An Iterative Technique for the Detection of Land-Cover Transitions in Multitemporal Remote-Sensing Images

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

***Abstract*—We propose a supervised nonparametric technique, based on the "compound classification rule" for minimum error, to detect land-cover transitions between two remote-sensing images acquired at different times. Thanks to a simplifying hypothesis, the compound classification rule is transformed into a form easier to compute. In the obtained rule, an important role is played by the probabilities of transitions, which take into account the temporal dependence between two images. In order to avoid requiring that training sets be representative of all possible types of transitions, we propose an iterative algorithm which allows the probabilities of transitions to be estimated directly from the images under investigation. Experimental results on two Thematic Mapper images confirm that the proposed algorithm may provide remarkably better detection accuracy than the "Post-Classification Comparison" algorithm, which is based on the separate classifications of the two images.**

I. Introduction

DETECTION of land-cover changes is one of the most interesting aspects of the analysis of multitemporal remote-sensing images [1]. In particular, it is very useful in many applications, like land use change analysis, study on shifting cultivation, monitoring of pollution, assessment of burned areas, assessment of deforestation, and so on [2]–[4]. Many of these applications relate to the analysis of large areas on the Earth surface; then, it is important to exploit automatic techniques to detect land-cover changes in order to reduce the effort required by manual image analysis.

A. Previous Work

Usually, change detection involves a couple of spatially registered remote-sensing images acquired on the same ground area at two different times. Two main approaches to detecting land-cover changes can be distinguished: changes can be detected by comparing the spectral reflectances of multitemporal raw satellite images; changes can be detected by using supervised classifiers.

Many change detection algorithms are based on the former approach [2], [5]–[9]. The *Univariate Image Differencing* algorithm [2], [7]–[9], for example, performs change detection by subtracting, on a pixel basis, the images acquired at two times to produce a further image ("difference image"). Under the hypothesis of few changes between the two times, changes can be detected in the tails of the probability density function of the pixel values in the difference image; this technique is usually applied to a single spectral band. Other techniques, like *Vegetation Index Differencing* [2], [5], make the same kind of comparison by using, instead of a spectral band, vegetation indices [1] or other linear (e.g., Tasselled Cap Transformation [1], [8]) or nonlinear combinations of original bands. Also the widely used *Change Vector Analysis* technique [2] exploits an analogous concept. In this case, however, the pixels at each time are represented by their vectors in the feature space. Then, for each couple of pixels, the so called "spectral change vector" is computed as the difference between the feature vectors at the two times. The statistical analysis of the magnitudes of the spectral change vectors allows one to detect the presence of changes, while their directions make it possible to distinguish among different kinds of transitions. Another technique similar to the above-described ones is *Image Ratioing* [2]; in this case, the comparison between spectral bands at two times is performed by computing the ratio, instead of the difference, between images. The techniques based on the *Principal Component Analysis* [2], [6], [9] can be used to perform change detection by applying the principal component transformation separately to the feature space at single times or to the merged feature space at two times. In the first case, change detection is performed according to *Vegetation Index Differencing* by using principal components instead of vegetation indices. In the second case, land-cover changes can be detected by analyzing the minor components of the transformed feature space [2].

The above techniques usually do not aim to identify explicitly what kinds of land-cover transitions have taken place in an area (e.g., the fact that a vegetated area has been urbanized). Only the *Change Vector Analysis* technique allows one to distinguish among different kinds of changes but, being not supervised, it does not explicitly identify the typologies of transitions. Then, the above techniques are suitable for

Manuscript received September 13, 1996; revised March 17, 1997. This research was conducted in part within the framework of the cooperative research project "Integrated assessment of environmental degradation connected with forest fires in European areas," which was supported by the EEC (Environment Programme II, Contract EV5V-CT94-0481), and in part within the framework of the Italian project "Sviluppo di metodi integrati di classificazione agroecologica tramite dati di telerilevamento per la gestione delle risorse naturali," supported by the Italian Space Agency.

The authors are with the Department of Biophysical and Electronic Engineering, University of Genoa, I-16145, Genova, Italy (e-mail: vulcano@dibe.unige.it).

Publisher Item Identifier S 0196-2892(97)04476-8.

applications like, for example, detection of burned areas, detection of pollution, detection of deforestation, and so on. However, they cannot be applied when the only information on the presence of change is not sufficient, like, for example, in the monitoring of shifting cultivation, where it is necessary to recognize the kinds of changes that have taken place in the agricultural area investigated. In addition, the performances of such techniques are generally degraded by several factors (like differences in illumination at two times, differences in atmospheric conditions, in sensor calibration and in ground moisture conditions) that make difficult a direct comparison between raw images acquired at different times.

In order to overcome these problems, one can use the techniques based on a supervised classification of multitemporal images [2]. The simplest technique of this category is *Post-Classification Comparison* [2]. It performs change detection by comparing the classification maps obtained by classifying independently two remote-sensing images of the same area acquired at different times. In this way, it is possible to detect changes and to understand the kinds of transitions that have taken place. Furthermore, the classification of multitemporal images avoids the need to normalize for atmospheric conditions, sensor differences and so on, between the two acquisitions. However, the performances of the *Post-Classification Comparison* technique critically depends on the accuracies of the classification maps. In particular, the final change detection map exhibits an accuracy close to the product of the accuracies yielded at the two times [2]. This is due to the fact that *Post-Classification Comparison* does not take into account the dependence existing between two images of the same area acquired at two different times. *Supervised Direct Multidata Classification* [2] is able to overcome this problem. In this technique, pixels are characterized by a vector obtained by "stacking" the feature vectors related to the images acquired at two times. Then, change detection is performed by considering each transition as a class and by training a classifier to recognize the transitions. Appropriate training sets are required: the training pixels at the two times should be related to the same points on the ground and should represent accurately the proportions of all the transitions in the whole images. Usually, in real applications, it is difficult to have training sets with such characteristics.

