A partially unsupervised cascade classifier for the analysis of multitemporal remote-sensing images

Lorenzo Bruzzone [*], Diego Fernández Prieto

Department of Information and Communication Technologies, University of Trento, Via Sommarive 14, Trento I-38050, Italy

Abstract

A partially unsupervised approach to the classification of multitemporal remote-sensing images is presented. Such an approach allows the automatic classification of a remote-sensing image for which training data are not available, drawing on the information derived from an image acquired in the same area at a previous time. In particular, the proposed technique is based on a cascade-classifier approach and on a specific formulation of the expectation-maximization (EM) algorithm used for the unsupervised estimation of the statistical parameters of the image to be classified. The results of experiments carried out on a multitemporal data set confirm the validity of the proposed approach. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Multitemporal classification; Cascade classifier; Unsupervised parameter estimation; Remote-sensing

1. Introduction

In the past few years, there has been a growing interest in the use of remote-sensing systems for a regular monitoring of the earth's surface. In this context, images acquired on the same area at different times (i.e., multitemporal images) represent a valuable source of information for the observation of the temporal behavior of the land-cover classes that characterize a given region of interest. From an operational point of view, the monitoring process can be carried out by applying specific supervised classification techniques to the analysis of multitemporal data (Swain, 1978; Kalayeh and Landgrebe, 1986; Khazenie and Crawford, 1990; Jeon and Landgrebe, 1992, 1999; Bruzzone et al., 1999). Unlike standard algorithms for the classification of single-date images, such techniques exploit the temporal correlation between images in order to increase the classification accuracy (a description of these techniques is provided in (Bruzzone et al., 1999)).

A problem arising from the above-mentioned supervised approaches to multitemporal remote-sensing data analysis is that, in general, they require the availability of suitable training data for each image to be categorized. Unfortunately, in most applications, this requirement is not satisfied. In fact, gathering a sufficient number of training samples for each specific image considered, by

[*] Corresponding author. Tel.: +39-0461-882056; fax: +39-0461-882671.

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

either photo-interpretation or the collection of "ground truth" information, is very expensive in terms of time and economic cost. Therefore, in many cases, it is not possible to rely on training data for each single image in a multitemporal data set. This prevents the generation of the corresponding land-cover maps by supervised approaches and, consequently, may affect the accurate and efficient monitoring of the considered site.

In this paper, we propose a partially unsupervised approach to the classification of temporal series of multispectral images that overcomes the aforesaid problem. In particular, it allows the automatic classification of a remote-sensing image, for which training data are not available, drawing on the information derived from earlier observations. The authors already faced this problem in (Bruzzone and Fernàndez Prieto, 2001), where an unsupervised method for the retraining of a maximum-likelihood (ML) classifier was described. The present work is an extension to such a method. In particular, the proposed technique, unlike the previous one, makes use of a cascade-classifier approach to the categorization of multitemporal remote-sensing images, thus allowing the exploitation of the temporal correlation between successive scenes. This approach is based on a specific formulation of the expectation-maximization (EM) algorithm (Dempster et al., 1977) in terms of the joint density function of pairs of sequential images. This formulation of the EM algorithm allows the unsupervised estimation of both the class-conditional density functions in the

second-date image (for which training data are not available) and the prior joint probabilities of classes in the two images considered. An interesting peculiarity of the proposed technique lies in the capability to include in the estimation process additional prior information (if available) about the possible land-cover transitions occurred in the area of interest between the considered dates; this may result in a more robust estimation procedure.

2. General formulation of the problem.

Let $X_1 = \{x_1^1, x_2^1, \ldots, x_{I \times J}^1\}$ and $X_2 = \{x_1^2, x_2^2, \ldots, x_{I \times J}^2\}$ denote two coregistered multispectral images (of dimensions $I \times J$) acquired in the same geographical area (area of interest) at two different times, t_1 and t_2, respectively (Table 1 provides a summary of the notations used in the paper). Let x_j^i be the feature vector associated with the jth pixel of the t_i ($i = 1, 2$) image, and let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_C\}$ be the set of C land-cover classes that characterize the geographical area. In developing our approach, we make two important assumptions, which are considered in several approaches to multitemporal classification (Jeon and Landgrebe, 1999; Solberg, 1999). One implies that the land-cover classes present in the area of interest are the same at the two different times (only the spatial distributions of such classes may change over time). The other implies that the two images are acquired in similar periods of the year in order to avoid incoherent responses from the correspond-

