FUZZY CLUSTERING ALGORITHMS ON LANDSAT IMAGES FOR DETECTION OF WASTE AREAS: A COMPARISON

A.M. Massone[(1)]   F. Masulli[(1,3)]   A. Petrosino[(2)]

(1) Istituto Nazionale per la Fisica della Materia
Via Dodecaneso 33, 16146 Genova, Italy

(2) Istituto Nazionale per la Fisica della Materia
Via S. Allende, I-84081 Baronissi (Salerno), Italy

(3) Dipartimento di Informatica e Scienze dell'Informazione
Università di Genova, Via Dodecaneso 35
16146 Genova, Italy

Abstract - Landsat data can be used to support a wide range of applications for monitoring the conditions of a selected land surface. For example, they can be used to map changes due to the effects of pollution and environmental degradation over different periods of time. In this paper we will present a comparison of fuzzy clustering algorithms for the segmentation of multi-temporal Landsat images. A relabeling stage is performed after the classification in such a way clusters of different segmentations, but corresponding to the same lithological area, are led to a homogeneous color-map.

Keywords: Fuzzy clustering algorithms, Landsat images segmentation, detection of waste.

1 Introduction

Remote sensing can be used to support a wide range of applications in Earth's land surface information management. Typical applications concern, e.g., the mapping of changes due to the effects of pollution and environmental degradation over different periods of time, thanks to the high frequency of coverage of the Earth surface by satellites.

An important class of algorithms used in remote sensing image analysis, is constituted by unsupervised classification (or clustering) algorithms [4]. As pointed out by the recent literature (see, e.g., Baraldi *et al.* [1]) clustering algorithms can overcome the limits of classical classifiers, such as the need of a priori hypothesis on the data distribution, sequentiality, etc. Moreover, the use of unsupervised algorithms is supported by the following arguments:

- Often clustering algorithms are faster and more stable than supervised classification models based on nonlinear optimization.

- The classification results obtained by unsupervised algorithms can provide a test on how good the feature extraction phase works.

- Training areas need not to be labeled during the system training.

In this paper, we shall discuss some relevant clustering algorithms proposed in literature, and then we will compare them with supervised techniques in the segmentation of multi-spectral LANDSAT thematic mapper (TM) images for the detection of waste areas.

In the comparison we will consider unsupervised classifiers based on Hard C-Means (HCM) [4], Fuzzy C-Means (FCM) [5], Possibilistic C-Means (PCM) [6, 7], and Deterministic Annealing (DA) [8].

HCM is an efficient approximation of the Maximum Likelihood technique for estimating clusters centers, using $\{0,1\}$ membership values of patterns to classes. We notice that HCM is subjected to the problem of confinement to local minima of the objective function during the descent procedure. Moreover, concerning the specific application, the crisp memberships for pixels to a class is a too strong constraint due to the limited resolution of sensors. This problem is especially critical for pixels in the border of regions.

In order to overcome the limits of HCM, the FCM algorithm generalizes the HCM objective function introducing the so called *fuzzifier parameter*, obtaining in such a way continuous membership values of patterns to classes.

The Deterministic Annealing (DA) is a different fuzzy approach to clustering based on the minimization of a Free Energy which has been demonstrated [9] equivalent to the FCM functional. The main difference with the FCM concerns the updating of fuzziness control parameter (that here has the meaning of a temperature) during the optimization of the objective function. Starting from a "high enough" value, the cost function is optimized at different scheduled temperature values (annealing procedure). It is worth of noting an on-line version of FCM, introducing also a scheduling of the fuzzifier parameter, has been recently proposed with the names of FKCN [10] and FLVQ [2].

HCM, FCM, DA and FLVQ use the *probabilistic constraint* that the memberships of a pattern across clusters must sum to 1, therefore the membership of a point in a cluster depends on the membership of the same point in all other classes. On the contrary, the PCM algorithm is based on the assumption that the membership value of a point in a cluster is *absolute* and it doesn't depend on the membership values of the same point in any other cluster.

After the classification step, carried out by the described algorithms, a second step of *relabeling* is performed. It is fundamental to lead clusters, coming from different segmentations, relative to the same kind of geographical area, to a homogeneous color-map.

