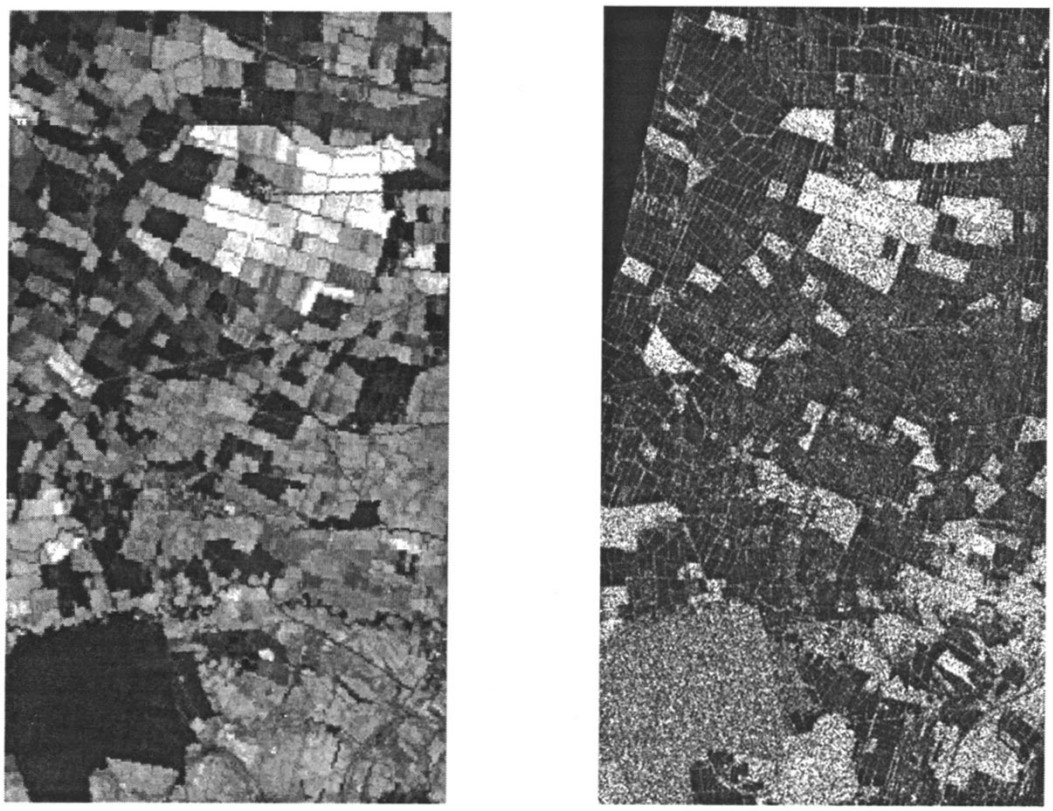
The two feature vectors X_1 and X_2 (related to time t_1 and t_2, respectively) consisted of 11 elements each, i.e., six TM bands (all bands but the infrared thermal one), the intensity of the ERS-1 image (C band, VV polarization), and four texture features computed from the ERS-1 image by means of the gray-level co-occurrence matrix (correlation, entropy, sum variance, difference entropy).

Several experiments were carried out by utilizing one sensor or both sensors at each of the two times in order to test the validity of the proposed approach. For the sake of brevity, in the following, we shall report on the two most significant experiments only.

B. First Experiment

The first experiment involved the fusion of the data provided by both sensors at both times.

Two feed-forward multilayer perceptron neural networks were trained on the training sets and used to estimate the posterior class probabilities $P(\omega_i/X_1)$ and $P(v_j/X_2)$. Fully connected architectures with one hidden layer were defined; in both neural networks, the number of hidden neurons was



(a) (b)

Fig. 3. Images acquired in April 1994 and utilized for the experiments: (a) band 2 of the Landsat TM; (b) ERS-1 SAR image.

fixed at eight. The number of input neurons was set equal to the size of the feature vectors (i.e., 11) and the number of output neurons to the number of classes (i.e., four at time t_1 and five at time t_2).

