Unsupervised Retraining of a Maximum Likelihood Classifier for the Analysis of Multitemporal Remote Sensing Images

Lorenzo Bruzzone and Diego Fernàndez Prieto

***Abstract*—An unsupervised retraining technique for a maximum likelihood (ML) classifier is presented. The proposed technique allows the classifier's parameters, obtained by supervised learning on a specific image, to be updated in a totally unsupervised way on the basis of the distribution of a new image to be classified. This enables the classifier to provide a high accuracy for the new image even when the corresponding training set is not available.**

***Index Terms*—Expectation maximization algorithm, land-cover map updating, maximum likelihood (ML) classification, remote sensing, unsupervised retraining.**

I. Introduction

In the past few years, supervised classification techniques have proven effective tools for automatic generation of land cover maps of extended geographical areas [1]–[5]. The capabilities of such techniques and the frequent availability of remote sensing images, acquired periodically in many regions of the world by spaceborne sensors, make it possible to develop monitoring systems aimed at mapping the land cover classes that characterize specific geographical areas on a regular basis. From an operational point of view, the implementation of a system of this type requires the availability of a suitable training set (and hence of ground truth information) for each new image to be categorized. However, the collection of a reliable ground truth is usually an expensive task in terms of time and economic cost. Consequently, in many cases, it is not possible to rely on training data as frequently as required to ensure an efficient monitoring of the site considered. This is a serious drawback that limits the operational capabilities of the aforementioned monitoring systems.

We face this problem by focusing on an important group of real world applications in which the considered test sites can be assumed to be characterized by fixed sets of land cover classes: only the spatial distributions of such land covers are supposed to vary over time. Examples of such applications include studies on forestry, territorial management, and natural resource monitoring on a national or even continental scale [6]–[8].

In this communication, we present an unsupervised retraining technique for maximum likelihood (ML) classifiers [9] that overcomes the aforementioned drawback of land cover monitoring systems. In particular, the proposed technique allows the existing statistical parameters of an ML classifier (estimated by supervised learning on a specific image) to be updated whenever a new image lacking the corresponding training set has to be analyzed. The unsupervised retraining phase permits the classifier to generate an accurate land cover map from the new image even when the related training set is not available.

II. General Formulation of the Problem

Let us consider an ML classifier for the periodical monitoring of a specific geographical area. Let $\mathbf{X}_1 = \{x_{1,1}, x_{1,2}, \cdots, x_{1,M}\}$ denote a multispectral image consisting of M pixels acquired in the area under analysis at the time, t_1, $x_{1,j}$ being the feature vector associated with the jth pixel of the image. Let $\Omega = \{\omega_1, \omega_2, \cdots, \omega_C\}$ be the set of C land cover classes that characterize the geographical area considered at t_1. Let X_1 be a multivariate random variable that represents the pixel values (i.e., the feature vector values) in $\mathbf{X}_1$. Finally, let us assume that a reliable ground truth (from which a training set $\mathbf{Y}_1$ can be derived) is available at t_1.

In the context of the Bayes decision theory, the decision rule adopted by an ML classifier is expressed as follows [1], [9]:

$$x_{1,j} \in \omega_k, \quad \text{if} \quad \omega_k = \arg\max_{\omega_i \in \Omega} \left\{ \hat{P}_1(\omega_i)\hat{p}_1(x_{1,j}/\omega_i) \right\} \tag{1}$$

where $\hat{P}_1(\omega_i)$ and $\hat{p}_1(X_1/\omega_i)$ are the estimates of the *a priori* probability and of the conditional density function of the class ω_i in the image $\mathbf{X}_1$, respectively. It is worth noting that the subscript 1 is used here to stress the dependencies of both statistical terms on the considered image $\mathbf{X}_1$. According to (1), the training phase of an ML classifier consists in the estimations of the *a priori* probability $P_1(\omega_i)$ and the conditional density $p_1(X_1/\omega_i)$ for each class $\omega_i \in \Omega$. Such estimates can be obtained by using classical supervised approaches that exploit the information that is present in the considered training set $\mathbf{Y}_1$ [1], [9].