In general, the approach based on supervised classification is more flexible than that based on the comparison of multitemporal raw data. In addition to the already mentioned capability to explicitly recognize land-cover transitions and to reduce the effects of different acquisition conditions at two times, it allows one to perform change detection also by using different sensors at two times. This is a useful property when change detection on a large temporal scale has to be performed and available images are provided by different sensors. Finally, by exploiting appropriate nonparametric classifiers, this approach is able to utilize also images of the multisensorial type acquired at both single times.

B. Content of the Paper

In this paper, we present a technique to explicitly identify land-cover transitions in multitemporal remote-sensing images. This technique belongs to the type of approaches based on supervised classification and is designed to reduce the problems of some of the techniques of this type. In particular, with respect to *Post-Classification Comparison*, we are able to take into account the time dependency between the *a priori* probabilities of classes at two times by applying a "compound classification" to the images acquired at two times. With respect to the *Supervised Direct Multidata Classification* technique, we aim to relax the constraints on training sets. To this end, we propose an iterative technique aimed to estimate the probabilities of land-cover transitions among classes not from training sets but directly from the images under analysis.

In order to evaluate the performances of our approach, we selected a multitemporal data set, composed of two Landsat Thematic Mapper images, related to an agricultural area. Experimental results are reported and compared with those provided by the *Post-Classification Comparison* technique.

This paper is organized into five sections. In the next section, we derive the simplified compound classification rule we utilize to perform the detection of land-cover transitions in multitemporal images. In Section III, we indicate how to implement the simplified compound classification rule; in particular, we propose an algorithm to estimate iteratively the probabilities of land-cover transitions. The data set used for experiments is described in Section IV, where experimental results are also reported. In Section V results are discussed and conclusions are drawn.

II. Detection of Land-Cover Transitions by Compound Classification

Two multispectral remote-sensing images acquired at times t_1 and t_2 on the same area on the ground are examined. Let us consider couples of pixels made up of a pixel of the multispectral image acquired at time t_1 and a spatially corresponding pixel of the multispectral image acquired at time t_2; let such pixels be characterized by the d-dimensional feature vectors X_1 and X_2, respectively. Let $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$ be the set of possible land-cover classes at time t_1, and let $N = \{\nu_1, \nu_2, \cdots, \nu_m\}$ be the set of possible land-cover classes at time t_2. A land-cover change in the considered couple of pixels is detected if the two classes ω_i and ν_j, to which such pixels are assigned, are different.

If we disregard contextual information in the spatial domain, i.e., if we classify each couple of pixels independently of any other on the basis only of its feature vectors X_1 and X_2, the optimal classification, in the sense of minimum error probability, is given by the Bayes rule for the case of compound classification problems [10]. Such a rule requires that the couple of classes (ω_i, ν_j) be selected that provides the maximum *a posteriori* probability, given the observed feature vectors X_1 and X_2 [10], [11]:

$$\max_{\omega_i, \nu_j} \left\{ P\left(\frac{\omega_i, \nu_j}{X_1, X_2} \right) \right\}. \tag{1}$$

The couple of classes (ω_i, ν_j) that provides the maximum in (1) is the same that provides the following maxima:

$$\max_{\omega_i,\nu_j}\left\{\frac{p\left(\frac{X_1, X_2}{\omega_i, \nu_j}\right)P\left(\frac{\nu_j}{\omega_i}\right)P(\omega_i)}{p(X_1, X_2)}\right\} \Leftrightarrow \max_{\omega_i,\nu_j}\left\{p\left(\frac{X_1, X_2}{\omega_i, \nu_j}\right)P\left(\frac{\nu_j}{\omega_i}\right)P(\omega_i)\right\} \quad (2)$$

where the term $p(X_1, X_2)$ can be neglected, as it is independent of ω_i and ν_j.

Both (1) and (2) involve the estimations of $n \times m$ functions which are defined in a $(2 \times d)$-dimensional space. These estimations could be carried out by using a set of training pixels ("training set"). Unfortunately, in real situations, it is difficult to have suitable training sets available, as a large number of training pixels for each possible combination of classes ω_i and ν_j are required. In order to simplify the estimation of such functions, we introduce the following hypothesis. Let us consider the feature vector $X_i\,(i = 1, 2)$, related to time t_i, be composed of a signal component S_i and of a noise component N_i, i.e.,

$$X_1 = S_1 + N_1$$

and

$$X_2 = S_2 + N_2 \quad (3)$$

Let us assume that the signal S_i depends only on the land-cover class at time t_i, and that the noise N_i depends only on the land-cover class at time t_i and possibly on S_i (as occurs, for example, for multiplicative noise in SAR images). Under this hypothesis, the probabilistic dependence between the images at the two times derives only from the dependence of the classes at the two times, and one can write

$$p\left(\frac{X_1, X_2}{\omega_i, \nu_j}\right) = p\left(\frac{X_1}{\omega_i}\right)p\left(\frac{X_2}{\nu_j}\right). \quad (4)$$

(An assumption analogous to that defined by (4) was introduced by Swain [11] in the context of multitemporal classification.)