The EM algorithm was applied and the convergence parameter ε was selected equal to 0.001. Six iterations of this algorithm were necessary to reach convergence in the estimations of the joint probabilities $P(\omega_i, \upsilon_j)$. We point out that such estimations can be performed directly on the unknown data to be classified (in our case, to assess performances, they were performed on the test sets). Tables II and III show, respectively, the true values of the joint probabilities for the test set and the estimates of these probabilities obtained at the last iteration of the EM algorithm. As one can see, the final estimates were quite satisfactory, as the maximum error on a single joint probability was equal to about 0.02. Such an error was incurred on the joint probability related to the pair of classes bare soil/dry rice fields. In order to better illustrate the behavior of the EM algorithm in this experiment, Fig. 4 shows the trend of the estimate of the joint probability P(*wood*, *wood*) versus the number of iterations. As one can see, the estimate evolves from a value of 0.363 to 0.394 in only two iterations and reaches its final value (i.e., 0.395) at the sixth iteration.

Classification results obtained on the test sets with the proposed technique at the two times are shown in Table IV in the columns "% Error EM." For a comparison, the performances of the neural networks applied separately to the images I_1 and I_2 are also included (columns "% Error NN").

TABLE I
TRAINING AND TEST SETS UTILIZED FOR THE EXPERIMENTS: (a) APRIL 1994; (b) MAY 1994

Land-cover classes (April 1994)	Number of pixels	
	Training set	Test set
Wet rice-fields	1798	1846
Bare soil	1258	1205
Cereals	123	108
Wood	2136	2069
Total	5315	5228

(a)

Land-cover classes (May 1994)	Number of pixels	
	Training set	Test set
Corn	333	344
Dry rice-fields	2139	2181
Wet rice-fields	584	526
Cereals	123	108
Wood	2136	2069
Total	5315	5228

(b)

The total percent error made with our technique is quite small and significantly smaller than that of the neural networks applied separately to I_1 and I_2. In particular, the overall classification error was reduced from 2.68% to 0.98% on the April data set and from 3.73% to 3.25% on the May data set. However, the improvement appears more remarkable if one considers single-class accuracies, in particular, for the two

TABLE II
TRUE VALUES OF THE PRIOR JOINT PROBABILITIES CONCERNING THE TEST SET

May 1994 / April 1994	Corn	Dry rice-fields	Wet rice-fields	Cereals	Wood
Wet rice-fields	0.000	0.285	0.068	0.000	0.000
Bare soil	0.066	0.132	0.033	0.000	0.000
Cereals	0.000	0.000	0.000	0.021	0.000
Wood	0.000	0.000	0.000	0.000	0.396

TABLE III
ESTIMATES OF THE PRIOR JOINT PROBABILITIES PROVIDED BY THE EM ALGORITHM IN THE FIRST EXPERIMENT, USING BOTH THE SAR AND TM IMAGES AT THE TWO DATES

May 1994 / April 1994	Corn	Dry rice-fields	Wet rice-fields	Cereals	Wood
Wet rice-fields	0.000	0.284	0.069	0.000	0.000
Bare soil	0.063	0.110	0.042	0.000	0.000
Cereals	0.000	0.015	0.001	0.020	0.002
Wood	0.000	0.000	0.000	0.000	0.395

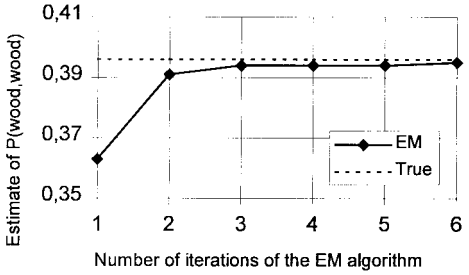


Fig. 4. Behavior of the estimate of the prior joint probability $P(_{wood,\,wood})$ versus the number of iterations of the EM algorithm in the first experiment, carried out using both the SAR and TM images at the two dates.

most difficult classes: cereals in April (error reduction from 50.0% to 18.52%) and corn in May (error reduction from 21.8% to 15.12%).

Fig. 5 allows one to gain a deeper insight into the effects of the accuracies of the joint-probability estimates on the classification results. The figure shows the behaviors of the overall classification errors versus the number of iterations of the EM algorithm at both dates. As one can see, when the number of iterations increases, the overall classification errors exhibit a decreasing trend, which results from the increasing accuracies of the estimates. In particular, at the first iteration, at which the estimates obtained are not very accurate, the overall classification errors in April and May are equal to 2.68% and 4.86%, respectively. As the algorithm evolves, the increasing accuracies of the estimates progressively reduce the classification errors to their final values (i.e., 0.98% in April and 3.25% in May).