Let us now assume that, at the time t_2, a new land cover map of the study area is required. Let $\mathbf{X_2} = \{x_{2,1}, x_{2,2}, \cdots, x_{2,N}\}$ (N may be different from M) be a new image acquired at t_2 in the study area, which is assumed to be characterized by the same set of land cover classes $\Omega = \{\omega_1, \omega_2, \cdots, \omega_C\}$. This means that only the spatial distributions of such classes may change between t_1 and t_2, but no new land covers can be noticed at t_2. Let us also assume that at t_2, the corresponding training set is not available. This prevents the generation of the required land cover map as the training of the classifier cannot be performed [i.e., both the *a priori* probability $P_2(\omega_i)$ and the conditional density function $p_2(X_2/\omega_i)$ of each class $\omega_i \in \Omega$ in the new image $\mathbf{X}_2$ cannot be estimated by traditional supervised approaches]. At the same time, it is not possible to apply the classifier trained on the image $\mathbf{X}_1$ to the image $\mathbf{X}_2$ because, in general, the estimates of statistical class parameters at t_1 do not provide accurate approximations for the same terms at t_2. This is due to several factors (e.g., differences in the atmospheric and light conditions at the image acquisition dates, sensor nonlinearities, different levels of soil moisture, etc.) that alter the spectral signatures of land cover classes in different images and consequently, the distributions of such classes in the feature space.

In this context, we propose an unsupervised retraining technique to derive, for each class $\omega_i \in \Omega$, reliable estimates of both $P_2(\omega_i)$ and $p_2(X_2/\omega_i)$, starting from the current classifier parameters obtained in a supervised way at the time t_1.

III. The Proposed Unsupervised Retraining Technique

The main idea of the proposed technique is that the first approximate estimates of the parameter values that characterize the classes considered at the time t_2 can be obtained by exploiting the classifier's parameters estimated at the time t_1 by supervised learning. In particular, for each class $\omega_i \in \Omega$, the initial values of both the prior probability $P_2^0(\omega_i)$ and the conditional density function $p_2^0(X_2/\omega_i)$ can be approximated by

$$P_2^0(\omega_i) = \hat{P}_1(\omega_i); \qquad p_2^0(X_2/\omega_i) = \hat{p}_1(X_1/\omega_i). \tag{2}$$

As already pointed out in the previous section, generally, such first estimates do not provide accurate approximations for the statistical pa-

Manuscript received January 21, 2000; revised April 6, 2000. This work was supported by the Italian Space Agency, Ispra, Italy.

The authors are with the Department of Civil and Environmental Engineering, University of Trento, Trento, Italy.

Publisher Item Identifier S 0196-2892(01)01158-5.

rameters of the classes at t_2. Therefore, we suggest improving such rough estimates by exploiting the information associated with the distribution $p_2(X_2)$ of the new image $\mathbf{X}_2$. In particular, the proposed method is based on the observation that the statistical distribution of the pixel values in $\mathbf{X}_2$ can be described by a mixed density distribution with as many components as the classes to be recognized

$$p_2(X_2) = \sum_{i=1}^{C} P_2(\omega_i) p_2(X_2/\omega_i) \tag{3}$$

where the mixing parameters and the component densities are the *a priori* probabilities and the conditional density functions of the classes, respectively [10]. In this context, the retraining of the ML classifier at the time t_2 becomes a mixture density estimation problem. In our case, this problem involves the estimation of the parameter vector $\theta_2 = [\theta_{2,1}, P_2(\omega_1), \theta_{2,2}, P_2(\omega_2), \cdots, \theta_{2,C}, P_2(\omega_C)]$, where each component $\theta_{2,i}$ represents the vector of parameters that characterizes the density function $p_2(X_2/\omega_i)$, which, given its dependence on $\theta_{2,i}$, can be rewritten as $p_2(X_2/\omega_i, \theta_{2,i})$. The components of θ_2 can be estimated by maximizing a log-likelihood function $\ell(\mathbf{X}_2/\boldsymbol{\theta}_2)$ defined as

$$\ell(\mathbf{X}_2/\boldsymbol{\theta}_2) = \sum_{j=1}^{N} \log \left\{ \sum_{i=1}^{C} P_2(\omega_i/\boldsymbol{\theta}_2) p_2(x_{2,j}/\omega_i, \boldsymbol{\theta}_2) \right\}. \tag{4}$$