In the next Section we will discuss the FCM, PCM and DA algorithms. In Section 3 we will describe the relabeling algorithm. In Section 4 we will present the experimental data set whereas in Section 5 we will compare and discuss our results. Conclusions are drawn in Section 6.

2 Fuzzy Clustering Algorithms

2.1 The Fuzzy C-Means Algorithm

The Fuzzy C-Means (FCM) algorithm proposed by Bezdek [5] aims to find fuzzy partitioning of a given training set, by minimizing a fuzzy generalization of the Least-Squares functional. Let us assume as Fuzzy C-Means functional:

$$J_m(\mathbf{U}, Y) = \sum_{k=1}^{n} \sum_{j=1}^{c} (u_{jk})^m E_j(\mathbf{x}_k) \tag{1}$$

where:

- $\Omega = \{\mathbf{x}_k | k \in [1, n]\}$ is the training set containing n unlabeled samples;

- $Y = \{\mathbf{y}_j | j \in [1, c]\}$ is the set of cluster centers;

- $E_j(\mathbf{x}_k)$ is a dissimilarity measure (distortion) between the sample $\mathbf{x}_k$ and the center $\mathbf{y}_j$ of a specific cluster j. In this paper we use the Euclidean distance: $E_j(\mathbf{x}_k) = \|\mathbf{x}_k - \mathbf{y}_j\|^2$;

- $\mathbf{U} = [u_{jk}]$ is the $c \times n$ fuzzy c-partition matrix, containing the membership values of all samples in all clusters;

- $m \in (1, \infty)$ is a control parameter of fuzziness.

The minimization of J_m, under the probabilistic constraint $\sum_{j=1}^{c} u_{jk} = 1$, leads to the iteration of the following formulas:

$$\mathbf{y}_j = \frac{\sum_{k=1}^{n} (u_{jk})^m \mathbf{x}_k}{\sum_{k=1}^{n} (u_{jk})^m} \qquad \forall j, \tag{2}$$

and

$$u_{jk} = \begin{cases} \left(\sum_{l=1}^{c} \left(\frac{E_j(\mathbf{x}_k)}{E_l(\mathbf{x}_k)} \right)^{\frac{2}{m-1}} \right)^{-1} & if \quad E_j(\mathbf{x}_k) > 0 \quad \forall j, k \\ 1 & if \quad E_j(\mathbf{x}_k) = 0 \quad (u_{lk} = 0 \quad \forall l \neq j) \end{cases} \tag{3}$$

It is worth noting that choosing $m = 1$ the Fuzzy C-Means functional J_m (Eq. 1) reduces to the expectation of the global error (which we denote as $< E >$):

$$< E > = \sum_{k=1}^{n} \sum_{j=1}^{c} u_{jk} E_j(\mathbf{x}_k), \tag{4}$$

and the FCM algorithm becomes the classic Hard C-Means algorithm [4].

2.2 The Deterministic Annealing Algorithm

The Deterministic Annealing algorithm is an approach to hierarchical cluster based on the minimization of the objective function depending on the temperature. Starting from a "high enough" value, the cost function is deterministically optimized at each temperature. The objective function to be minimized is the Free Energy:

$$F = \sum_{j=1}^{c} \sum_{k=1}^{n} u_{jk} E_j(\mathbf{x}_k) + \frac{1}{\beta} \sum_{j=1}^{c} \sum_{k=1}^{n} (u_{jk} \log u_{jk}) \tag{5}$$

where $E_j(\mathbf{x}_k) = \|\mathbf{x}_k - \mathbf{y}_j\|^2$ and the parameter β can be interpreted as the inverse of temperature T ($\beta = 1/T$) [8] [11] from the statistical mechanics point of view.

For an assigned temperature, the resulting *association degree* is a Gibbs distribution:

$$u_{jk} = \frac{e^{-\beta E_j(\mathbf{x}_k)}}{\sum_{l=1}^{c} e^{-\beta E_l(\mathbf{x}_k)}} \tag{6}$$

and

$$\mathbf{y}_j = \frac{\sum_{k=1}^{n} u_{jk} \mathbf{x}_k}{\sum_{k=1}^{n} u_{jk}} \tag{7}$$

For $\beta \to 0^+$ (starting point of the annealing process), $u_{jk} = 1/c \ \forall j, k$ i.e., each sample is equally associated to each cluster. When β increases, the associations of samples to clusters become crisper and for $\beta \to +\infty$, $u_{jk} = 1$ if $\mathbf{x}_k$ belongs to the cluster j, and $u_{ik} = 0 \ \forall i \neq j$, i.e., each sample is associated to exactly one cluster (*hard limit*).

