A Context-Sensitive Clustering Technique Based on Graph-Cut Initialization and Expectation-Maximization Algorithm

Mayank Tyagi, Francesca Bovolo, *Member, IEEE*, Ankit K. Mehra, Subhasis Chaudhuri, and Lorenzo Bruzzone, *Senior Member, IEEE*

***Abstract*—This letter presents a multistage clustering technique for unsupervised classification that is based on the following: 1) a graph-cut procedure to produce initial segments that are made up of pixels with similar spatial and spectral properties; 2) a fuzzy c-means algorithm to group these segments into a fixed number of classes; 3) a proper implementation of the expectation-maximization (EM) algorithm to estimate the statistical parameters of classes on the basis of the initial seeds that are achieved at convergence by the fuzzy c-means algorithm; and 4) the Bayes rule for minimum error to perform the final classification on the basis of the distributions that are estimated with the EM algorithm. Experimental results confirm the effectiveness of the proposed technique.**

***Index Terms*—Clustering, expectation-maximization (EM) algorithm, remote sensing, segmentation, unsupervised classification.**

I. Introduction

IN THE pattern recognition literature, two main approaches to the classification problem have been proposed: the supervised and the unsupervised ones [1]. Supervised classification methods require the availability of a suitable ground truth (i.e., a training set) for the learning of the classifier. Unsupervised techniques (which are based on clustering algorithms) perform classification by considering only the information that is contained in the data, without requiring any training set. Generally, the supervised methods offer a higher accuracy compared to the unsupervised ones. However, in many applications, the training information is not available, as the collection of ground-truth information is a complex and expensive process. In these cases, it is mandatory to adopt unsupervised algorithms.

In remote-sensing applications, clustering techniques, like the fuzzy c-means and K-means [2], [3], have been widely used, owing to their reasonable tradeoff between classification accuracy and simplicity. However, other more complex procedures have been considered (like the ISODATA[1] algorithm [4]) or proposed to capture the properties of multispectral remote-sensing images [3]–[7]. Among these techniques, we recall the methods that are described in [7] and [8], which integrate the spatial-contextual information of the image in the clustering procedure, in order to properly model the spatial properties of the scene. For example, in [7], a method which adopts Markov random fields in modeling the spatial-context information is presented. However, despite the fact that the method can provide good accuracy, its results are insensitive to the geometrical details that are present in the scene or to small homogeneous regions due to the use of a neighborhood system of fixed size and shape.

In this letter, we propose a simple yet effective clustering technique that produces classification maps by the following: 1) considering the spatial-context information and 2) preserving the small areas and details present in the scene. This technique merges the advantages of segmentation, hierarchical methods, and expectation-maximization (EM) algorithm within a Bayesian framework.

II. Problem Formulation

Let $\mathbf{X} = \{\mathbf{x}_{1,1}, \mathbf{x}_{1,2}, \ldots, \mathbf{x}_{R,S}\}$ denote a multispectral image that is composed of $R \cdot S$ pixels, and $\mathbf{x}_{r,s}$ is the d-variate feature vector (where d is the number of spectral bands) that is associated with the pixel at position (r, s) of the image $\mathbf{X}$. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ be the set of M land-cover classes that characterize the geographical area that is represented in the image $\mathbf{X}$. In this letter, for simplicity, the value of M is assumed to be known. However, like any other clustering technique, the method can be applied by assuming that the value of M is unknown and by adopting validation criteria that are present in the literature to identify its best value [9].

In the context of the Bayes classifier [10], the optimal decision rule for the minimum error is given by

$$\mathbf{x}_{r,s} \in \omega_m \Leftrightarrow \omega_m = \arg\max_{\omega_i \in \Omega} \{P(\omega_i) p(\mathbf{x}_{r,s}|\omega_i)\} \qquad (1)$$

where $P(\omega_i)$ and $p(\mathbf{x}_{r,s}|\omega_i)$ are the estimates of the prior probability and of the conditional density function (for the pixel $\mathbf{x}_{r,s}$) of class ω_i, respectively. The training phase of the Bayes classifier consists in the estimation of the *a priori* probability and the conditional density function for each class $\omega_i \in \Omega$. These estimates are typically achieved by using supervised approaches that exploit the information that is included in the training set that is associated with the considered image. However, in our case, we propose to use the Bayesian decision rule in the assumption that no training data are available. Thus, the probabilities should be estimated from the data in an unsupervised way.