The expectation maximization (EM) algorithm [11]–[13] is one of the most powerful solutions to this type of problem. It consists of two main steps: an expectation step and a maximization step. Both steps are iterated so that, at each iteration, the estimated parameters provide an increase in the log-likelihood function $\ell(\mathbf{X}_2/\boldsymbol{\theta}_2)$ until a local maximum is reached.

To further explain the proposed approach, let us consider, for simplicity, the case in which all classes included in Ω can be described by Gaussian distributions. In this context, the density function associated with each class ω_i at t_2 can be completely described by the mean vector $\mu_{2,i}$ and the covariance matrix $\Sigma_{2,i}$. Therefore, the vector of parameters to be estimated becomes

$$\theta_2 = [\mu_{2,1}, \Sigma_{2,1}, P_2(\omega_1), \cdots, \mu_{2,C}, \Sigma_{2,C}, P_2(\omega_C)]. \tag{5}$$

It can be proven that the equations for estimating the statistical terms associated with a generic class ω_i are the following [11]–[13]:

$$P_2^{s+1}(\omega_i) = \frac{\displaystyle\sum_{x_{2,j} \in \mathbf{X_2}} \frac{P_2^s(\omega_i) p_2^s(x_{2,j}/\omega_i)}{p^s(x_{2,j})}}{N} \tag{6}$$

$$[\mu_{2,i}]^{s+1} = \frac{\displaystyle\sum_{x_{2,j} \in \mathbf{X_2}} \frac{P_2^s(\omega_i) p_2^s(x_{2,j}/\omega_i)}{p_2^s(x_{2,j})} x_{2,j}}{\displaystyle\sum_{x_{2,j} \in \mathbf{X_2}} \frac{P_2^s(\omega_i) p_2^s(x_{2,j}/\omega_i)}{p_2^s(x_{2,j})}} \tag{7}$$

$$[\Sigma_{2,i}]^{s+1} = \frac{\displaystyle\sum_{x_{2,j} \in \mathbf{X_2}} \frac{P_2^s(\omega_i) p_2^s(x_{2,j}/\omega_i)}{p_2^s(x_{2,j})} \left\{x_{2,j} - [\mu_{2,i}]^{s+1}\right\}^2}{\displaystyle\sum_{x_{2,j} \in \mathbf{X_2}} \frac{P_2^s(\omega_i) p_2^s(x_{2,j}/\omega_i)}{p_2^s(x_{2,j})}} \tag{8}$$

where the superscripts s and $s+1$ refer to the values of the parameters at the current and next iterations, respectively. The estimates are obtained starting from the initial values of the considered parameters and iterating the above equations up to convergence.

An important aspect of the EM algorithm concerns its convergence properties. Even though convergence can be ensured, it is impossible to guarantee that the algorithm will converge to the global maximum of the log-likelihood function (only in few specific cases is it possible to ensure the convergence to the global maximum). A detailed description of the EM algorithm and its theoretical aspects is beyond the scope of this paper. We refer the reader to the literature for a more detailed analysis of such an algorithm and its properties [11], [13].

The estimates obtained for each class $\omega_i \in \Omega$ at convergence (i.e., $P_2^{conv}(\omega_i)$, $\mu_{2,i}^{conv}$ and $\Sigma_{2,i}^{conv}$) are the new parameters of the ML classifier at the time t_2, i.e., $\hat{P}_2(\omega_i) = P_2^{conv}(\omega_i)$; $\hat{\mu}_{2,i} = \mu_{2,i}^{conv}$; $\hat{\Sigma}_{2,i} = \Sigma_{2,i}^{conv}$.