It is worth noting that, whereas standard clustering algorithms need to specify the number of clusters, the Deterministic Annealing algorithm can start with an over-dimensioned number of clusters. At high temperatures, all centers collapse to a unique point (the center of mass of the distribution), and then, during annealing, "natural" clusters differentiate.

2.3   The Possibilistic C-Means Algorithm

In order to allow a *possibilistic* interpretation of the membership function as a *degree of typicality*, in the Possibilistic C-Means (PCM) the probabilistic constraint is relaxed so that the elements of the fuzzy membership matrix U must simply verify:

$$\bigvee_j u_{jk} > 0 \quad \forall k. \tag{8}$$

In [6], [7], Krishnapuram and Keller presented two versions of the Possibilistic C-Means algorithm. In this paper we consider the second one.

This formulation of PCM [7] is based on a modification to the cost function of the HCM: the objective function contains two terms, the first one is the objective function of the HCM, while the second is a regularizing term, forcing the values u_{jk} to be greatest as possible, in order that points with a high degree of typicality with respect to a cluster may have high u_{jk} values, and points not very representative may have low u_{jk} values in all the clusters:

$$J(U, \mathbf{Y}) = \sum_{j=1}^{c} \sum_{k=1}^{n} u_{jk} E_j(\mathbf{x}_k) + \sum_{j=1}^{c} \eta_j \sum_{k=1}^{n} (u_{jk} \log u_{jk} - u_{jk}), \tag{9}$$

where $Y = \{\mathbf{y}_j \mid j = 1, ..., c\}$ is the set of centers of clusters, $E_j(\mathbf{x}_k)$ is the Euclidean distance $(E_j(\mathbf{x}_k) = \|\mathbf{x}_k - \mathbf{y}_j\|^2)$, and the parameter η_j depends on the distribution of points in the j-th cluster and is assumed to be proportional to the mean value of the *intra-cluster distance*.

If clusters with similar distributions are expected, η_j could be set to the same value for each cluster. In general, it is assumed that η_j depends on the average size and on the shape of the j-th cluster.

As demonstrated in [7], the couple $(U, \mathbf{Y})$ minimizes J, under the constraint (8) only if $\mathbf{y}_j$ and u_{jk} are given by:

$$\mathbf{y}_j = \frac{\sum_{k=1}^{n} u_{jk}\, \mathbf{x}_k}{\sum_{k=1}^{n} u_{jk}} \quad \forall j, \qquad u_{jk} = \exp\left\{-\frac{E_j(\mathbf{x}_k)}{\eta_j}\right\} \quad \forall j, k. \tag{10}$$

A bootstrap clustering algorithm is anyway needed before starting PCM, in order to obtain an initial distribution of prototypes in the feature space and to estimate parameters η_j. In this paper we will use outputs of a FCM in order to estimate η_j parameters according to [6]:

$$\eta_j = K \frac{\sum_{k=1}^{n} (u_{jk})^m E_j(\mathbf{x}_k)}{\sum_{k=1}^{n} (u_{jk})^m} \tag{11}$$

where K is a constant.

3   The Relabeling Algorithm

In order to compare the segmentation results obtained using two different clustering algorithms on the same dataset, it is necessary to find a one-to-one mapping between clusters generated by two different algorithms.

For this purpose we used the *relabeling algorithm* proposed in [10]. Given a reference classification, obtained by one of the two clustering techniques, the relabeling algorithm calculates a *co-occurrence matrix* $C = [c_{ij}]$, where the rows are the labels of regions in the reference segmentation and the columns are the labels of regions in the segmentation to be re-labeled. The generic element c_{ij} represents the number of points labeled i in the reference

1. $k = 0$;

2. **do until** $k < nclass$;

    (a) $(i^*, j^*) = arg\ max_{i,j}\ c_{i,j}$;

    (b) $A(j^*) = i^*$;

    (c) $c_{i^*j} = 0\ \ \forall j$;

    (d) $c_{ij^*} = 0\ \ \forall i$;

3. $k + +$;

4. **end do.**

Table 1: Relabeling Algorithm.

segmentation and j in the other segmentation. Then the relabeling algorithm compiles the association vector A, as shown in Table 1.