By substituting (4) into (2) and by applying some transformations, we obtain that the following maximum can be used in the decision rule:

$$\max_{\omega_i,\nu_j}\left\{\frac{P\left(\frac{\omega_i}{X_1}\right)P\left(\frac{\nu_j}{X_2}\right)P\left(\frac{\nu_j}{\omega_i}\right)}{P(\nu_j)}\right\}. \quad (5)$$

According to (5), under the above-defined hypothesis, to perform the compound classification of two multitemporal remote-sensing images we need to estimate the *a priori* probabilities $P(\nu_j)$ of the classes at time t_2, the single-date, multivariate, conditional probabilities $P(\omega_i/X_1)$ and $P(\nu_j/X_2)$ at the two times, and the probabilities of transitions $P(\nu_j/\omega_i)$.

III. Iterative Detection of Land-Cover Transitions

The approach we propose to the detection of land-cover transitions is of the supervised nonparametric type, that is, we assume to have a training set at our disposal for each of the image acquired at the two times and we do not hypothesize any model of probability density functions of data. We require that these two training sets be representative of the data in the whole images, in the sense that: i) the number of pixels belonging to each class is approximately proportional to the prior probability of that class; ii) the probability distribution of training data is the same as that of the data in the whole images. It is worth noting that these two assumptions are the same that are usually adopted in the supervised classification of single-date images. The two training sets may refer to different points on the ground.

A. Class Prior Probabilities

In the above assumptions about the training sets, the *a priori* class probabilities $P(\nu_j)$ at t_2 can be easily estimated from the related training set by considering the relative frequency of the pixels of each class:

$$\hat{P}(\nu_j) = \frac{\text{no. of training pixels at } t_2 \text{ that belong to } \nu_j}{\text{total no. of training pixels at } t_2}. \quad (6)$$

B. Posterior Class Probabilities

According to theoretical results reported in [12] and [13], a neural network appropriately trained by using a mean-square-error criterion gives network outputs that, after normalization, approximate posterior class probabilities. In order to estimate $P(\omega_i/X_1)$ and $P(\nu_j/X_2)$, we utilize two different multilayer perceptron neural networks [14], [15] trained with the data sets available at times t_1 and t_2, respectively. Training is carried out by using the backpropagation learning procedure [14].

C. Probabilities of Land-Cover Transitions

We have to estimate the elements of the following $n \times m$ matrix of probabilities of land-cover transitions:

$$L = \begin{bmatrix} P\left(\frac{\nu_1}{\omega_1}\right) & \cdots & P\left(\frac{\nu_j}{\omega_1}\right) & \cdots & P\left(\frac{\nu_m}{\omega_1}\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P\left(\frac{\nu_1}{\omega_i}\right) & \cdots & P\left(\frac{\nu_j}{\omega_i}\right) & \cdots & P\left(\frac{\nu_m}{\omega_i}\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P\left(\frac{\nu_1}{\omega_n}\right) & \cdots & P\left(\frac{\nu_j}{\omega_n}\right) & \cdots & P\left(\frac{\nu_m}{\omega_n}\right) \end{bmatrix} \quad (7)$$

where the element l_{ij} represents the probability that a pixel belonging to the class ω_i at t_1 may belong to the class ν_j at t_2.

In general, it is not possible to estimate the above-mentioned probabilities of land-cover transitions directly from the training

set because in real applications it is difficult to have training sets that verify the following two necessary conditions: i) training pixels at the two dates should correspond to the same points on the ground; ii) the relative frequency of the $n \times m$ land-cover transitions in the training sets should provide a reasonable approximation to the probabilities of land-cover transitions in the whole image. In many real cases, this does not occur. In particular, the latter of the above conditions is not equivalent to the former condition we assumed at the beginning of this section, and it is more difficult that it should be verified by training sets.

Some information about the probabilities of land-cover transitions could be obtained by interviews to experts in the applications or by analyzing historical databases. However, also in this case, it is difficult that, for all the probabilities of land-cover transitions, reasonable estimates may be obtained.

Consequently, we propose to estimate the probabilities of transitions from the whole images under investigation. The problem is that, in order to estimate such probabilities from the images, one needs to detect the transitions that occur inside them, but the proposed technique to detect transitions, in its turn, requires such estimates of probabilities of transitions as inputs. To solve this problem, we propose, in the following, an iterative algorithm which does not require any *a priori* knowledge about probabilities of transitions.

D. The Iterative Compound Classification Algorithm

Initialization: We start by estimating $P(\omega_i/X_1)$ and $P(\nu_j/X_2)$ (e.g., by using two neural networks). Then we compute $\hat{P}(\nu_j)\,(j = 1, \cdots, m)$ from the training set as in (6) and, assuming the independence of each class ν_j at t_2 of each class ω_i at t_1, we initialize the matrix of probabilities of transitions as follows:

$$l_{ij}^{k} = \hat{P}\left(\frac{\nu_j}{\omega_i}\right) = \hat{P}(\nu_j) \qquad (k = 0) \tag{8}$$

where k indicates the number of the iteration.

First Iteration: The above initialization allows the application of the first iteration of the compound classification as defined by (5). As a result, a classification map is obtained for each of the images at the two times; it can be used to compute the first estimate of probabilities of transitions (see (9), shown at the bottom of the page) for $k = 1$ and all ω_i and ν_j.

If the first estimate l_{ij}^{1} is more accurate than the initialization in (8), we expect that it will be used to obtain more accurate classification maps and a more precise detection of land-cover transitions by applying again the compound classification rule. This process can be further iterated.

kth Iteration: In general, at the kth iteration, a couple of updated classification maps can be generated by using the estimate L^{k-1} of the L matrix computed at the $(k-1)$th iteration. $\hat{P}(\omega_i/X_1)$, $\hat{P}(\nu_j/X_2)$, and $\hat{P}(\nu_j)$, which are already computed in the initialization step, are used to this purpose (without any need to compute them again). Then, a new estimate L^k can be derived from the new classification maps.