At this point, a land cover map of the analyzed area at the time t_2 can be generated by labeling each pixel $x_{2,j}$ in accordance with the ML decision rule:

$$x_{2,j} \in \omega_k, \quad \text{if} \quad \omega_k = \arg\max_{\omega_i \in \Omega} \left\{ \hat{P}_2(\omega_i) \hat{p}_2(x_{2,j}/\omega_i) \right\}. \tag{9}$$

IV. Experimental Results

Different experiments were carried out on a data set made up of two multispectral images acquired by the Thematic Mapper (TM) sensor of the Landsat 5 satellite. The selected test site was a section (412 × 382 pixels) of a scene showing Lake Mulargias on the Island of Sardinia, Italy. The two images used in the experiments were acquired in September 1995 (t_1) and July 1996 (t_2). Fig. 1 shows channels 5 of both images. The available ground truth was used to derive a training set and a test set for each image (see Table I). In particular, five land cover classes that characterized the test site at the above dates (i.e., urban area, vineyard, forest, pasture, water) were identified. It is worth noting that the images were acquired in slightly different periods of the year. This involves some differences in the spectral responses of the classes in the two images. Such differences are mainly due to the different light conditions at the image acquisition dates and to the different levels of growth of the vegetation shown in the images. Therefore, the unsupervised retraining problem turned out to be rather complex.

To carry out the experiments, we assumed that only the training set associated with the image acquired in September 1995 was available. The training set for the July 1996 image was only used to evaluate the performances of the proposed technique trough a comparison with a supervised approach.

The ML classifier was trained (in a supervised way) on the September 1995 image to estimate the *a priori* probabilities and the parameters that characterize the density functions of the classes at the time t_1. The assumption of normal distributions was made for the density functions of the classes (this was a reasonable assumption, as we considered TM images). After training, the effectiveness of the classifier was evaluated on the test sets for both images. The classification accuracies obtained are given in Table II. On the one hand, as expected, the classifier provided a high classification accuracy (90.97%) for the test set related to the September 1995 image. On the other hand, it exhibited very poor performances for the July 1996 test set. In particular, the overall classification accuracy for the July test set was equal to 50.43%, which cannot be considered an acceptable result.

In order to better understand the behavior of the classifier, in Fig. 2(a) and (b), the distributions (in the feature space) of subsets of training samples corresponding to the September 1995 and July 1996 images are represented, respectively. Bands 2 and 5 were selected for the representation, as they provide the best separation of classes in the feature space (a feature-selection process based on the Jeffreys–Matusita distance [1], [14] was applied to identify such bands). In greater detail, Fig. 2(a) shows the estimates of the distributions of the classes