After the application of the relabeling algorithm we can use homogeneous (consistent) color-maps in the different segmentations.

4   Experimental Data Set and Methods

The experimental data set consist of three multi-spectral Landsat thematic mapper (TM) images acquired in May 1994, March 1997 and October 1997. The selected geographical area is located between Monte San Michele and Piana di San Marco Vecchio, near Caserta (Italy), and the specific goal was the discrimination and monitoring of caves and wasting areas present in the scene. In our case we use only six out of the seven available bands (we exclude the thermal infrared sixth band) and we analyzed several combinations of three bands. Among the possible combinations of Landsat bands, the most significant for our aims have been:

1. The bands 4, 5 and 7 which allow the discrimination of urban areas from forest areas.

2. The bands 4, 3 and 2 which allow the discrimination of bare areas from grass.

3. The bands 5, 4 and 1 for the discrimination of vegetation moisture content and soil moisture, determining vegetation types and delineating water bodies and roads.

We tested the combination of bands 5, 4 and 1 which is of great efficacy for the aims of our analysis. In Figures 1 and 2 the set of bands 5, 4 and 1 are depicted respectively for the month of May 1994 and March 1997. The fusion of selected bands defines a three-dimensional feature space whose point coordinates represent the intensity values of each band; the detection of clusters in the feature space corresponds to a possible segmentation of the input image in agglomerative areas.

For the HCM and FCM algorithms we fixed the number of clusters to be found to be 8, whereas the Deterministic Annealing algorithm found itself the same number of classes starting from an over-dimensioned number (in our case 10 clusters). Furthermore, the starting point for the PCM algorithm was the FCM output.
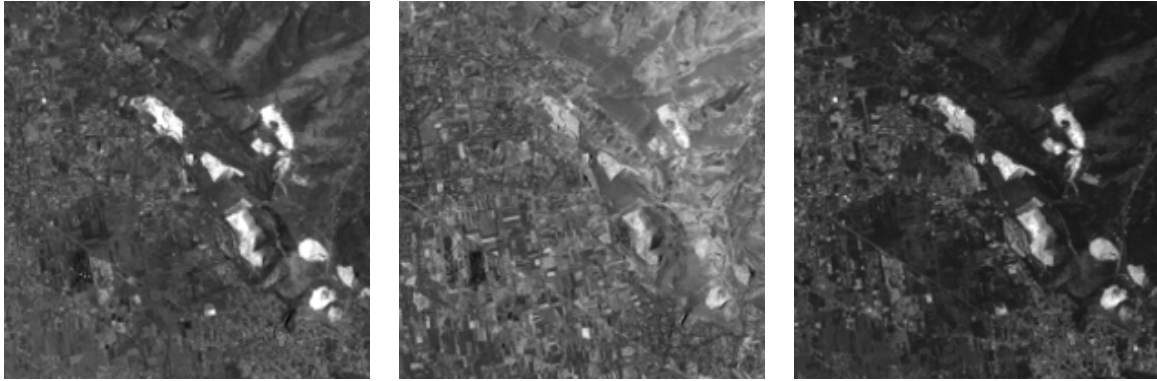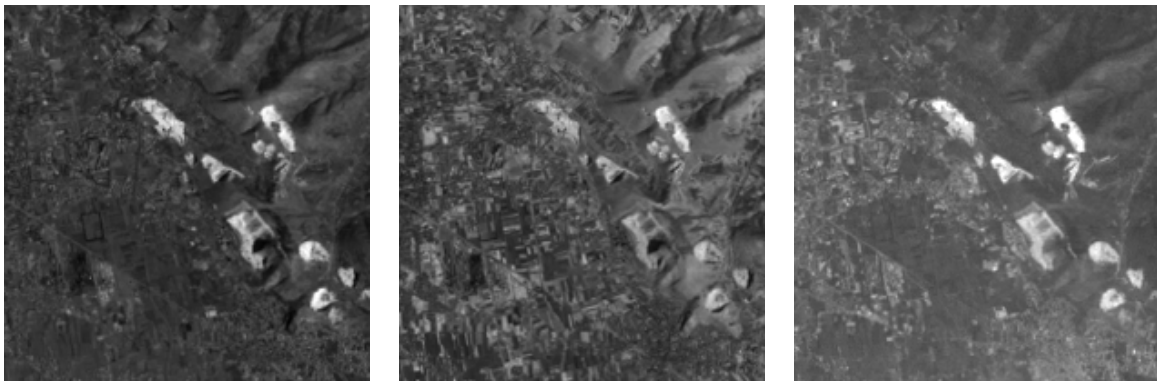
Figure 1: Band 5 (a), Band 4 (b), and Band 1 (c). May 1994.