Stop Criterion: The estimation process stops when, between two iterations, the largest difference among the estimates of all probabilities of transitions is below a selected threshold ε:

$$\max_{i,j} \{|l_{ij}^{k} - l_{ij}^{k-1}|\} < \varepsilon \tag{10}$$

where ε is in the range 0–1 and tunes the estimation accuracy.

The outputs of the algorithm are the land-cover transitions that are detected by comparing the two classification maps generated at the last iteration.

With reference to the classification provided by the first iteration, it is worth noting that the assumption of the hypothesis of independence of *a priori* class probabilities in (8), together with the previous hypothesis on feature vectors we introduced in Section II, makes the compound classification rule in (5) equivalent to the single-time Bayes classification rule applied separately to the two images [as is easy to verify by substituting (8) into (5)]. In other words, (5) becomes equivalent to the independent maximization of $P(\omega_i/X_1)$ and $P(\nu_j/X_2)$.

Concerning convergence, we cannot provide a proof that the estimates obtained by the above algorithm converge to the true values of the probabilities of transitions. Further experiments showed that, if the first estimate is very far from the true value of the matrix L, for example, due to very inaccurate estimates of the posterior class probabilities in (5), the estimates of some elements of L may not converge to true values. However, starting from reasonable first estimates, we always obtained convergence. We note that a similar concept to that applied in our iterative algorithm was applied in *Recursive Contextual Classification* [16]; also in that case, the author left the study of the convergence proof as an open point for further research.

IV. Experimental Results

A. Data Set Description

The study area is located in the delta of the Po River in the north of Italy. The climate in the area is Xerofil with two main rainy seasons: autumn and spring. July is the driest month. The study area is mostly characterized by extended cultivation of wheat, corn, soybean, sugar beet, alpha-alpha, meadows and horticulture.

We considered a section (350 $\times$ 500 pixels) of a couple of multispectral images (Fig. 1) acquired with the Thematic Mapper (TM) scanner, mounted on board of Landsat, in May 1988 (time t_1) and July 1988 (time t_2), respectively. The

$$l_{ij}^{k} = \frac{\text{no. of couples of pixels assigned to class } \omega_i \text{ at } t_1 \text{ and to class } \nu_j \text{ at } t_2}{\text{total no. of couples of pixels assigned to class } \omega_i \text{ at } t_1} \tag{9}$$

(a) (b)

Fig. 1. Data utilized for experiments: channel 4 of the Landsat Thematic Mapper images acquired (a) in May 1988 and (b) in July 1988.

images were registered to an average accuracy on the ground control points of about 0.5 pixel. The ground-truth of the study area was collected for some of the land-cover types during the days of the passages of the Landsat. The available ground truth was used to prepare two thematic maps of the selected test site (one for each date) which were used as reference maps to assess the accuracy of the detection of the land-cover transitions. To this end, we selected couples of pixels corresponding to the same positions on the ground and for which the ground truth was available at both times. Such pixels, which were used to generate both the training and the test sets, belonged to one among three possible classes at time t_1 (i.e., bare soil, wheat, and urban) and five classes at time t_2 (i.e., sugar beet, corn, soybean, bare soil, and urban). Between the two times, the following transitions took place (see Table I): bare soil in May became corn or soybean or sugar beet in July, and wheat in May became bare soil in July; urban areas did not change. To better approximate the independence condition between training and test sets, we subdivided into regions the image areas with ground truth available at both times, then we grouped these regions into two disjoint sets, and we took 7215 training pixels from the regions of one set and 6308 test pixels from the regions of the other set.

TABLE I
LAND-COVER TRANSITIONS OF THE PIXELS OF THE TEST SETS BETWEEN MAY 1988 AND JULY 1988. THE NUMBER OF PIXELS FOR EACH TRANSITION IS GIVEN

Data Class (May 1988)	Data Class (July 1988)	Number of pixels in the test sets
Urban	Urban	810
Bare soil	Corn	387
	Soybean	3322
	Sugar beet	255
Wheat	Bare soil	1534

Each pixel was represented by a vector of six features corresponding to six channels of the TM in the visible and in the infrared spectrum (the thermal channel was disregarded).

B. Results

The performances of the proposed *Iterative Compound Classification* (ICC) technique are assessed in the following and compared with those of the *Post-Classification Comparison* (PCC) technique [2] on the above-described data set.

Concerning the ICC algorithm, we used the training set related to the image acquired at time t_2 to derive the estimations

TABLE II
TRUE VALUE OF THE MATRIX L OF THE PROBABILITIES OF LAND-COVER TRANSITIONS

May 1988 \ July 1988	Urban	Corn	Bare soil	Soybean	Sugar beet
Urban	1.000	0.000	0.000	0.000	0.000
Bare soil	0.000	0.098	0.000	0.838	0.064
Wheat	0.000	0.000	1.000	0.000	0.000

TABLE III
ESTIMATE OF THE VALUE OF THE MATRIX L OF THE PROBABILITIES OF LAND-COVER TRANSITIONS PROVIDED BY THE ICC ALGORITHM AT THE FIRST ITERATION

May 1988 \ July 1988	Urban	Corn	Bare soil	Soybean	Sugar beet
Urban	0.727	0.000	0.159	0.114	0.000
Bare soil	0.137	0.073	0.014	0.718	0.058
Wheat	0.190	0.003	0.768	0.039	0.000

TABLE IV
ESTIMATE OF THE VALUE OF THE MATRIX L OF THE PROBABILITIES OF LAND-COVER TRANSITIONS PROVIDED BY THE ICC ALGORITHM AT THE CONVERGENCE