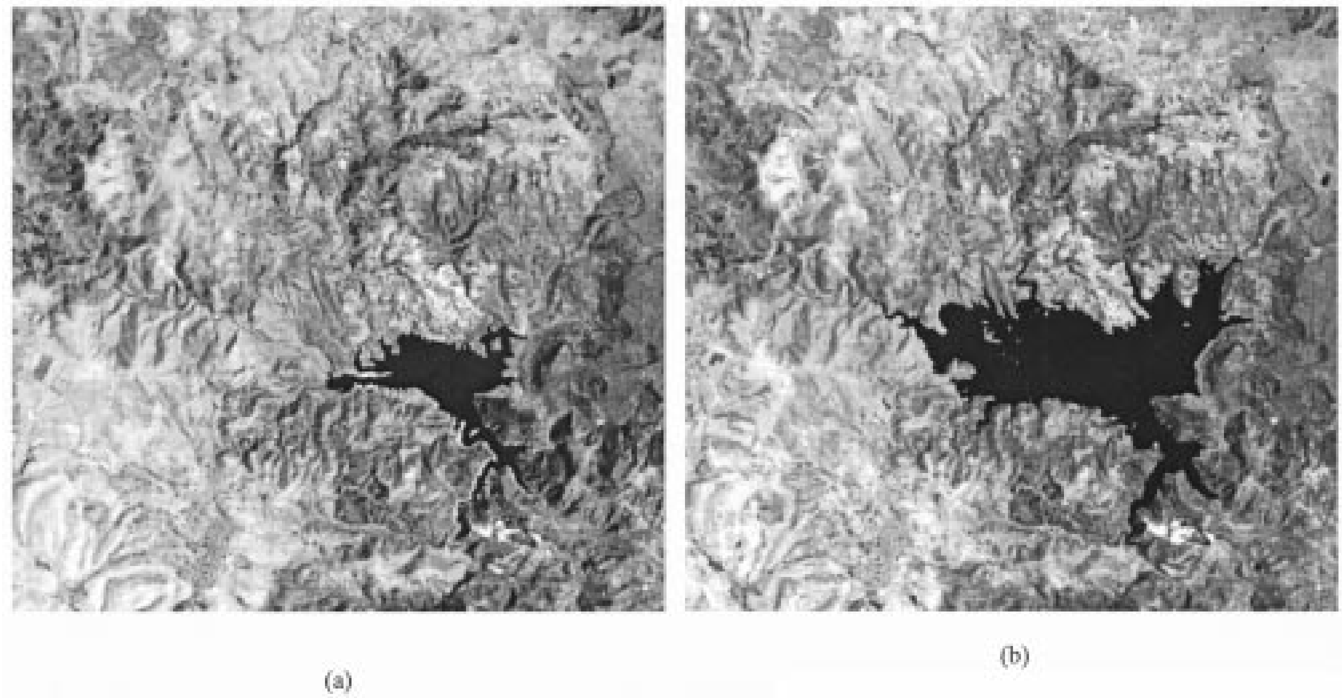


Fig. 1. Bands 5 of the Landsat-5 TM images utilized for the experiments. (a) Image acquired in September 1995 and (b) image acquired in July 1996.

TABLE I
THE TRAINING AND THE TEST SETS USED FOR THE EXPERIMENTS CARRIED OUT ON THE IMAGES ACQUIRED IN SEPTEMBER 1995 AND JULY 1996

Land-cover classes	Number of patterns (September 1995)		Number of patterns (July 1996)	
	Training set	Test set	Training set	Test set
Pasture	554	589	554	589
Forest	304	274	304	274
Urban area	408	418	408	418
Water body	804	551	804	551
Vineyard	179	117	179	117
Overall	**2249**	**1949**	**2249**	**1949**

TABLE II
CLASSIFICATION ACCURACIES OBTAINED FOR THE SEPTEMBER 1995 AND JULY 1996 TEST SETS BY A CLASSICAL SUPERVISED ML CLASSIFIER TRAINED ON THE SEPTEMBER 1995 TRAINING SET

Land-cover class	Classification accuracy (%)	
	September 1995 test set	July 1996 test set
Pasture	82.51	19.52
Forest	97.44	95.62
Urban area	94.73	90.43
Water body	100.00	36.11
Vineyard	62.39	24.78
Overall	**90.97**	**50.43**

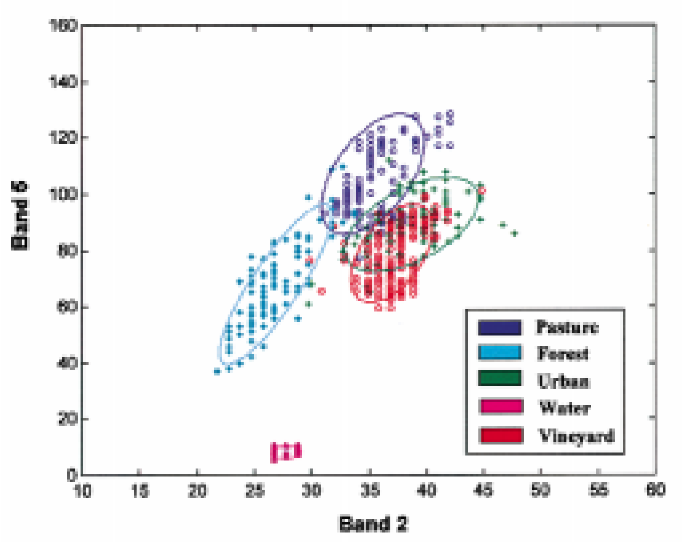
obtained by supervised learning at the time t_1 (the contours of the estimated Gaussian conditional densities are represented for a value equal to 3σ) and the distribution of a subset of training samples of the t_1 image in the feature space. Analogously, Fig. 2(b) presents the aforementioned contours obtained by supervised learning at t_1 compared to the distribution of a subset of training samples of the classes at t_2. A comparison of the two figures points out the intrinsic complexity of the faced problem. These differences explain the low accuracy obtained for the July 1996 test set by the ML classifier trained on the September 1995 image.