Figure 2: Band 5 (a), Band 4 (b), and Band 1 (c). March 1997.

The *fuzzifier parameter m* in the FCM was chosen to 2, while the other fundamental parameters were set after several trials. In the PCM algorithm the parameter K (Eq. 11) was set to 0.1. In the Deterministic Annealing algorithm the initial value of β (Eq. 5) was set to 10^{-4} and the scheduling equation was:

$$\beta^{t+1} = 1.1 \; \beta^t \tag{12}$$

The results of the unsupervised methods were compared to those obtained from the application of the supervised techniques *Maximum Likelihood* and *K-Nearest Neighbour* [4]. The supervised methods were trained over five areas extracted by a photo-interpreter, each characterizing a specific class: shadow, waste/quarry, urban area, cultivated area and forest.

5   Results and Discussion

The classification obtained over the images dated May 1994 by using unsupervised clustering are shown in Fig. 3 [1]. In Fig. 4, the same algorithms are applied to the images dated March 1997; while in Fig. 5 we show the results generated from the same data set by using the Maximum Likelihood and K-Nearest Neighbour techniques.

As shown, the results generated by the supervised and unsupervised methods well compare each other, in terms of correctly classified pixels. In particular, the results obtained by using fuzzy clustering methods outperform the crisp ones and are more comparable to those resulted by the supervised classification methods.

The fuzzy clustering methods allow to classify in a semi-automatic manner images where the content is not known *a priori*; only the information about the maximum number of classes is needed. In particular, the fuzzy methods have allowed to identify objects in a more flexible manner, assigning to each pixel degree of membership to the object-classes in the scene.

Due to these characteristics, the classification results produced by fuzzy methods have allowed to identify a neglected waste site in the geographical area under exam, which was not known before the present study. Specifically, the waste site is located in the lower-left part of the image and it is evident how it is less wide in the image dated May 1994 with respect to the image dated March 1997.

6   Conclusions

In the study reported in this paper we have applied and compared different supervised and unsupervised classification algorithms for the detection of waste areas using LANDSAT TM images.

It is worth of noting that the 30 meters spatial resolution of the Landsat-TM sensor makes the process of detecting waste areas effective only for medium (10,000-60,000 m^2) to large (200,000-300,000 m^2) landfills, thus being unusable for small (40-50 m^2) ones. This limitation has not allowed us to identify more sites than those reported here.

It is however under study the application of the methods presented here to high-resolution images obtained by the bispectral infrared scanner ATL-80 and the panchromatic images sensed by the IKONOS II satellite, where the land resolution is nearly one meter square; this should allow more refined detection results, also for small waste disposal areas.
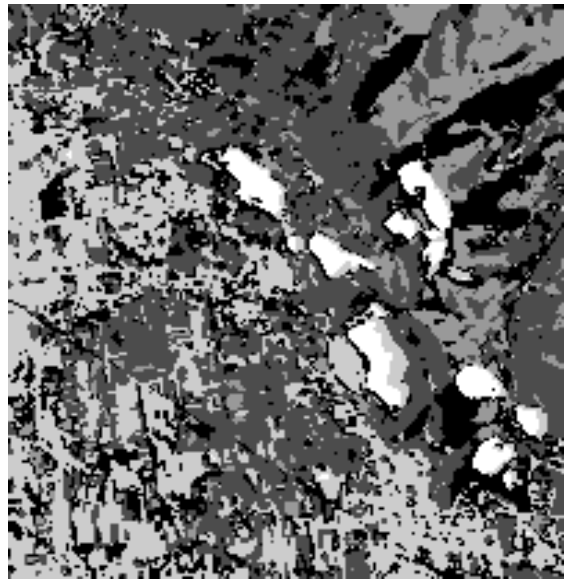