C. Second Experiment

For different reasons, in real applications, the available sensors at two dates may be different. For example, at one date, both optical and SAR images may be available, whereas, at the other date, the optical image may convey no information because of cloud covering. As another example, one of the sensors might not yet (or any more) be operating at the time of the first (or second) acquisition. Finally, the costs of images provided by different sensors may be very different;

TABLE IV
CLASSIFICATION RESULTS IN THE FIRST EXPERIMENT, USING BOTH THE SAR AND TM IMAGES AT THE TWO DATES. (a) APRIL 1994; (b) MAY 1994. THE CLASSIFICATION ERRORS INCURRED WHEN USING NEURAL NETWORKS SEPARATELY ON EACH SINGLE DATE (% ERROR NN) AND WHEN USING THE PROPOSED TECHNIQUE (% ERROR EM) ARE GIVEN

Land-cover classes (April 1994)	%Error NN (test set)	%Error EM (test set)
Wet rice-fields	0.49	0.38
Bare soil	4.23	1.58
Cereals	50.0	18.52
Wood	1.26	0.24
Overall	2.68	0.98

(a)

Land-cover classes (May 1994)	%Error NN (test set)	%Error EM (test set)
Corn	21.80	15.12
Dry rice-fields	1.33	1.56
Wet rice-fields	12.93	11.22
Cereals	15.74	18.52
Wood	0.29	0.24
Overall	3.73	3.25

(b)

as a consequence, one might prefer to buy only one image acquired by the most expensive sensor. Therefore, we deem it interesting to report on an experiment carried out by using different sensors at the two acquisition dates considered.

To this end, for the second experiment, we included only the ERS-1 image in the April 1994, data set and both the TM and the ERS-1 images in the May 1994 data set. The feature vector X_1 was reduced to five components corresponding to the above-defined features derived from the ERS-1 image. The neural network applied to the April 1994, data was modified accordingly: the number of input neurons was reduced to five, the number of hidden neurons was maintained at eight, and, obviously, the number of output neurons at four. Nothing was changed for the neural network applied to the May 1994 data set (neither did we need to retrain the network).

In this experiment, the value of the convergence parameter ε was the same as in the previous experiment (i.e., 0.001), but the EM algorithm required eight iterations to reach convergence.

TABLE V
ESTIMATES OF THE PRIOR JOINT PROBABILITIES PROVIDED BY THE EM ALGORITHM IN THE SECOND EXPERIMENT, USING THE SAR IMAGE AT THE FIRST DATE AND BOTH THE SAR AND TM IMAGES AT THE SECOND DATE

April 1994 \ May 1994	Corn	Dry rice-fields	Wet rice-fields	Cereals	Wood
Wet rice-fields	0.000	0.285	0.070	0.000	0.011
Bare soil	0.063	0.118	0.043	0.000	0.000
Cereals	0.000	0.003	0.000	0.019	0.001
Wood	0.000	0.004	0.000	0.000	0.384

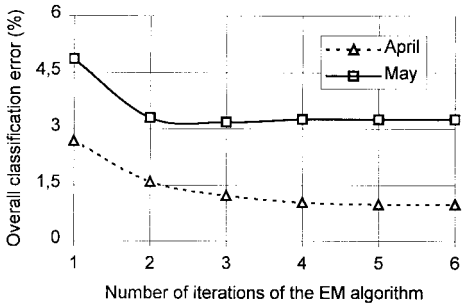


Fig. 5. Behavior of the classification error versus the number of iterations of the EM algorithm in the first experiment, carried out using both the SAR and TM images at the two dates.

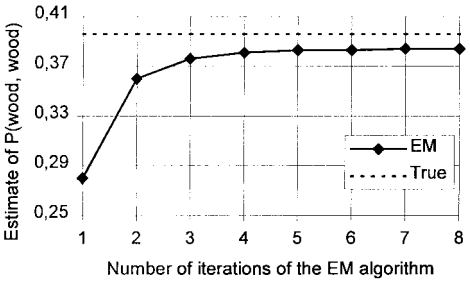


Fig. 6. Behavior of the estimate of the prior joint probability $P(wood, wood)$ versus the number of iterations of the EM algorithm in the second experiment, carried out using the SAR image at the first date and both the SAR and TM images at the second date.