May 1988 \ July 1988	Urban	Corn	Bare soil	Soybean	Sugar beet
Urban	1.000	0.000	0.000	0.000	0.000
Bare soil	0.016	0.084	0.000	0.831	0.069
Wheat	0.091	0.000	0.895	0.014	0.000

of the *a priori* probabilities of classes $P(\nu_j)$ at that time. Then, we used two different neural networks trained on the training sets related to times t_1 and t_2, respectively, in order to estimate the posterior class probabilities at the two times [i.e., $P(\omega_i/X_1)$ and $P(\nu_j/X_2)$]. Both neural networks were fully-connected multilayer perceptrons with three layers of neurons: an input layer, a hidden layer, and an output layer. In the input layer we had as many neurons as the dimension of the vector of features (i.e., six neurons). The choice of the number of neurons in the hidden layer was made after performing trials with different numbers of neurons. We selected three neurons as, when we increased the number of neurons, the classification accuracies on the training sets did not significantly improve. In the output layer we used as many neurons as the number of data classes, that is, three and five neurons for the networks related to times t_1 and t_2, respectively. Finally, we initialized the matrix L of the probabilities of land-cover transitions with the estimates of the *a priori* probabilities of classes at time t_2 [i.e., $l_{ij}^0 = \hat{P}(\nu_j) \forall i, j$] given by (6). The ICC technique was then run on the test sets to assess performances; a value equal to 0.01 was selected for the threshold ε in (10).

It is interesting to analyze the behavior of the estimates of probabilities of land-cover transitions. To this end, Tables II–IV give the true value of the matrix L of the probabilities of land-cover transitions and the related estimates provided by the ICC at the first iteration and at convergence, respectively. Some of the estimates of probabilities of land-cover transitions obtained at the first iteration of the ICC (Table III) are quite different from the corresponding real values (see, for example, the urban-urban and wheat-bare soil transitions, which exhibit errors of 0.273 and 0.232, respectively). Nonetheless, the estimates of the probabilities of land-cover transitions provided by the ICC technique at convergence are, on an average, close to the true values, with a maximum error of 0.105. In Fig. 2, the diagrams of the behaviors of the estimates of probabilities of land-cover transitions are plotted versus the number of iterations of the ICC algorithm for the five transitions that really occurred. On the considered data set, the convergence was reached in four iterations.

Table V shows the error matrix of the land-cover transitions for the ICC algorithm. This matrix was computed by comparing, for each couple of pixels in the test sets, the classification maps provided by the ICC algorithm at convergence with the corresponding ground truth. In this matrix, we give the true land-cover transitions (as determined from the ground truth) on the rows and the land-cover transitions detected by the considered algorithms on the columns. The terms on the diagonal of this matrix give correctly recognized land-cover transitions, while the other terms give the errors on the recognition of transitions.