Table 1
Legend of notations used in this paper

Symbol	Description
X_i	Image acquired at the time t_i
X_i	Multivariate random variable that represents the pixel values in X_i
$I \times J$	Dimensions of the selected images
x_j^i	Feature vector associated with the jth pixel of X_i
Ω	Set of land-cover classes that characterize the region of interest
ω_i	ith land-cover class
T_1	Training set available for time t_1
M_2	Required land-cover map at time t_2
l_j^2	Classification label of the jth pixel at time t_2
$p(x_j^i|\omega_k)$	Value of the conditional probability density function for pixel x_j^i, given the class ω_k
$P(\omega_n, \omega_k)$	Prior joint probability of the pair of classes (ω_n, ω_k)

ing types of surface covers. In fact, land-cover classes, especially vegetation classes, may present different spectral behaviors depending on the particular season considered.

Let us assume that only the training set T_1, corresponding to the first image X_1, is available. We aim at classifying the area of interest at the time t_2 by exploiting the information derived from the previous observations X_1 and the corresponding training set T_1. This process involves the generation of a land-cover map $M_2 = \{l_1^2, l_2^2, \ldots, l_{I \times J}^2\}$, where $l_j^2 \in \Omega$ is the classification label of the jth pixel at the time t_2.

We face this problem by applying the Bayes rule for the minimum error to a "cascade classifier" (Swain, 1978), i.e.

$l_j^2 = \omega_m \in \Omega$ if and only if

$$P(\omega_m | x_j^1, x_j^2) = \max_{\omega_k \in \Omega} \left\{ P\left(\omega_k | x_j^1, x_j^2\right) \right\}, \tag{1}$$

where $P(\omega_k | x_j^1, x_j^2)$ is the value of the probability that the jth pixel belongs to class ω_k at t_2, given the observations x_j^1 and x_j^2. Under the conventional assumption of class-conditional independence (Swain, 1978; Bruzzone et al., 1999), the above decision rule can be rewritten as:

$l_j^2 = \omega_m \in \Omega$ if and only if

$$\sum_{n=1}^{C} p\left(x_j^1 | \omega_n\right) p\left(x_j^2 | \omega_m\right) P(\omega_n, \omega_m)$$
$$= \max_{\omega_k \in \Omega} \left\{ \sum_{n=1}^{C} p\left(x_j^1 | \omega_n\right) p\left(x_j^2 | \omega_k\right) P(\omega_n, \omega_k) \right\}, \tag{2}$$

where $p(x_j^i | \omega_k)$ is the value of the conditional density function for the pixel x_j^i, given the class $\omega_k \in \Omega$, and $P(\omega_n, \omega_k)$ is the prior joint probability of the pair of classes (ω_n, ω_k). The latter term takes into account the temporal correlation between the two images.

On the basis of the previous expression (2), the classification of X_2 involves the estimation of the class-conditional densities at time t_1, the class-conditional densities at time t_2, and the prior joint probability for each possible pair of classes. These estimates cannot be obtained by using classical

supervised techniques, as the lack of training data for the second image X_2 prevents the conventional application of such techniques. In this context, we adopt a partially unsupervised approach to the estimation of such probabilistic terms. On the one hand, the class-conditional densities at time t_1 can be estimated from the available training set T_1 by using a supervised approach to density estimation problems (Duda and Hart, 1973). On the other hand, an unsupervised approach based on the EM algorithm (Dempster et al., 1977) is proposed for the estimation of the remaining terms: the class-conditional densities at time t_2 and the prior joint probabilities of classes.

3. The proposed partially unsupervised estimation procedure

The proposed estimation procedure is based on the observation that, under the assumption of class-conditional independence over time, the joint density function of images X_1 and X_2 (i.e., $p(X_1, X_2)$) can be described as a mixture density with $C \times C$ components (as many components as possible pairs of classes):

$$p(X_1, X_2) = \sum_{n=1}^{C} \sum_{m=1}^{C} p(X_1, X_2 | \omega_n, \omega_m) P(\omega_n, \omega_m)$$
$$\cong \sum_{n=1}^{C} \sum_{m=1}^{C} p(X_1 | \omega_n) p(X_2 | \omega_m) P(\omega_n, \omega_m), \tag{3}$$

where X_1 and X_2 are two multivariate random variables that represent the pixel values (i.e., the feature vector values) in X_1 and X_2, respectively. In this context, the estimation of the above terms becomes a mixture density estimation problem, which can be solved by applying the EM algorithm (Dempster et al., 1977; Moon, 1996; Shahshahani and Langrebe, 1994; Bruzzone et al., 1999). In particular, we propose an iterative technique based on a specific formulation of the EM algorithm in terms of the prior joint probabilities of classes in the two images considered. This formulation allows one to derive accurate estimates of both the

class-conditional densities of classes $p(X_2|\omega_m)$ at time t_2 and the prior joint probabilities $P(\omega_n, \omega_m)$.