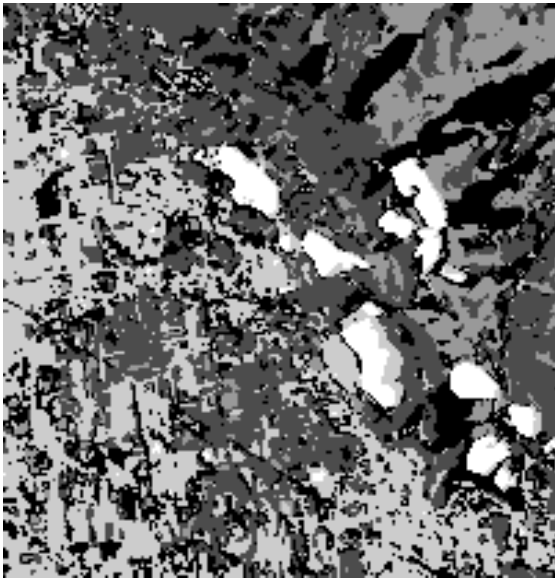[1]Color versions of all segmentation results presented in this paper are available at http://www.ge.infm.it/~massone/TELEMA.

Legend

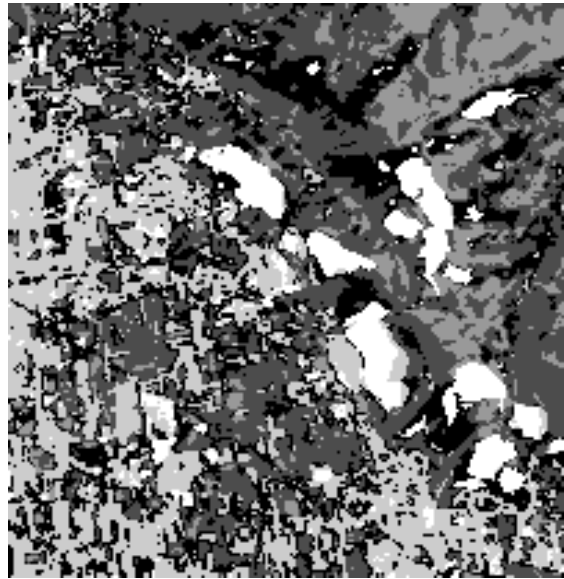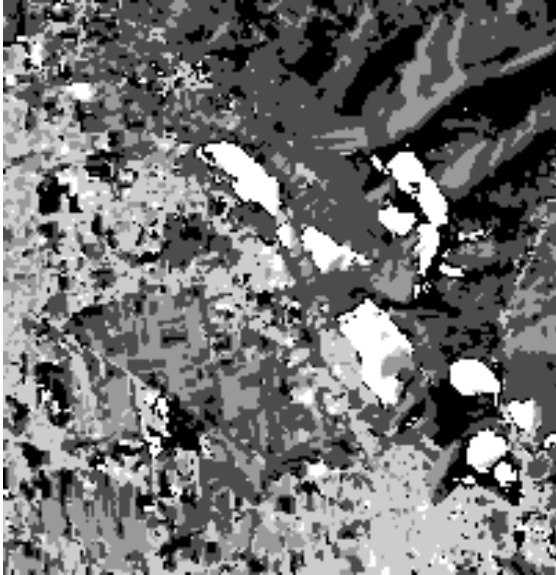|	
⬛	Forest areas
⬛	Cultivated areas
⬛	Shadow
⬛	Urban areas
⬜	Quarry and waste areas
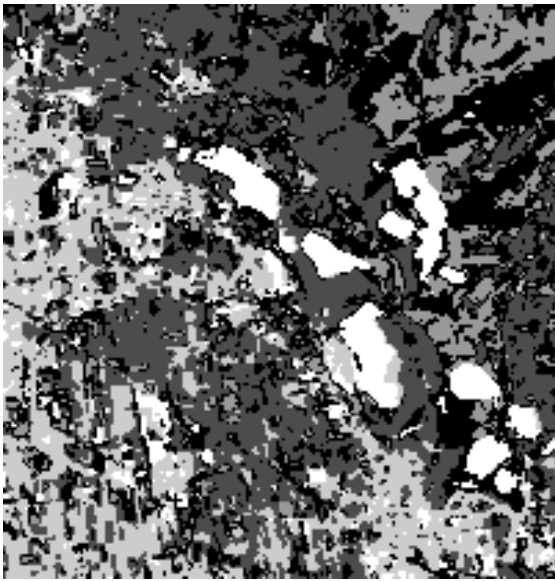
(a)