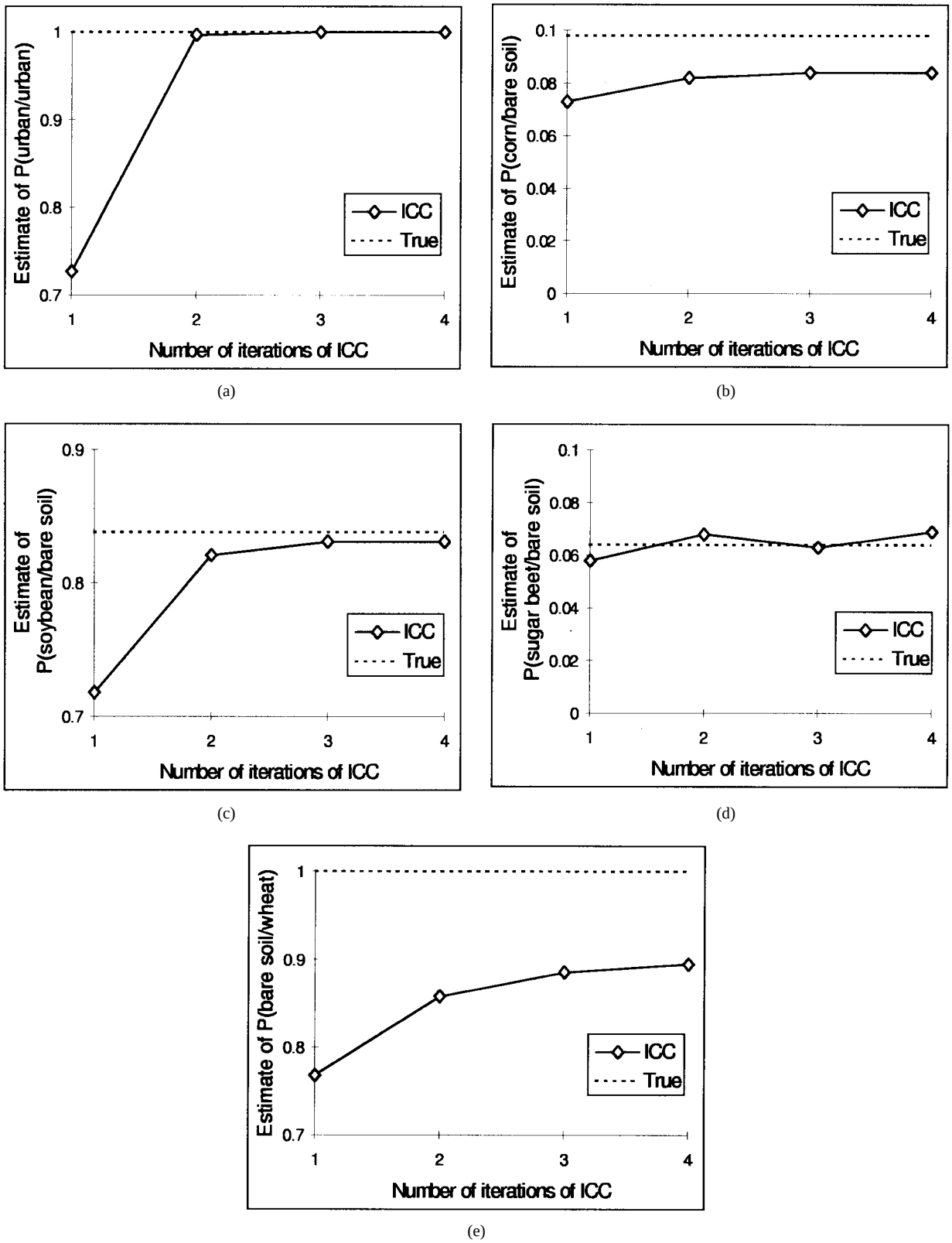

Fig. 2. Diagrams of the behaviors of the estimates of probabilities of land-cover transitions plotted versus the number of iterations of the ICC algorithm. Only the transitions that really occurred are considered: (a) urban–urban, (b) bare soil–corn, (c) bare soil–soybean, (d) bare soil–sugar beet, and (e) wheat–bare soil. Note that different scales are utilized for the vertical axes.

For the comparison with the PCC technique, we needed classification maps obtained by independent classifications of the test sets at the two times. To this end, we utilized the estimates of posterior class probabilities [i.e., $P(\omega_i/X_1)$ and $P(\nu_j/X_2)$] provided by the two neural networks described above; then, we performed the classification (independently at the two times) according to the maximum *a posteriori* probability. By comparing the classification maps at the two

TABLE V

ERROR MATRIX OF THE LAND-COVER TRANSITIONS OF THE PIXELS OF THE TEST SETS FOR THE ICC ALGORITHM. THE TRUE LAND-COVER TRANSITIONS (AS DETERMINED FROM THE GROUND TRUTH) ARE GIVEN IN THE ROWS AND THE LAND-COVER TRANSITIONS DETECTED BY THE ICC ALGORITHM ARE GIVEN IN THE COLUMNS. (MAY 1988: ω_1 = URBAN, ω_2 = BARE SOIL, ω_3 = WHEAT; JULY 1988: ν_1 = URBAN, ν_2 = CORN, ν_3 = BARE SOIL, ν_4 = SOYBEAN, ν_5 = SUGAR BEET)

Ground truth / ICC	$\omega_1\to\nu_1$	$\omega_1\to\nu_2$	$\omega_1\to\nu_3$	$\omega_1\to\nu_4$	$\omega_1\to\nu_5$	$\omega_2\to\nu_1$	$\omega_2\to\nu_2$	$\omega_2\to\nu_3$	$\omega_2\to\nu_4$	$\omega_2\to\nu_5$	$\omega_3\to\nu_1$	$\omega_3\to\nu_2$	$\omega_3\to\nu_3$	$\omega_3\to\nu_4$	$\omega_3\to\nu_5$
$\omega_1\to\nu_1$	**647**	0	0	0	0	0	2	0	0	4	0	0	13	0	0
$\omega_1\to\nu_2$	0	**0**	0	0	0	0	0	0	0	0	0	0	0	0	0
$\omega_1\to\nu_3$	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0	0
$\omega_1\to\nu_4$	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0	0
$\omega_1\to\nu_5$	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	0
$\omega_2\to\nu_1$	57	0	0	0	0	**0**	2	0	4	1	0	0	0	0	0
$\omega_2\to\nu_2$	21	0	0	0	0	0	**302**	0	6	9	0	0	1	0	0
$\omega_2\to\nu_3$	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	0
$\omega_2\to\nu_4$	16	0	0	0	0	0	27	0	**3243**	72	0	0	1	0	0
$\omega_2\to\nu_5$	1	0	0	0	0	0	52	0	58	**169**	0	0	0	0	0
$\omega_3\to\nu_1$	43	0	0	0	0	0	2	0	1	0	**0**	0	99	0	0
$\omega_3\to\nu_2$	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	0
$\omega_3\to\nu_3$	22	0	0	0	0	0	0	0	2	0	0	0	**1408**	0	0
$\omega_3\to\nu_4$	3	0	0	0	0	0	0	0	8	0	0	0	12	**0**	0
$\omega_3\to\nu_5$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**

times with the ground truth, we computed, also for the PCC, the error matrix of land-cover transitions (Table VI).

Comparing the two matrices, it is easy to observe that the ICC algorithm provided a better accuracy than the PCC one. In particular, a notable improvement was obtained for the urban-urban transition. The reason is that, at time t_1 the distribution of the urban class in the feature space strongly overlaps that of the bare-soil class. Therefore, it is very difficult for a classifier to distinguish between these two classes on the basis of the single-date spectral information acquired at t_1. On the contrary, at time t_2, it is easier to separate the urban class from the other classes. Consequently, the PCC exhibits low accuracy on the urban class at t_1 and high accuracy at t_2. The ICC, by jointly utilizing information at the two times, may yield improved classification accuracy for the urban class also at time t_1.

In order to synthesize the differences between the performances provided by the two algorithms, we computed the Kappa coefficients of agreement [17]–[19] (Table VII) related to the two error matrices in Tables V and VI. A considerable improvement in the Kappa coefficient with respect to the PCC (i.e., 0.19) was obtained by using the proposed algorithm.

We can also consider the accuracy of each classification map (related to a single time) independently of the other. At both times, the accuracy provided by the ICC algorithm is better than that provided by the PCC (Table VIII): improvements in the overall classification accuracies of about 10% and 3.4% were obtained at times t_1 and t_2, respectively.