In order to further explain the presented technique, let us assume that the probability density function of each class can be described by a Gaussian distribution (i.e. by a mean vector μ and a covariance matrix Σ). Under this common assumption (widely adopted for multispectral image classification problems), the estimation of the above-mentioned statistical terms involves the computation of the following vector parameter ϑ:

$$\vartheta = [\mu_1^2, \Sigma_1^2, \ldots, \mu_C^2, \Sigma_C^2, P(\omega_1, \omega_1), P(\omega_1, \omega_2), \ldots,$$
$$P(\omega_C, \omega_{C-1}), P(\omega_C, \omega_C)]. \tag{4}$$

In this context, it can be proved that the recursive equations to be used in order to estimate the required parameters are (Dempster et al., 1977):

$$[\mu_m^2]^{t+1} = \frac{\sum_{j=1}^{I \times J} \left\{ \sum_{n=1}^{C} P^t(\omega_n, \omega_m | x_j^1, x_j^2) \right\} x_j^2}{\sum_{j=1}^{I \times J} \left\{ \sum_{n=1}^{C} P^t(\omega_n, \omega_m | x_j^1, x_j^2) \right\}}, \tag{5}$$

$$[\Sigma_m^2]^{t+1} = \frac{\sum_{j=1}^{I \times J} \left\{ \sum_{n=1}^{C} P^t(\omega_n, \omega_m | x_j^1, x_j^2) \right\} ||x_j^2 - [\mu_m^2]^t||^2}{\sum_{j=1}^{I \times J} \left\{ \sum_{n=1}^{C} P^t(\omega_n, \omega_m | x_j^1, x_j^2) \right\}}, \tag{6}$$

$$P(\omega_n, \omega_m)^{t+1} = \frac{\sum_{j=1}^{I \times J} P^t(\omega_n, \omega_m | x_j^1, x_j^2)}{I \times J}, \tag{7}$$

where, the superscripts t and $t+1$ refer to the values of the parameters at the current and next iterations, respectively. Under the adopted assumption of class-conditional independence in the time domain, the term $P^t(\omega_n, \omega_m | x_j^1, x_j^2)$ can be expressed as

$$P^t(\omega_n, \omega_m | x_j^1, x_j^2)$$
$$= \frac{p(x_j^1|\omega_n) p^t(x_j^2|\omega_m) P^t(\omega_n, \omega_m)}{\sum_{n=1}^{C} \sum_{m=1}^{C} p(x_j^1|\omega_n) p^t(x_j^2|\omega_m) P^t(\omega_n, \omega_m)}. \tag{8}$$

It is worth noting that the terms associated with the class-conditional densities $p(X_1|\omega_n)$ at time t_1 do not present any superscript, as their values are estimated by using a classical supervised procedure and remain fixed during the estimation process.

It is possible to prove (Dempster et al., 1977) that, at each iteration, the estimated parameters evolve from their initial values, thus providing an increase in the log-likelihood function $L(X_1, X_2|\vartheta)$:

$$L(X_1, X_2|\vartheta)$$
$$= \sum_{j=1}^{I \times J} \log \left\{ \sum_{n=1}^{C} \sum_{m=1}^{C} p(x_j^1|\omega_n) p(x_j^2|\omega_m) P(\omega_n, \omega_m) \right\} \tag{9}$$

until a local maximum is reached. The estimates of the parameters obtained at convergence and those achieved by the classical supervised procedure are then substituted into (2) in order to derive the required set of pixel labels M_2.