Table V gives the final estimates of the joint probabilities obtained at the last iteration of the EM algorithm. As in the first experiment, the obtained estimates were quite accurate (the maximum error on a single probability estimate was equal to 0.014). Fig. 6 shows the value of the joint probability $P(wood, wood)$ versus the number of iterations of the EM algorithm. As one can see, a value equal to 0.280 was obtained for $P(wood, wood)$ at the first iteration, which implies an error equal to 0.116, as compared with the true value equal to 0.396. This error was steadily reduced to 0.012 at the last iteration, obtaining a final value equal to 0.384 for $P(wood, wood)$.

The classification accuracies obtained in this experiment are shown in Table VI. As expected, in comparison with the results of the first experiment, the classification error at the first date (April), based only on the ERS-1 April image, sharply increased; obviously, the column "% Error NN" for May 1994 is identical. The exploitation of the temporal correlation among the images by means of the EM estimations of the joint class probabilities allowed a significant reduction in the classification errors on the April data set (from 6.87% to 2.54%). The analysis of single class errors confirms the improvement in the classification provided by our method: most single-class percent errors were reduced; in particular, the unacceptable error on the cereals class at the first date was sharply reduced (from 83.33% to 25.0%). As the error on the single-date independent classification of the May data set was already quite small and much smaller than that on the April data set, only a slight improvement was provided by our technique in the classification of the May data set (from 3.73% to 3.37%).

TABLE VI
CLASSIFICATION RESULTS IN THE SECOND EXPERIMENT USING THE SAR IMAGE AT THE FIRST DATE AND BOTH THE SAR AND TM IMAGES AT THE SECOND DATE. (a) APRIL 1994; (b) MAY 1994. THE CLASSIFICATION ERRORS INCURRED WHEN USING NEURAL NETWORKS SEPARATELY ON EACH SINGLE DATE (% ERROR NN) AND WHEN USING THE PROPOSED TECHNIQUE (% ERROR EM) ARE GIVEN

Land-cover classes (April 1994)	%Error NN (test set)	%Error EM (test set)
Wet rice-fields	1.46	0.54
Bare soil	6.89	2.82
Cereals	83.33	25.00
Wood	7.68	3.00
Overall	6.87	2.54

(a)

Land-cover classes (May 1994)	%Error NN (test set)	%Error EM (test set)
Corn	21.80	14.83
Dry rice fields	1.33	1.56
Wet rice fields	12.93	11.22
Cereals	15.74	25.00
Wood	0.29	0.24
Overall	3.73	3.37

(b)

VI. DISCUSSION AND CONCLUSIONS

In this paper, we have presented a data-fusion approach to the classification of multitemporal and multisource remote-sensing images. The proposed approach is based on the application of the Bayes rule for minimum error to the compound classification of two image data sets acquired at two different dates. Concerning the multitemporal aspect, we assumed that, for simplicity, the temporal correlation between the two data sets can be taken into account by the prior joint probabilities of classes at the two dates. Multisource data fusion was then performed by means of multilayer perceptron neural networks, which provide nonparametric estimates of posterior class probabilities on the basis of single-date multisource data.

In addition to the flexibility provided by nonparametric techniques, the use of neural networks for data fusion offers the general advantages of every neural-network approach: intrinsic parallelism, adaptability to data, and robustness to noise and errors on training data. Unfortunately, the neural-network approach also exhibits some well-known drawbacks. In particular, for the neural model we adopted, no general rules exist to define the network topology and to establish the procedure of the training process; moreover, it is difficult to interpret the network behavior [27], [28].

From a methodological viewpoint, the main innovation of this paper lies in the use of the EM algorithm for the estimation of the prior joint probabilities of classes. The importance of these probabilities for the classification of multitemporal data was already pointed out by Swain [29]. In the literature, prior joint probabilities are usually chosen manually by a human expert on the basis of *a priori* knowledge derived from the characteristics of the geographical area considered and from the time intervals between acquisitions. The advantage of the proposed technique based on the EM algorithm consists in the possibility of computing the estimates of joint probabilities directly from the data set to be classified and in a fully automatic way. This overcomes the drawbacks resulting from the need for a human intervention and from the dependence on the accuracy of *a priori* knowledge. In addition, it has been proven that the EM algorithm converges to a local maximum of the likelihood function [41], [43]. Even though the convergence to a global maximum cannot, in general, be demonstrated, it has been shown that, in many applications, the EM algorithm provides accurate estimates [41], [43].