Manuscript received October 23, 2006; revised May 29, 2007.

M. Tyagi, A. K. Mehra, and S. Chaudhuri are with the Electrical Engineering Department, IIT-Bombay, Mumbai 400 076, India (e-mail: sc@ee.iitb.ac.in).

F. Bovolo and L. Bruzzone are with the Department of Information Technology, University of Trento, 38100 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

Digital Object Identifier 10.1109/LGRS.2007.905119

[1] Iterative Self-Organizing Data Analysis Technique.

1545-598X/$25.00 © 2007 IEEE

III. Proposed Context-Sensitive Clustering Technique

The proposed unsupervised classifier is a hybrid technique that works at the pixel level, but also exploits the spatial information that is included in homogeneous regions that are present in the image. It is made up of four steps: 1) a graph-cut technique [11] to cluster pixels having similar spectral properties and to obtain a set of initial segments that properly model the geometrical detail information that is present in the considered scene; 2) a fuzzy c-means clustering [2] to partition these segments into a given number M of land-cover classes; 3) an estimation module based on the EM algorithm [12], [13] to improve the estimations of the class parameters resulting from the fuzzy c-means; and 4) a decision module that generates the classification map on the basis of the Bayesian optimal rule. In the following, each step is described in a detailed way.

A. Step 1: Graph-Cut-Based Context-Sensitive Initialization

In the first step, the proposed classifier applies the graph-cut-based segmentation technique [11] to $\mathbf{X}$ to obtain an over-segmentation G of the image (it is worth noting that over segmentation of $\mathbf{X}$ is acceptable at this stage of the process). This segmentation G consists of different regions C_k, such that each C_k is a connected subgraph of the original graph. The use of the graph-cut-based segmentation method comes from its simplicity and low computational time. This is an important constraint as we do not need to optimize the segmentation, but just to seize some context information to initialize, in a robust way, the statistical representation of the problem for the next steps of the procedure.

The graph theoretic approaches to image segmentation model the image $\mathbf{X}$ as an undirected graph $G : (V, E)$, where a generic image pixel in spatial position (r, s) has a corresponding vertex $v_{r,s} \in V$, and an edge $e_{r,s;t,u} \in E$ that connects it to its neighboring pixel (vertex) in the spatial position (t, u). Each edge $e_{r,s;t,u} \in E$ has a weight $w(e_{r,s;t,u})$ based on the measure of similarity between the two considered pixels. Since the purpose of the initial segmentation is to get a quick over segmentation of the images without any emphasis on the spectral signatures of classes, each band is modeled as a separate graph and individually segmented. The results that are obtained on individual spectral channels are merged together by means of a logical OR operator to produce the final segmentation of $\mathbf{X}$. An appropriate modeling of the spectral signature of each class is demanded to the fuzzy c-means clustering and the EM steps. As the image is modeled as a graph, the feature vector $\mathbf{x}_{r,s}$ that is associated with image pixels is used to calculate the edge weights $w(e_{r,s;t,u})$. The main objective of this approach is to find a segmentation G of $\mathbf{X}$, i.e., a partition of V into components such that each region (or component) $C_k \in G$ corresponds to a connected region in the segmentation $G : (V, E')$, where $E' \subset E$. In other words, a segmentation of the image is induced by a subset of edges in E. In general, it is desired to have similar elements in a component and dissimilar elements in different components. This implies that the edges between two vertices of the same component should have relatively lower weights than the edges between two vertices of different components.