Concerning the initialization of the considered statistical terms, we adopt the following strategies. The initial values of the parameters that characterize the class-conditional densities at time t_2 are obtained by exploiting the corresponding values estimated at time t_1 by supervised learning, as proposed in (Bruzzone and Fernàndez Prieto, 2001). Concerning the prior joint probabilities, two possible initialization strategies can be followed depending on the prior knowledge available. One strategy can be used in the cases where no prior knowledge exists concerning the possible land-cover transitions that occurred in the area of interest between the two dates considered. In particular, this strategy assigns equal probabilities to each pair of classes. The other strategy can be adopted when the end-user relies on prior information about the temporal evolution of some land-cover classes. Such information can be translated into probabilistic terms by determining if some of the possible land-cover transitions are likely to have occurred between the two considered dates (e.g., in many cases, urban areas do not change to a forest during short and medium-term time periods; this means that $P(\text{urban}, \text{forest}) = 0$). In particular, we can impose explicit constraints on the permitted values of the prior joint probabilities of classes $P(\omega_n, \omega_m)$. Such constraints may be integrated in the proposed estimation process by fixing the values of the related prior joint probabilities on the basis of the existing knowledge. Then these values remain constant

during the iterative estimation process; only the values of the prior joint probabilities for which no information is available are allowed to vary. This results in a more robust and more accurate estimation process.

4. Experimental results

In order to assess the effectiveness of the proposed technique, different experiments were carried out on a data set composed 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 including Lake Mulargias on the Island of Sardinia, Italy. The two images used in the experiments were acquired in September 1995 (t_1) and in 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 2). In particular, five classes of interest (i.e., pasture, forest, urban area, water body, and vineyard), which characterize the study area over time, were identified. To carry out the experiments, we assumed that only the training set associated with the image acquired in September 1995 was available.

We applied the proposed cascade-classifier approach to the above-described images in order to

Table 2
Number of patterns in the training and test sets of both the September 1995 and July 1996 images

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

analyze the July 1996 image by using the estimates of the statistical parameters obtained for the September 1995 image (thanks to the available September 1995 training set), but without using the July 1996 training set. In our experiments, the assumption of normal distribution was made for the density functions of the classes (this was a reasonable assumption, as we considered TM images).

The mean values and the covariance matrices of the Gaussian density functions of the classes at t_1 were computed by using the related training set. Concerning the parameters of the density functions of the classes at t_2 and the prior joint probabilities of the classes, they were estimated in an unsupervised way by using the proposed formulation of the iterative EM algorithm. In the first experiment, the parameters of the density
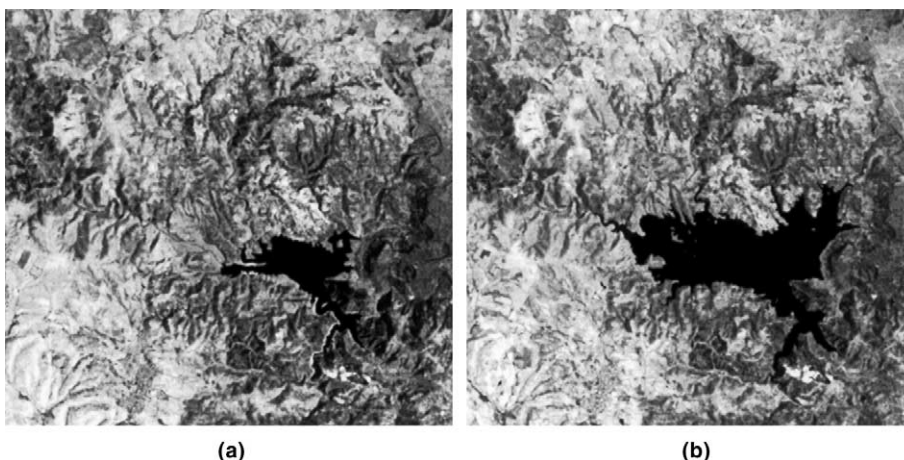

Fig. 1. Bands 5 of the Landsat-5 TM images utilized for the experiments: (a) image acquired in September 1995; (b) image acquired in July 1996.

functions of the classes at t_2 were initialized with the values achieved at t_1, whereas the values of the prior joint probabilities of the classes were assumed to be the same (no prior information on the land-cover transitions were used in this experiment). The EM algorithm converged in 11 iterations (258 s on a Sun Workstation Ultra-Sparc 80). At the end of the iterative process, the resulting estimates were used to perform the classification of the July 1996 image. The results obtained are shown in Table 3. As one can see, we obtained a high overall classification accuracy at t_2 (i.e., 91.48%), even though the training set was not used. The value of the coefficient of accuracy K (i.e., 0.88) also confirms the effectiveness of the presented technique.