In the present formulation, the estimation of the prior joint probabilities depends on the training data, as it exploits the estimates of the single-date prior class probabilities and of the single-date posterior class probabilities previously computed from training data. It is possible to generalize the use of the EM algorithm to estimate also the latter probabilities directly from the data to be classified; this would make the algorithm less sensitive to differences in the statistical distributions of training and test data. On the other hand, it might reduce the reliability of the estimates obtained.

Experimental results on multisource and multitemporal data have been reported that attest the effectiveness of our technique. In particular, the proposed technique may provide a significant improvement on an independent classification of multitemporal data sets, even when the sensors available at two times are different (e.g., both optical and SAR sensors at one time and only one SAR sensor at the other time). It is worth noting that, in any case, our technique is not able to improve very poor classification results obtained at one date, not even if a better classification is performed at the other date.

In consideration of the above-discussed characteristics and of the interesting experimental results obtained, in our opinion, the EM-based technique proposed can be considered a sufficiently reliable tool for the estimation of prior joint probabilities.

Concerning the fusion of multitemporal information, we have considered only the case of two data sets (corresponding to two acquisition dates). A generalization to more than two multitemporal data sets could be obtained by extensions to the proposed algorithms, such that may allow different tradeoffs between simplicity and accuracy. Additional research on this issue is necessary to reach the most effective solution.

Finally, in this paper, we have focused on the temporal context, whereas we have disregarded the spatial context. As an interesting development, the proposed method might be extended to include the information conveyed by the spatial context (e.g., by using Markov random fields [30], [35]).

ACKNOWLEDGMENT

The authors wish to thank Dr. M. A. Gomarasca (C.N.R.—I.R.R.S.—Telerilevamento, Milan, Italy) for providing the multisource images and for assisting in the related agronomic interpretation.

REFERENCES

[1] D. G. Leckei, "Synergism of synthetic aperture radar and visible/infrared data for forest type discrimination," *Photogramm. Eng. Remote Sens.*, vol. 56, pp. 1237–1246, 1990.

[2] S. E. Franklin, "Ancillary data input to satellite remote sensing of complex terrain phenomena," *Comput. Geosci.*, vol. 15, pp. 799–808, 1989.

[3] C. F. Hutchinson, "Techniques for combining Landsat and ancillary data for digital classification improvement," *Photogramm. Eng. Remote Sens.*, vol. 48, pp. 123–130, 1982.

[4] D. R. Peddle and S. E. Franklin, "Image texture processing and data integration for surface pattern discrimination," *Photogramm. Eng. Remote Sens.*, vol. 57, pp. 413–420, 1991.

[5] A. H. Strahler and N. A. Bryant, "Improving forest cover classification accuracy from Landsat by incorporating topographic information," in *Proc. 12th Int. Symp. Remote Sensing Environment*, Ann Arbor, MI, 1978, pp. 927–942.

[6] A. H. Strahler, "The use of prior probabilities in maximum likelihood classification of remotely sensed data," *Remote Sens. Environ.*, vol. 10, pp. 135–163, 1980.

[7] J. A. Richards, D. A. Landgrebe, and P. H. Swain, "A means for utilizing ancillary information in multispectral classification," *Remote Sens. Environ.*, vol. 12, pp. 463–477, 1982.

[8] J. A. Richards, *Remote Sensing Digital Image Analysis*, 2nd Ed. New York: Springer-Verlag, 1993.

[9] P. H. Swain, J. A. Richards, and T. Lee, "Multisource data analysis in remote sensing and geographic information processing," in *Proc. 11th Int. Symp. Machine Processing of Remotely Sensed Data 1985*, West Lafayette, IN, June 1985, pp. 211–217.

[10] T. Lee, J. A. Richards, and P. H. Swain, "Probabilistic and evidential approaches for multisource data analysis" *IEEE Trans. Geosci. Remote Sensing*, vol. GE-25, pp. 283–293, May 1987.