At this point, the proposed technique was applied to the t_2 image (July 1996) in order to compute, in an unsupervised way, the new estimates of the *a priori* probabilities and density function parameters of the considered land cover classes. The parameters of the ML classifier trained on the t_1 image (September 1995) were exploited to initialize the EM algorithm. At the end of the iterative process, the resulting estimates were associated with the new parameters of the ML classifier. In order to evaluate the accuracies of the new estimates, the classifier was tested again on the July 1996 test set. For the sake of comparison, a supervised ML classifier was trained and subsequently tested on the July 1996 image by using the classical approach (i.e., exploiting the training set for a supervised parameter estimation). The results obtained are given in Table III. As one can see, the classification accuracy provided by the proposed classifier for the July test set increased by about 42%, compared to the one exhibited by the classifier trained on the September image (92.76% versus 50.43%). It is worth noting that this improvement was shared by most of the considered classes. A comparison with the supervised ML classifier trained and tested on the July image showed that such a classifier provided an overall accuracy (92.66%) very similar to the one yielded by the proposed technique (92.76%).

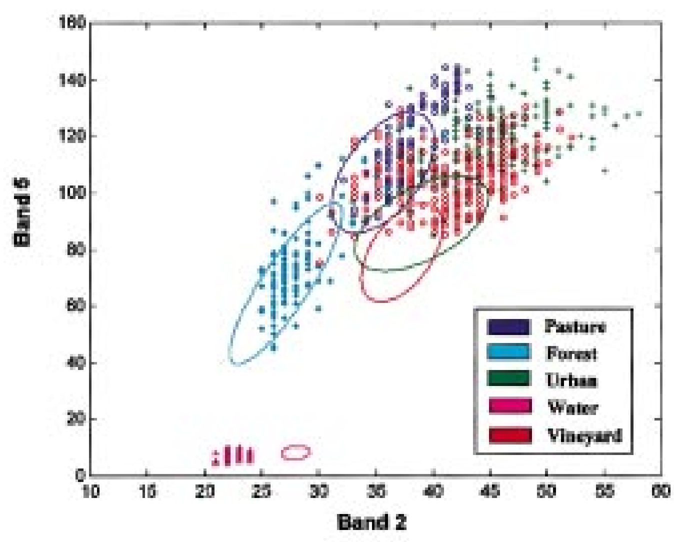
If we analyze the situation in the feature space after the unsupervised retraining [see Fig. 2(c)], we can observe that the new estimates achieved by the proposed unsupervised technique fit rather accurately the distributions of the training samples in the t_2 image.

A further insight into the behavior of the proposed method is provided by Fig. 3, where the trend of the overall classification accuracy versus the number of EM-algorithm iterations is plotted. As can be seen, the overall classification accuracy increases significantly from 50.43% (i.e., for the initial estimates) to 88.19% in only ten iterations and reaches the final value of 92.76% in 23 iterations.