In order to better understand the capabilities of the proposed approach, we carried out a second experiment in which we exploited some a priori information in the initialization of the prior joint probabilities of the classes and, consequently, in the estimation process. In particular, as in the considered region no changes affected the vineyard class between the two dates, we assumed $P(\text{vineyard, vineyard}) = P(\text{vineyard})$, whereas all the remaining prior joint probabilities related to the vineyard class were fixed to zero (e.g., $P(\text{vineyard, urban}) = 0$, $P(\text{urban, vineyard}) = 0$, etc.). The results obtained by using the above-mentioned prior information (see Table 3), show a slight increase in the overall classification accuracy (i.e., about 1%) and a sharp increase in the clas-

sification accuracy of the vineyard class (i.e., about 4.3%). In addition, the value of the coefficient of accuracy K increased significantly from 0.88 to 0.90. This behaviour is also confirmed by a comparison with the accuracies provided by the previous version of the method (Bruzzone and Fernàndez Prieto, 2001), where no prior knowledge was used. For example, with respect to that method, an increase in the classification accuracy of the vineyard class of about 2.6% was obtained (see Bruzzone and Fernàndez Prieto, 2001).

A further insight into the capabilities of the proposed technique can be derived from Tables 4(a) and (b), where the confusion matrices that resulted from the aforementioned experiments are shown. In the matrices, the true land-cover classes (determined according to the ground truth) are given in the rows, and the land-cover classes identified with the proposed technique are given in the columns. Therefore, the values on the diagonals of such matrices represent correctly recognized land-cover classes, while the other values represent errors on the recognition of the classes. As one can see, the use of the prior information about the vineyard class (i.e., $P(\text{vineyard, vineyard}) = P(\text{vineyard})$) in the estimation process resulted in a significant reduction in the omission errors for such a class.

In the third experiment, for the sake of comparison, a supervised ML classifier was trained and subsequently tested on the July 1996 image by the classical approach (i.e., by using a training set for the supervised parameter estimation). The results obtained are presented in Tables 5 and 6, where the classification errors and the confusion matrix are given, respectively. As one can see, the overall classification accuracies achieved by the proposed approach on the July test set (91.48% and 92.51% in Table 3) are very similar to that yielded by the supervised ML classifier (92.66% in Table 5). In greater detail, the accuracies are comparable for all the classes. In addition, the value of the coefficient of accuracy K obtained by the proposed technique when the prior knowledge was used (i.e., 0.90) was equal to the value obtained by the classical supervised approach.

Finally, a further experiment was carried out to assess the accuracies of the initial estimates of the

Table 3
Classification accuracies obtained by using the proposed technique with two different initialization strategies for the joint probabilities of classes: (a) equal probabilities; (b) prior knowledge of the vineyard class

Land-cover class	Classification accuracy (%)	
	Equal probabilities	Using prior knowledge
Pasture	83.53	88.28
Forest	97.44	97.44
Urban area	95.70	92.58
Water body	100.00	100.00
Vineyard	62.38	66.67
Overall	91.48	92.51

Table 4
Confusion matrices that resulted from the classification of the July 1996 test set by using the proposed technique with two different intialization strategies for the joint probabilities of classes

	Pasture	Forest	Urban area	Water body	Vineyard
(a) Equal probabilities					
Pasture	*492*	12	85	0	0
Forest	2	*267*	2	0	3
Urban area	5	5	*400*	0	8
Water body	0	0	0	*551*	0
Vineyard	23	11	10	0	*73*
(b) Prior knowledge of the vineyard class					
Pasture	*520*	13	56	0	0
Forest	2	*267*	2	0	3
Urban area	7	7	*387*	0	17
Water body	0	0	0	*551*	0
Vineyard	22	9	8	0	*78*

Table 5
Classification accuracies obtained on the July 1996 test set by using a classical supervised ML classifier trained with the July 1996 training set

Land-cover class	Classification accuracy (%)
Pasture	92.02
Forest	92.70
Urban area	93.30
Water body	100.00
Vineyard	58.97
Overall	92.66

Table 7
Classification accuracies obtained for the July 1996 test set by using a classical supervised ML classifier trained on the September 1995 training set

Land-cover class	Classification accuracy (%)
Pasture	19.52
Forest	95.62
Urban area	90.43
Water body	36.11
Vineyard	24.78
Overall	50.43

class-conditional densities at time t_2. (According to our approach, such estimates were assumed to be equal to the corresponding values obtained for the t_1 image.) To this end, a classical ML classifier was trained on the t_1 image (i.e., the September 1995 image). After training, the effectiveness of the classifier was evaluated on the test set of the t_2 image (i.e., the July 1996 image). The classification

accuracies obtained (see Table 7) were very low. In particular, the overall classification accuracy for the July test set was equal to 50.43%, which cannot be considered an acceptable result. The meaning of this result is twofold. On the one hand, it highlights the implicit difficulty of the multitemporal data set considered, as the estimates on the September training set proved unsuitable to providing