There exist many approaches in achieving such a segmentation. One of the faster and more accurate algorithms was proposed in [11]. This algorithm requires similarity and dissimilarity between the pixels to be quantitatively defined according to a proper criterion. First, the internal variation $\mathrm{Int}(C_k)$ of a component $C_k \subset G$ is defined as the maximum edge weight present in any minimum spanning tree of the set of edges which compose a region, i.e.,

$$\mathrm{Int}(C_k) = \max_{e_{r,s;t,u} \in \mathrm{MST}(C_k,E)} \{w(e_{r,s;t,u})\} \tag{2}$$

where $\mathrm{MST}(C_k, E)$ denotes the minimum spanning tree of the component that is built with respect to the set of edges E. This is a measureof similarity within a region. Similarly, the external variation $\mathrm{Ext}(C_i, C_j)$ between two regions $C_i, C_j \subset G$ is defined as the smallest weighted edge connecting them, i.e.,

$$\mathrm{Ext}(C_1, C_2) = \min_{\substack{v_{r,s} \in C_i \\ v_{t,u} \in C_j}} \{w(e_{r,s;t,u})\}. \tag{3}$$

Given the similarity measures, the algorithm checks whether there is any evidence for a boundary between a pair of components. This is accomplished by comparing the external variation between components $\mathrm{Ext}(C_i, C_j)$ with the internal variation of both components $\mathrm{Int}(C_i)$ and $\mathrm{Int}(C_j)$. A boundary is detected if $\mathrm{Ext}(C_i, C_j)$ is greater than at least one between $\mathrm{Int}(C_i)$ and $\mathrm{Int}(C_j)$. This can be defined as

$$\mathrm{Ext}(C_i, C_j) \geq \mathrm{MInt}(C_i, C_j) \tag{4}$$

where the minimum internal difference MInt is defined as

$$\mathrm{MInt}(C_i, C_j) = \min\left(\mathrm{Int}(C_i) + \frac{K}{|C_i|}, \mathrm{Int}(C_j) + \frac{K}{|C_j|}\right). \tag{5}$$

Here, $|C_k|$ denotes the size of component C_k, and K is a nonnegative constant. The threshold value $K/|C_k|$ controls the degree of which the external variation between the components should be greater than their internal variations, such that only similar pixels are put in any component. In greater detail, as K increases, the segmentation process results in a lower number of large segments; on the opposite, as K decreases, the segmentation leads to a high number of small segments. In this letter, K is empirically set to 5000 in order to obtain a reasonable over segmentation of the considered image.

This algorithm obtains a segmentation $G = (C_1, C_2, \ldots, C_N)$ by performing the following iterations.

1) Sort E on the basis of nondecreasing edge weights.
2) Start with a segmentation G^0, where each vertex $v_{r,s}$ is considered a separate component.
3) For $q = 1, 2, \ldots, |E|$ (i.e., the number of components in the edge set E), construct G^q given G^{q-1} as follows. Let C_i^{q-1} be the component of G^{q-1} containing $v_{r,s}$ and C_j^{q-1} the component containing $v_{t,u}$. If $C_i^{q-1} \neq C_j^{q-1}$ and $w(e_{r,s;t,u}) \leq \mathrm{MInt}(C_i^{q-1}, C_j^{q-1})$, then G^q is obtained from G^{q-1} by merging C_i^{q-1} and C_j^{q-1}. Otherwise, $G^q = G^{q-1}$.
4) Return $G = G^{|E|}$.

After the completion of the first step, the number of image regions is always greater than the number of land cover classes M characterizing the given classification problem. At this stage, all the components of the over segmentation G are individually labeled with different labels, i.e., all segments are considered belonging to different classes. In order to reduce the number of segment labels to M, segments with similar statistical properties are merged by applying the fuzzy c-means clustering algorithm.