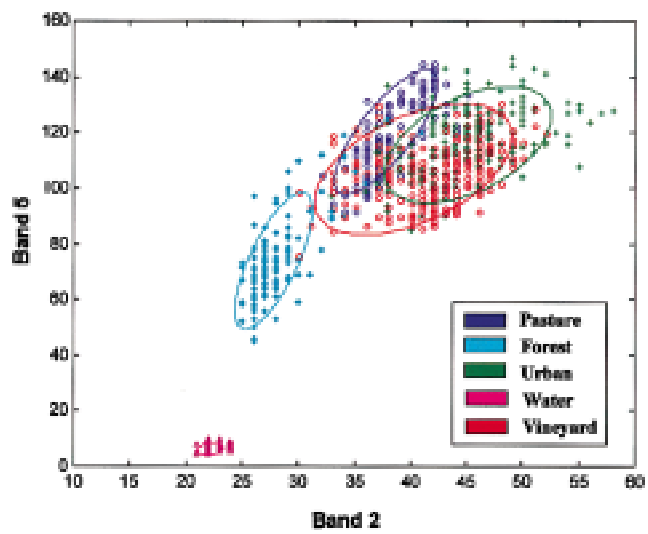
As far as the computational load is concerned, in our experiments, carried out on a Sun Workstation Ultra-Sparc 80, the time taken by the algorithm to reach convergence (in 23 iterations) was equal to 478 s. This seems rather a reasonable time, considering both the complexity of the problem and the high accuracy of the results obtained.



(a)



(b)



(c)

Fig. 2. Comparisons between the distributions of classes in the images considered (represented by a subset of samples randomly extracted from the corresponding training set) and their estimates (the contours of the estimated Gaussian conditional densities are represented for a value equal to 3σ). (a) Distributions of classes in the September 1995 image and their estimates achieved by supervised learning carried out on the September 1995 training set and (b) distributions of classes in the July 1996 image and their estimates achieved by supervised learning carried out on the September 1995. (c) Distributions of classes in the July 1996 image and their final estimates achieved by the proposed unsupervised technique applied to the July 1996 image.

V. Discussion and Conclusions

An unsupervised retraining technique for ML classifiers has been presented that constitutes a useful support for remote sensing monitoring systems based on multitemporal images. The proposed technique allows the generation of accurate land cover maps of a specific study area also from images for which a reliable ground truth (hence a suitable training set) is not available. This is made possible by an unsupervised updating of the existing classifier's parameters (obtained by classical supervised training on a specific image) on the basis of the distribution of the new image to be classified.

The proposed technique can be used in particular applications in which the areas of interest are characterized by the same land cover classes over time. This means that only the spatial distributions of land covers are assumed to change over time.

Our approach exploits the capabilities of the EM algorithm to estimate the prior probabilities and conditional density functions of classes at the time t_2 on the basis of both the estimates performed at the time t_1 and the distribution $p(X_2)$ of the new image in the feature space. It is worth noting that this approach does not consider the spatial distributions of classes in the new scene but only the distributions of patterns in the feature space. This results in a robust behavior with respect to land cover changes that may occur in the analyzed area between the two times considered.

The presented method is based on the assumption that the estimates of the distributions of classes derived from a supervised training on a previous image of the considered area can represent rough estimates of the class distributions in the new image to be categorized. Then the EM algorithm is applied in order to improve such estimates iteratively on the basis of the global density function of the new image. It is worth noting that when the initial estimates are very different from the true ones (e.g., when the considered image has been acquired under atmospheric or light conditions very different from the ones in the image exploited for the supervised initial training of the classifier), the EM algorithm may lead to inaccurate final values. Therefore, in order to overcome this problem, we recommend the application of a suitable preprocessing phase aimed at reducing the main differences between images due to the aforementioned factors (simple relative calibration techniques, which do not require any atmospheric data, can be adopted [15], [16]). In addition, the sequence of images to be classified should be acquired in similar periods of the year, as the spectral responses of the related land covers (hence, the corresponding distributions in the feature space) may significantly change in different seasons.