Table 6
Confusion matrix that resulted from the classification of the July 1996 test set by using a classical supervised ML classifier trained on the July 1996 training set

	Pasture	Forest	Urban area	Water body	Vineyard
Pasture	*542*	26	19	0	2
Forest	16	*254*	1	0	3
Urban area	11	2	*390*	0	15
Water body	0	0	0	*551*	0
Vineyard	36	3	9	0	*69*

acceptable classification accuracies for the July test set. On the other hand, it confirms the capability of the proposed approach to iteratively improve the accuracies of the estimates of the class-conditional densities at time t_2.

5. Discussion and conclusions

In this paper, a partially unsupervised approach to the classification of multitemporal remote-sensing images has been presented. Such an approach allows the classification of a remote-sensing image acquired in a specific geographical area at a given time, in the cases where training data are not available. The classification is performed using the statistical parameters estimated for an image acquired in the same area before the one under analysis.

The proposed method is based on a cascade-classifier approach and on a specific formulation of the EM algorithm. The iterative EM algorithm allows the unsupervised estimation of both the prior joint probabilities of classes and the density functions of classes at time t_2 on the basis of the available information: i.e., the known density functions of classes at time t_1 (derived from the available training set) and the joint density function $p(X_1, X_2)$ of the two images considered. In addition, the proposed technique allows the exploitation of the prior information (if available) about the possible land-cover transitions that occurred in the area of interest between the two considered times; this may increase the robustness of the classification procedure.

It is worth noting that, in some cases, the values of the parameters that characterize the density functions of classes at time t_1 may not provide accurate approximations for the same terms at time t_2. This may depend on differences in atmospheric and light conditions that alter the spectral signatures of land-cover classes in different images and consequently the distributions of the classes in the feature space. Such differences may lead to the use of initialization values of the parameters of the density functions of classes at t_2 significantly different from the true values. Therefore, in these cases, before applying the proposed approach, we recommend performing a simple pre-processing phase aimed at reducing, if possible, the effects of the aforesaid differences. This may provide better starting points for the estimation procedure.

Experiments carried out on a multitemporal data set confirmed the validity of the proposed technique. In particular, they pointed out its capability to accurately estimate the class-conditional densities at time t_2 as well as the prior joint probabilities of classes. Consequently, the adopted classifier proved very effective and attained high classification accuracies for the second image, without relying on the corresponding training set.

Acknowledgements

References

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

Bruzzone, L., Fernàndez Prieto, D., 2001. Unsupervised retraining of a maximum-likelihood classifier for the analysis of multitemporal remote-sensing images. IEEE Trans. Geosci. Remote Sensing 39 (2), 456–460.

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

Duda, R.O., Hart, P.E., 1973. Pattern Classification and Scene Analysis. Wiley, New York.

Jeon, B., Landgrebe, D.A., 1992. Classification with spatio-temporal interpixel class dependency contexts. IEEE Trans. Geosci. Remote Sensing 30 (4), 663–672.

Jeon, B., Landgrebe, D.A., 1999. Data fusion approach to multitemporal classification. IEEE Trans. Geosci. Remote Sensing 37 (3), 1227–1233.

Kalayeh, H.M., Landgrebe, D.A., 1986. Utilizing multitemporal data by a stochastic model. IEEE Trans. Geosci. Remote Sensing GE-24 (5), 792–795.

Khazenie, N., Crawford, M.M., 1990. Spatial-temporal autocorrelated model for contextual classification. IEEE Trans. Geosci. Remote Sensing 28 (4), 529–539.

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

Solberg, A.H.S., 1999. Contextual data fusion applied to forest map revision. IEEE Trans. Geosci. Remote Sensing 37 (3), 1234–1243.

Shahshahani, B.M., Langrebe, D.A., 1994. The effect of unlabeled samples in reducing the small size problem and mitigating the Hughes phenomenon. IEEE Trans. Geosci. Remote Sensing 32 (5), 1087–1095.

Swain, P.H., 1978. Bayesian classification in a time-varying environment. IEEE Trans. Geosci. Remote Sensing SMC-8 (12), 880–883.