V. CONCLUSIONS

In this paper, we have presented an iterative approach to the detection of land-cover transitions which is based on the compound classification of multitemporal remote-sensing images. The introduction of a simplifying hypothesis about the feature vectors allowed the compound classification rule for the minimum error to be transformed into a form which can be more easily computed. In particular, the dependence between the images is taken into account only in the probability of land-cover transitions between the two times.

The novelty of the proposed approach lies in the iterative technique to estimate the probabilities of land-cover transitions directly from the images under investigation. As a consequence, both in the presentation of the approach and in the description of experimental results, we focused on such an iterative estimation rather than on the other probability estimations required, for which suitable techniques can be found in the literature. In particular, we suggested adopting neural networks to estimate posterior class probabilities at a single time; however, one can utilize different techniques for estimating these probabilities. Furthermore, one may rewrite the compound classification rule as a function of the class-conditional probability density functions of the feature vectors [i.e., $p(X_1/\omega_i)$ and $p(X_2/\nu_j)$] and utilize parametric (e.g., hypothesizing Gaussian distributions) or nonparametric estimation techniques.

Experimental results on the selected real data set confirmed that, even if the dependence between the multitemporal images may be partially lost due to the simplifying hypothesis we introduced, the use of the iterative probability estimates of land-cover transitions inside the classification rule allows a significant improvement in the detection of such transitions with respect to the independent classification of the multitemporal images performed by the *Post-Classification Comparison*.

Like any other classification-based technique, our approach requires that a training set be available for each of the multitemporal images; on the other hand, it also exhibits the

TABLE VI
ERROR MATRIX OF THE LAND-COVER TRANSITIONS OF THE PIXELS OF THE TEST SETS FOR THE PCC ALGORITHM. THE TRUE LAND-COVER TRANSITIONS (AS DETERMINED FROM THE GROUND TRUTH) ARE GIVEN IN THE ROWS AND THE LAND-COVER TRANSITIONS DETECTED BY THE PCC ALGORITHM ARE GIVEN IN THE COLUMNS. (MAY 1988: ω_1 = URBAN, ω_2 = BARE SOIL, ω_3 = WHEAT JULY 1988: ν_1 = URBAN, ν_2 = CORN, ν_3 = BARE SOIL, ν_4 = SOYBEAN, ν_5 = SUGAR BEET)

Ground truth / PCC	$\omega_1\rightarrow\nu_1$	$\omega_1\rightarrow\nu_2$	$\omega_1\rightarrow\nu_3$	$\omega_1\rightarrow\nu_4$	$\omega_1\rightarrow\nu_5$	$\omega_2\rightarrow\nu_1$	$\omega_2\rightarrow\nu_2$	$\omega_2\rightarrow\nu_3$	$\omega_2\rightarrow\nu_4$	$\omega_2\rightarrow\nu_5$	$\omega_3\rightarrow\nu_1$	$\omega_3\rightarrow\nu_2$	$\omega_3\rightarrow\nu_3$	$\omega_3\rightarrow\nu_4$	$\omega_3\rightarrow\nu_5$
$\omega_1\rightarrow\nu_1$	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\omega_1\rightarrow\nu_2$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\omega_1\rightarrow\nu_3$	6	0	0	0	0	0	0	0	0	0	0	0	1	0	0
$\omega_1\rightarrow\nu_4$	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\omega_1\rightarrow\nu_5$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\omega_2\rightarrow\nu_1$	607	0	0	0	0	0	6	0	8	5	0	0	12	0	0
$\omega_2\rightarrow\nu_2$	24	0	0	0	0	0	302	0	5	10	0	0	0	0	0
$\omega_2\rightarrow\nu_3$	59	0	0	0	0	0	1	0	2	0	0	0	2	0	0
$\omega_2\rightarrow\nu_4$	14	0	0	0	0	0	28	0	3213	74	0	0	1	0	0
$\omega_2\rightarrow\nu_5$	1	0	0	0	0	0	46	0	56	165	0	0	0	0	0
$\omega_3\rightarrow\nu_1$	55	0	0	0	0	0	2	0	0	0	0	0	251	0	0
$\omega_3\rightarrow\nu_2$	0	0	0	0	0	0	2	0	0	0	0	0	3	0	0
$\omega_3\rightarrow\nu_3$	2	0	0	0	0	0	0	0	0	0	0	0	1244	0	0
$\omega_3\rightarrow\nu_4$	5	0	0	0	0	0	0	0	38	1	0	0	20	0	0
$\omega_3\rightarrow\nu_5$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

TABLE VII
KAPPA COEFFICIENT OF AGREEMENT RELATED TO THE ERROR MATRICES FOR THE ICC (SEE TABLE V) AND PCC (SEE TABLE VI) ALGORITHMS

Kappa coefficient of agreement	
ICC	**PCC**
0.86	0.67

TABLE VIII
CLASSIFICATION ACCURACIES OF CLASSIFICATION MAPS AT EACH TIME PROVIDED BY THE ICC AND PCC ALGORITHMS

Algorithm	Overall Classification Accuracy (%)	
	May 1988	July 1988
ICC	96.88	93.17
PCC	86.91	89.74

typical advantages of this kind of techniques, which have already been pointed out in the Introduction.

In the context of classification-based methods, the advantage of the proposed approach over to the *Post-Classification Comparison* lies mainly in the higher accuracy it may provide. With respect to the *Supervised Direct Multidate Classification*, in our case the training sets can be generated independently, as they need not refer to the same points on the ground, nor be representative of all possible types of land-cover transitions (as, instead, is required by the *Supervised Direct Multidate Classification*).