B. Step 2: Fuzzy c-Means Algorithm

At the end of Step 1, the final segmented image is composed of N regions $C_1, C_2, \ldots, C_N$, where $N > M$. In order to group the N regions in M clusters, the fuzzy c-means algorithm is used [2]. Let $\{\mu_1, \mu_2, \ldots, \mu_N\}$ denote the mean vectors of segments that are obtained at the previous step, and $\{\mu_1, \mu_2, \ldots, \mu_M\}$ denotes the mean vectors of the resulting clusters. Our purpose is to find the optimum matrix U of the membership u_{ij} of segment i to class j (with respect to M classes) by minimizing the following objective function:

$$J(U,S) = \sum_{j=1}^{M}\sum_{i=1}^{N}(u_{ij})^m d^2(C_i, S_j)$$

$$\text{given} \quad \sum_{j=1}^{M} u_{ij} = 1$$

$$\text{and} \quad u_{ij} \in [0,1] \quad \forall \quad j = 1, \ldots, M \tag{6}$$

where $S = \{S_1, \ldots, S_M\}$ is the set of prototypes that represents the M classes, m is a user-defined parameter, and $d(C_i, S_j)$ is the distance between C_i and S_j. In the following, it will be assumed that $d(.,.)$ is the Euclidean distance (but any other distance can be chosen); thus, the solution of the defined constrained optimization problem can be achieved with the following iterative procedure.

1) Randomly initialize the matrix U of memberships u_{ij}.
2) At t-step, calculate the new centroids $\hat{S}_i$ as

$$\hat{S}_i = \frac{\sum_{j=1}^{n}(u_{ij})^m \mu_j}{\sum_{j=1}^{M}(u_{ij})^m}. \tag{7}$$

3) Update the memberships u_{ij} as

$$\hat{u}_{ij} = \left[\sum_{k=1}^{M}\left(\frac{d(C_i, S_j)}{d(C_i, S_k)}\right)^{\frac{2}{m-1}}\right]^{-1}. \tag{8}$$

4) If $\max \|u_{ij} - \hat{u}_{ij}\| < \epsilon$ (where ϵ is a threshold value), stop; otherwise, repeat Steps 2) and 3).

In order to have a proper representation of the prior terms in the clustering, we decided to weight each segment with a parameter that is related to its size (segments with a large size play a "more important" role in the clustering than segments with a small size). To achieve this, we have modified Step 3) of the fuzzy c-means clustering algorithm explained before as

$$\hat{S}_i = \frac{\sum_{j=1}^{n}(u_{ij} \cdot Sz_j)^m \mu_j}{\sum_{j=1}^{n}(u_{ij} \cdot Sz_j)^m} \tag{9}$$

where Sz_j is the size of the jth component. In the context of the classifier, size Sz_j refers to the number of pixels that is present in the jth cluster. Note that, if all the classes are fairly well represented in the data, there would be no need to modify the fuzzy c-means technique, as suggested in (9). If some of the classes are very sparsely represented, the use of Sz_j improves the capability of modeling the problem of the fuzzy c-means technique.

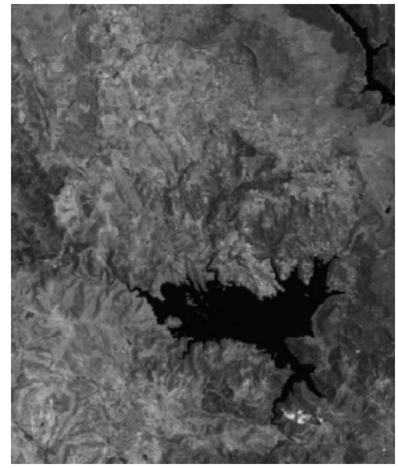

Fig. 1. Band 5 of the Landsat-5 TM image acquired in July 1996 on Sardinia Island, Italy.

C. Step 3: Iterative Estimation of Statistical Parameters of Classes Based on the EM Algorithm

The third step of the proposed procedure consists in optimizing the class representation obtained with the fuzzy c-means clustering. As stated in Section II, to apply the Bayes decision rule, it is necessary to estimate the *a priori* probability $P(\omega_i)$ and conditional density function $p(\mathbf{x}_{r,s}|\omega_i)$ for each class $\omega_i \in \Omega$. Reliable estimates for these statistical terms can be obtained by starting from the estimates that are derived earlier and by using the iterative EM algorithm (which is a well-known technique for the estimation of parameters in incomplete-data problems [12], [13]).