[11] J. A. Benediktsson and P. H. Swain, "A method of statistical multisource classification with a mechanism to weight the influence of the data sources," in *Proc. IEEE Symp. Geoscience and Remote Sensing (IGARSS)*, Vancouver, B.C., Canada, July 1989, pp. 517–520.

[12] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1979.

[13] T. D. Garvey, J. D. Lowrance, and M. A. Fischler, "An inference technique for integrating knowledge from disparate sources," in *Proc. 7th Int. Conf. Artificial Intelligence*, Vancouver, B.C., Canada, 1981, pp. 319–325.

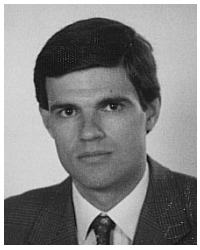
[14] M. Goldberg, D. G. Goodenough, M. Alvo, and G. M. Karam, "A hierarchical expert system for updating forestry maps with Landsat data," *Proc. IEEE*, vol. 73, pp. 1054–1063, 1985.

[15] H. Kim and P. H. Swain, "A method for classification of multisource data using interval-valued probabilities and its application to HIRIS data," in *Proc. Workshop Multisource Data Integration in Remote Sensing*, NASA Conf. Publ. 3099, Greenbelt, MD, June 1990, pp. 27–38.

[16] N. Lehrer, G. Reynolds, and J. Griffith, "Initial hypothesis formation in image understanding using an automatically generated knowledge base," in *Proc. Workshop Image Understanding*, Los Angeles, CA, Feb. 1987, pp. 521–537.

[17] S. Le Hègarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in

multisource remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, pp. 1018–1031, July 1997.
[18] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*. Boston, MA: Addison-Wesley, 1991.
[19] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Clarendon, 1995.
[20] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural networks approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 540–552, July 1990.
[21] NASA, in *Proc. Int. Workshop on Multisource Data Integration in Remote Sensing*, NASA Conference Publications 3099, Maryland, 1990.
[22] IAPR TC7, in *Proc. Int. Workshop on Multisource Data Integration in Remote Sensing for Land Inventory Applications*, Wageningen Agricultural Univ. Pub., Delft, The Netherlands, 1992.
[23] X. Zhuang, B. A. Engel, M. F. Baumgardner, and P. H. Swain, "Improving classification of crop residues using digital land ownership data and Landsat TM imagery," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 11, pp. 1487–1492, 1991.
[24] D. L. Civco and Y. Wang, "Classification of multispectral, multitemporal, multisource spatial data using artificial neural networks," in *Proc. 1994 Annu. ASPRS/ACSM Conv.*, Reno, NE, 1994, vol. 1, pp. 123–133.
[25] P. Gong, "Integrated analysis of spatial data from multiple sources: An overview," *Can. J. Remote Sens.*, vol. 40, no. 4, pp. 349–359, 1994.
[26] L. Bruzzone, C. Conese, F. Maselli, and F. Roli, "Multisource classification of complex rural areas by statistical and neural-network approaches," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 5, pp. 523–533, May 1997.
[27] S. B. Serpico, L. Bruzzone, and F. Roli, "An experimental comparison of neural and statistical nonparametric algorithms for supervised classification of remote-sensing images," *Pattern Recognit. Lett.*, vol. 17, no. 13, pp. 1331–1341, Nov. 1996.
[28] S. B. Serpico and F. Roli, "Classification of multisensor remote-sensing images by structured neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, pp. 562–578, May 1995.
[29] P. H. Swain, "Bayesian classification in a time-varying environment," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 880–883, Dec. 1978.
[30] H. M. Kalayeh and D. A. Landgrebe, "Utilizing multitemporal data by a stochastic model," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-24, pp. 792–795, Sept. 1986.
[31] N. Khazenie and M. M. Crawford, "Spatial-temporal autocorrelated model for contextual classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 529–539, July 1990.
[32] J. Middelkoop and L. L. F. Jansen, "Implementation of temporal relationships in knowledge based classification of satellite images," *Photogramm. Eng. Remote Sens.*, vol. 57, pp. 937–945, 1991.
[33] B. Jeon and D. A. Landgrebe, "Classification with spatio-temporal interpixel class dependency contexts," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, pp. 663–672, July 1992.
[34] A. H. S. Solberg, A. K. Jain, and T. Taxt, "Multisource classification of remotely sensed data: Fusion of Landsat TM and SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 768–778, July 1994.
[35] A. H. S. Solberg, T. Taxt, and A. K. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, pp. 100–113, Jan. 1996.
[36] L. Bruzzone and S. B. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 35, pp. 858–867, July 1997.
[37] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[38] V. Granville and J. P. Rasson, "Multivariate discriminant analysis and maximum penalized likelihood density estimation," *J. R. Stat. Soc.*, vol. 57, no. 3, pp. 501–517, 1995.
[39] H. Gish, "A probabilistic approach to the understanding and training of neural network classifiers," in *Proc. 1990 Int. Conf. Acoustics, Speech, Signal Processing*, Apr. 3–6, 1990, pp. 1361–1364.
[40] M. Richard and Lippmann, "Neural network classifiers estimate Bayesian a posteriory probabilities," *Neural Comput.*, vol. 3, no. 4, pp. 461–463, 1991.
[41] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.
[42] B. M. Shahshahani and D. A. Langrebe, "The effect of unlabeled samples in reducing the small size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 1087–1095, Sept. 1994.
[43] T. K. Moon, "The expectation-maximization algorithm," *Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.