Concerning the computational load, the only time-consuming step of the ICC is the initialization, which requires the separate estimation of the posterior class probabilities $P(\omega_i/X_1)$ and $P(\nu_j/X_2)$ (e.g., by neural networks). The other iterations of the ICC utilize the estimation of the posterior probabilities computed in the initialization step and, therefore, they are very fast. Since the PCC also requires the estimation of posterior class probabilities, we may conclude that the proposed approach does not require a significant increase in the computational load with respect to the PCC.

As further developments, we are studying how to make the convergence of the iterative estimates of transition probabilities more reliable. In addition, we are investigating if some of the hypotheses we made on the feature vectors and on the training sets can be removed. For example, we are considering the estimation of prior class probabilities directly from the images under investigation in order to remove the hypothesis that the number of training pixels of each class is proportional to the corresponding class probabilities.

ACKNOWLEDGMENT

The authors wish to thank Dr. F. Roli for his helpful suggestions and comments, and Dr. M.A. Gomarasca (C.N.R. – I.R.R.S. – Telerilevamento, Milan, Italy) for providing the Thematic Mapper data and for assisting in the related agronomic interpretation.

REFERENCES

[1] J. A. Richards, *Remote Sensing Digital Image Analysis*, 2nd ed. New York: Springer-Verlag, 1993.
[2] A. Singh, "Digital change detection techniques using remotely-sensed data," *Int. J. Remote Sensing*, vol. 10, no. 6, pp. 989–1003, 1989.

[3] P. R. Coppin and M. E. Bauer, "Processing of multitemporal Landsat TM imagery to optimize extraction of forest cover change features," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 4, pp. 918–927, 1994.
[4] K. Green, D. Kempka, and L. Lackey, "Using remote sensing to detect and monitor land-cover and land-use change," *Photogramm. Eng. Remote Sens.*, vol. 60, no. 3, pp. 331–337, 1994.
[5] J. R. G. Townshend and C. O. Justice, "Spatial variability of images and the monitoring of changes in the Normalized Difference Vegetation Index," *Int. J. Remote Sensing*, vol. 16, no. 12, pp. 2187–2195, 1995.
[6] T. Fung and E. LeDrew, "Application of principal component analysis to change detection," *Photogramm. Eng. Remote Sens.*, vol. 53, no. 12, pp. 1649–1658, 1987.
[7] P. S. Chavez, Jr., and D. J. MacKinnon, "Automatic detection of vegetation changes in the southwestern united states using remotely sensed images," *Photogramm. Eng. Remote Sens.*, vol. 60, no. 5, pp. 571–583, 1994.
[8] T. Fung, "An assessment of TM imagery for land-cover change detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, no. 4, pp. 681–684, 1990.
[9] D. M. Muchoney and B. N. Haack, "Change detection for monitoring forest defoliation," *Photogramm. Eng. Remote Sens.*, vol. 60, no. 10, pp. 1243–1251, 1994.
[10] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis.* New York: Wiley, 1973.
[11] P. H. Swain, "Bayesian classification in a time-varying environment," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 879–883, 1978.
[12] M. Richard and R. Lippmann, "Neural network classifiers estimate Bayesian a posteriory probabilities," *Neural Computat.*, vol. 3, no. 4, pp. 461–463, 1991.
[13] H. Gish, "A probabilistic approach to the understanding and training of neural network classifiers," in *Proc. 1990 Int. Conf. Acoustic, Speech, Sig. Proc.*, 3–6 Apr. 1990, pp. 1361–1364.
[14] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*. Boston, MA: Addison-Wesley, 1991.
[15] J. D. Paola and R. A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 981–996, July 1995.
[16] K. S. Fu and T. S. Yu, *Statistical Pattern Classification Using Contextual Information.* New York: Research Studies, 1980.
[17] R. G. Congalton, R. G. Oderwald, and R. A. Mead, "Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques," *Photogramm. Eng. Remote Sens.*, vol. 49, no. 12, pp. 1671–1678, 1983.
[18] W. D. Hudson and C. W. Ramm, "Correct formulation of the Kappa coefficient of agreement," *Photogramm. Eng. Remote Sens.*, vol. 53, no. 4, pp. 421–422, 1987.
[19] G. H. Rosenfield and K. Fitzpatrick-Lins, "A coefficient of agreement as a measure of thematic classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 52, no. 2, pp. 223–227, 1986.

Lorenzo Bruzzone (S'95) received the "Laurea" (M.S.) degree in electronic engineering from the University of Genoa, Italy, in 1993. Since 1994, he has been Ph.D. student in Electronics and Computer Science.

He is presently with the Signal Processing and Understanding group at the Department of Biophysical and Electronic Engineering, University of Genoa. His main research activity is in the area of remote-sensing image processing and recognition; in particular, his research interests include: feature selection for classification purposes, classification, and change detection in multitemporal images. In these fields, he conducts and supervises research works within several national and international projects.

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

Sebastiano B. Serpico (M'87) received the "Laurea" (M.S.) degree in electronic engineering and the Ph.D. degree in telecommunications, both from the University of Genoa, Italy, in 1982 and 1989, respectively.

Since 1982, he cooperated with the Department of Biophysical and Electronic Engineering (DIBE) of the University of Genoa in the field of image processing and recognition. At the present, he is doing research work on comparison and integration of different approaches to classification (statistical, knowledge-based, and neural-network approaches). The selected application domain is remote-sensing images. Since 1990, he has been a Research Fellow at DIBE and an assistant for the course of Telecommunications; since 1994, he has been Assistant Professor of the course of Pattern Recognition. He is currently the Manager of the research team on Signal Processing and Understanding at DIBE. He is author (or coauthor) of more than 100 scientific publications and is referee of the international journals IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Signal Processing* and *Pattern Recognition Letters*.

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