The EM algorithm is an iterative procedure that consists of two alternating steps: an expectation step, followed by a maximization step. The expectation is with respect to the unknown underlying variables, using the current estimates of the parameters and conditioned upon the observations. The maximization step provides new estimates of the parameters. At each iteration, the estimated parameters provide an increase in the maximum-likelihood (ML) function until a local maximum is achieved. Despite the convergence properties, there is no guarantee that the algorithm converges to a global maximum of the likelihood function. This depends on the initial starting points, i.e., the seeds given as input to the algorithm.

To better explain the proposed approach, for simplicity, let us consider the case in which all classes included in Ω can be described by Gaussian distributions. In this case, the density function that is associated with each class $\omega_i \in \Omega$ can be completely described by a mean vector μ_i and a covariance matrix Σ_i. Thus, the set of parameters to be estimated for the application of the Bayes rule is

$$\theta = [\mu_1, \Sigma_1, P(\omega_1), \ldots, \mu_M, \Sigma_M, P(\omega_M)]. \tag{10}$$

The initial values of the aforementioned terms are derived from the previous segmentation and fuzzy c-means clustering steps. These values are then updated according to the EM algorithm, which models the distribution of the image $\mathbf{X}$ as

TABLE I
OVERALL ACCURACIES OBTAINED ON THE LANDSAT TM DATA SET WITH THE PROPOSED CONTEXT-SENSITIVE CLUSTERING, THE CONTEXT-INSENSITIVE AND CONTEXT-SENSITIVE FUZZY c-MEANS CLUSTERING, AND THE SUPERVISED ML CLASSIFIER

Land-Cover Class	Number of Test Samples	Overall Accuracy (%)			
		Proposed Classifier	Fuzzy c-means Classifier		Supervised ML Classifier
			Context-Insensitive	Context-Sensitive	
Pasture+Vineyard	328	59.45	66.46	63.72	94.27
Forest	304	65.13	77.63	78.29	86.51
Urban Area	408	96.81	47.30	62.00	98.28
Water	804	100.00	100.00	100.00	100.00
Overall	1844	86.33	78.69	81.56	96.38

a mixture made up of M components. It can be proved that the equations to be used at each iteration of the EM algorithm for the estimation of the means and the covariance matrices for different classes in Ω are the following [10]:

$$P^{j+1}(\omega_i) = \frac{1}{R \cdot S} \sum_{\mathbf{x}_{\mathbf{r},\mathbf{s}} \in \mathbf{X}} \frac{P^j(\omega^i) p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}}|\omega_i)}{p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}})} \tag{11}$$

$$[\mu_i]^{j+1} = \frac{\sum_{\mathbf{x}_{\mathbf{r},\mathbf{s}} \in \mathbf{X}} \frac{P^j(\omega_i) p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}}|\omega_i)}{p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}})} \mathbf{x}_{\mathbf{r},\mathbf{s}}}{\sum_{\mathbf{x}_{\mathbf{r},\mathbf{s}} \in \mathbf{X}} \frac{P^j(\omega_i) p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}}|\omega_i)}{p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}})}} \tag{12}$$

$$[\Sigma_i]^{j+1} = \frac{\sum_{\mathbf{x}_{\mathbf{r},\mathbf{s}} \in \mathbf{X}} \frac{P^j(\omega_i) p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}}|\omega_i)}{p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}})} \left(\tilde{X}_{r,s,i} \cdot \tilde{X}_{r,s,i}^T \right)}{\sum_{\mathbf{x}_{\mathbf{r},\mathbf{s}} \in \mathbf{X}} \frac{P^j(\omega_i) p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}}|\omega_i)}{p^j(\mathbf{x}_{\mathbf{r},\mathbf{s}})}} \tag{13}$$

where

$$p(\mathbf{x}_{\mathbf{r},\mathbf{s}}) = \sum_{i=1}^{M} P(\omega_i) p(\mathbf{x}_{r,s}|\omega_i) \quad \tilde{\mathbf{X}}_{r,s,i} = \mathbf{x}_{r,s} - [\mu_i]^j \tag{14}$$

j is the iteration index, and $R \cdot S$ is the total number of pixels in the image $\mathbf{X}$.