Lorenzo Bruzzone (S'95–M'99) received the "Laurea" (M.S.) degree in electronic engineering and the Ph.D. degree in telecommunications, both from the University of Genoa, Italy, in 1993 and 1998, respectively.

At present, he is the Scientific Coordinator of the activities on remote sensing image analysis of the Signal Processing and Telecommunications Group in the Department of Biophysical and Electronic Engineering, University of Genoa. His main research contributions are in the area of remote-sensing image processing and recognition. In particular, his interests include feature selection, classification of multisensor and multitemporal images, and change detection in multitemporal images. He conducts and supervises research on these topics within the frameworks of several national and international projects. He is the author (or co-author) of more than 40 scientific publications and a referee for the international journals IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON NEURAL NETWORKS, *Signal Processing*, and *International Journal of Remote Sensing*.

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'98), Seattle, WA, July, 1998. He served on the Scientific Committee of the EUROPTO International Conference on "Signal and Image Processing for Remote Sensing IV" (Barcelona, September 1998). He is a member of the International Association for Pattern Recognition (IAPR).

Diego Fernández Prieto (S'99) received the "Licenciatura" (M.S.) degree in physics from the University of Santiago de Compostela, Spain, in 1994. In 1997, he received the master degree in business administration (M.B.A.) from the University of Deusto, Spain, and the master degree in European business administration (M.E.B.A.) from the University of Kent, U.K.

In 1994 and 1996, he was involved in the activities of the "Istituto per la Matematica Applicata" (I.M.A.) of the National Research Council (C.N.R.), Genoa, Italy, where he studied parallel-computing applications to geographic information system (G.I.S.). At present, he is a Ph.D. student and cooperates with the Signal Processing & Telecommunications Group, Department of Biophysical and Electronic Engineering, University of Genoa. His main research activity is in the area of the remote-sensing image processing and recognition; in particular, his current interests include neural networks for classification purposes, incremental learning, data fusion, and change detection techniques.

Sebastiano B. Serpico (M'87) received the Laurea degree in electronic engineering in 1982 and the Doctorate in telecommunications in 1989 from the University of Genoa, Italy.

Since 1982, he has cooperated with the Department of Biophysical and Electronic Engineering (DIBE) of the University of Genoa in the field of image processing and recognition. His current research interests include the application of pattern recognition (feature selection, classification, and change detection) to remotely sensed images. He was an Assistant Professor of telecommunications from 1990 to 1994 and of pattern recognition from 1995 to 1997. He is currently an Associate Professor of telecommunications, University of Genoa, where he teaches signal theory. Since 1995, he has been the responsible of the Signal Processing and Telecommunications Group (SP&T) of DIBE. He is author (or coauthor) of more than 100 scientific publications, including journals and conferences. He is a Reviewer of the international journals IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Signal Processing*, and *Pattern Recognition Letters*. He was the Chairman of the EUROPTO Conference on Signal and Image Processing for Remote Sensing (Barcelona, Spain, September, 1998).

Dr. Serpico a member of the International Association for Pattern Recognition (IAPR).