Experiments carried out on different multitemporal data sets confirmed the validity of the proposed technique (for brevity, only the results obtained on a data set have been reported in this communication). In particular, they pointed out the capability of the proposed technique to update, in a fully unsupervised way, the classifier's parameters in order to match the statistical class distributions of the new images to be classified. Consequently, the resulting classifiers revealed were very effective and attained high classification accuracies for the new images without relying on the corresponding training sets. It is worth noting that, despite some differences in the atmospheric and light conditions

TABLE III
CLASSIFICATION ACCURACIES OBTAINED FOR THE JULY 1996 TEST SET BY THE PROPOSED CLASSIFIER RETRAINED ON THE JULY 1996 IMAGE. FOR THE SAKE OF COMPARISON, THE CLASSIFICATION ACCURACIES ACHIEVED BY A CLASSICAL SUPERVISED ML CLASSIFIER TRAINED AND TESTED ON THE JULY 1996 IMAGE ARE ALSO GIVEN

Land-cover class	Classification accuracy (%)	
	Proposed unsupervised retraining technique	Classical ML classifier trained on the July training set
Pasture	94.06	92.02
Forest	87.22	92.70
Urban area	93.06	93.30
Water body	100.00	100.00
Vineyard	64.10	58.97
Overall	**92.76**	**92.66**

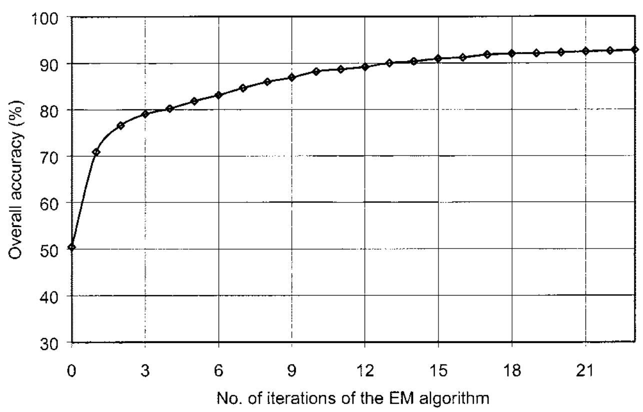


Fig. 3. Classification accuracy for the July 1996 test set versus number of iterations of the EM algorithm.

between images, in all the experiments carried out, we obtained high accuracies without applying any preprocessing technique to the data.

ACKNOWLEDGMENT

The authors wish to thank the European Space Agency (ESA-ESRIN), Frascati, Italy, for providing the images used in the experiments. They are also grateful to the anonymous reviewers for their constructive criticism.

REFERENCES

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

[2] 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, 1990.

[3] L. Bruzzone, D. Fernández Prieto, and S. B. Serpico, "A neural statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 1350–1359, May 1999.

[4] L. Bruzzone and D. Fernández Prieto, "A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 1179–1184, Mar. 1999.

[5] L. Bruzzone, "An approach to feature selection and classification of remote sensing images based on the Bayes rule for minimum cost," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 429–438, Jan. 2000.

[6] F. Maselli, M. A. Gilabert, and C. Conese, "Integration of high and low resolution NDVI data for monitoring vegetation in mediterranean environments," *Remote Sens. Environ.*, vol. 63, pp. 208–218, 1998.

[7] A. Grignetti, R. Salvatori, R. Casacchia, and F. Manes, "Mediterranean vegetation analysis by multi-temporal satellite sensor data," *Int. J. Remote Sensing*, vol. 18, no. 6, pp. 1307–1318, 1997.

[8] M. A. Friedl, C. E. Brodley, and A. H. Strahler, "Maximizing land cover accuracies produced by decision trees at continental to global scales," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 969–977, Mar. 1999.

[9] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974.

[10] B. Jeon and D. Landgrebe, "Partially supervised classification using weighted unsupervised clustering," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 1073–1079, Mar. 1999.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[12] B. M. Shahshahani and D. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 1087–1095, Sept. 1994.

[13] T. K. Moon, "The expectation-maximization algorithm," *Signal Processing Mag.*, vol. 13, no. 6, pp. 47–60, 1996.

[14] L. Bruzzone, F. Roli, and S. B. Serpico, "An extension of the Jeffreys-Matusita distance to multiclass cases for feature selection," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 1318–1321, Nov. 1995.

[15] P. S. Chavez, Jr., "Radiometric calibration of landsat thematic mapper multispectral images," *Photogramm. Eng. Remote Sens.*, vol. 55, pp. 1285–1294, Sept. 1989.

[16] 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, pp. 1285–1294, May 1994.