D. Step 4: Bayesian Decision

The estimates that are obtained for each class $\omega_i \in \Omega$ at convergence, i.e., $P(\omega_i)$, μ_i, and Σ_i, are taken as the final values of the parameters to be used in the classification. The final classification map is obtained by using these values in the framework of the Bayes decision rule in (1).

IV. EXPERIMENTAL RESULTS

In order to assess the effectiveness of the proposed approach, different experiments were carried out on various data sets. For space constraints, results are presented only for two of them.

A. Landsat Thematic Mapper (TM) Data Set

The first data set is made up of a multispectral image that is acquired by the TM multispectral sensor of the Landsat-5 satellite. The selected test site was a section (493×412 pixels) of a scene, including Lake Mulargia on the Sardinia Island, Italy. The image that is used in the experiments was acquired in July 1996. Fig. 1 shows band 5 of the image.

Five land-cover classes (i.e., pasture, forest, urban area, water, and vineyard) characterize the aforementioned image. However, preliminary clustering experiments carried out on the five-class data set showed low accuracy on both pasture and vineyard classes. An analysis of the distributions (in the feature space) of the pixels of these classes points out that they have similar spectral characteristics and that their distributions are significantly overlapped. Thus, in order to better understand the effectiveness of the proposed unsupervised clustering technique in reasonable conditions, the five-class problem was reduced to a four-class problem, where pasture and vineyard were considered as belonging to the same information class. The accuracy that is yielded by the proposed approach is reported in Table I, where it is compared with those obtained by the unsupervised standard (i.e., context-insensitive) fuzzy c-means clustering, the unsupervised context-sensitive fuzzy c-means clustering (i.e., applying Steps 1 and 2 of the proposed method), and the supervised ML classifier. Classification accuracies for all the four classifiers were evaluated on a test set that is defined on the basis of the available ground-truth information (see Table I). A training set was also considered for the estimation of the parameters of supervised ML classifier.

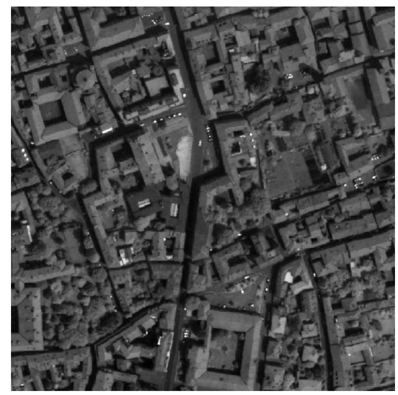

Fig. 2. Pancromatic channel of the Quickbird image acquired in 2002 on Pavia, Italy.

From Table I, one can see that the proposed approach resulted in an overall accuracy of 86.33%. This accuracy is sharply higher than the one that is obtained by both the unsupervised context-insensitive (78.69%) and context-sensitive (81.56%) fuzzy c-means clustering and, as expected, lower than that exhibited by the supervised ML classifier (96.38%). In general, it represents a good result by considering that no ground truth is used with the proposed method.

It is worth noting that this Landsat TM data set is quite challenging for the proposed technique because the moderate geometrical resolution of the TM sensor results in images that do not contain significant homogeneous regions. This reduces the accuracy of the initialization step of the proposed method, which is based on segmentation.

TABLE II
OVERALL ACCURACY OBTAINED ON THE QUICKBIRD DATA SET BY THE PROPOSED CONTEXT-SENSITIVE CLUSTERING, THE CONTEXT-INSENSITIVE AND CONTEXT-SENSITIVE FUZZY c-MEANS CLUSTERING, AND THE SUPERVISED ML CLASSIFIER

Land-Cover Class	Number of Test Samples	Overall Accuracy (%)			
		Proposed Classifier	Fuzzy c-means Classifier		Supervised ML Classifier
			Context-Insensitive	Context-Sensitive	
Tree	192	91.70	34.90	57.81	95.28
Forest	133	66.90	99.30	85.71	87.26
Road	251	73.70	81.30	83.63	31.47
Shadow	262	99.60	100.00	97.33	79.77
Red Roof	939	60.40	37.59	46.43	100.00
Gray Roof	14	71.40	50.00	71.40	66.70
White Roof	92	85.90	92.00	90.22	100.00
Overall	1883	72.60	58.94	64.74	86.42

B. Quickbird Data Set

To further assess the effectiveness of the proposed technique, a second data set was considered, which is made up of an image of 512 × 512 pixels that is acquired by the Quickbird satellite over the city of Pavia (north of Italy) in 2002 (see Fig. 2). Owing to the high geometrical resolution of the Quickbird sensor, this data set is intrinsically more suitable to the properties of the proposed approach.

The accuracy that is obtained by applying the proposed approach to this data set is shown in Table II, together with the definition of the seven considered land-cover classes and with the number of test samples for each of them. The accuracy of the proposed clustering technique was compared with the ones provided by the unsupervised context-insensitive and context-sensitive fuzzy c-means clustering algorithms and by the supervised ML classifier. In addition, for this second image, a training set was defined for the estimation of the parameters of the ML classifier.

From Table II, one can observe that the proposed classifier exhibited an overall accuracy of 72.6%, which is smaller than that obtained by the supervised ML classifier (86.42%) but sharply higher than the one yielded by the unsupervised context-insensitive (58.94%) and context-sensitive (64.74%) fuzzy c-means clustering. These results confirm the effectiveness of the proposed unsupervised clustering technique and the usefulness of the joint use of both the context information in the initialization phase of an unsupervised classifier and the EM algorithm in the classification step.

V. DISCUSSION AND CONCLUSION

A novel context-sensitive clustering technique based on the Bayes decision theory has been presented. The proposed technique is composed of a context-sensitive initialization and an iterative procedure aimed at estimating in an unsupervised way the statistical parameters of classes to be used in the Bayesian decision rule. The initial steps exploit a graph-cut segmentation algorithm, followed by a fuzzy c-means clustering. The iterative procedure is based on the EM algorithm, which, starting from the estimates derived in the initialization steps, achieves the final values of the statistical parameters of classes to be used to accomplish the Bayesian classification. In our experimental results, the proposed technique provided significantly higher accuracies than the context-insensitive and context-sensitive fuzzy c-means clustering algorithms, resulting in more reliable land-cover maps. Comparisons with the supervised ML classifier also pointed out that the effectiveness of the proposed method decreases in the presence of classes that strongly overlapped in the feature space. However, this is a typical drawback of unsupervised methods that should be addressed with the use of more complex features, where classes are better separated or with supervised algorithms.

ACKNOWLEDGMENT

The authors would like to thank A. Garzelli of University of Siena, Italy, for providing the Quickbird image of Pavia and the anonymous referees for their constructive criticism. This work was carried out in the framework of the India-Trento Program for Advanced Research (ITPAR).

REFERENCES

[1] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 2001.
[2] J. Bezedek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2/3, pp. 191–203, 1984.
[3] F. Dell'Acqua and P. Gamba, "Detection of urban structures in SAR images by robust fuzzy clustering algorithms: The example of street tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2287–2297, Oct. 2001.
[4] Y. Zhong, L. Zhang, B. Huang, and P. Li, "An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 420–431, Feb. 2006.
[5] S. Le Hegarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Application of Dempster–Shafer evidence theory to unsupervised classification in multisource remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 1018–1031, Jul. 1997.
[6] A. Baraldi and F. Parmiggiani, "A neural network for unsupervised categorization of multivalued input patterns: An application to satellite image clustering," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 305–316, Mar. 1995.
[7] T. Tran, R. Wehrens, D. Hoekman, and L. Buydens, "Initialization of Markov random field clustering of large remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 8, pp. 1912–1919, Aug. 2005.
[8] S. Lee, "Efficient multistage approach for unsupervised image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2004, vol. 3, pp. 1581–1584.
[9] V. Cherkassky and F. Mulier, *Learning From Data: Concepts, Theory, and Methods*. New York: Wiley, 1998.
[10] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 456–460, Feb. 2001.
[11] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
[12] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.
[13] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.