(b)

(c)

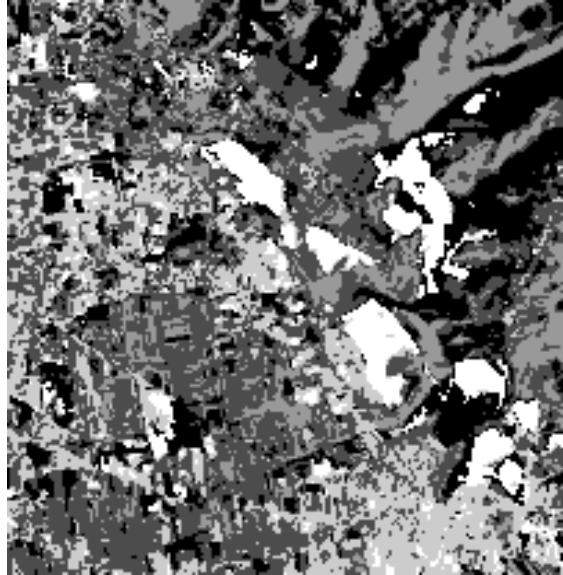(d)

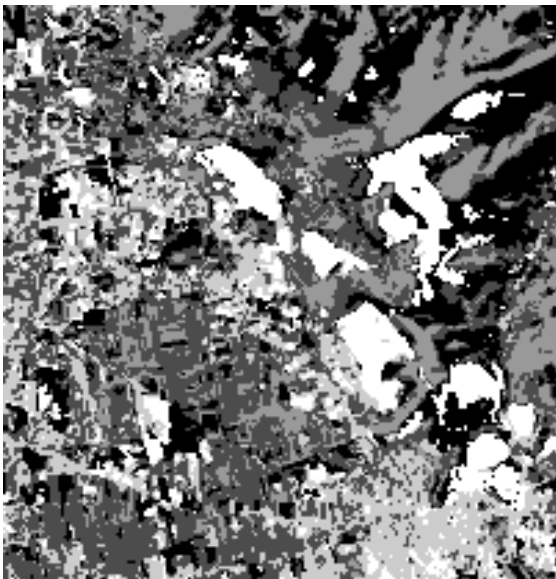*Figure 3: Segmentations obtained using HCM (a), FCM (b), PCM (c), and Deterministic Annealing (d). May 1994.*
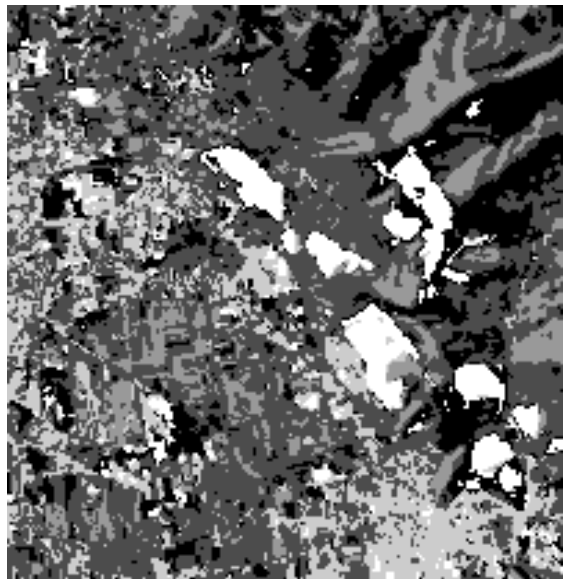
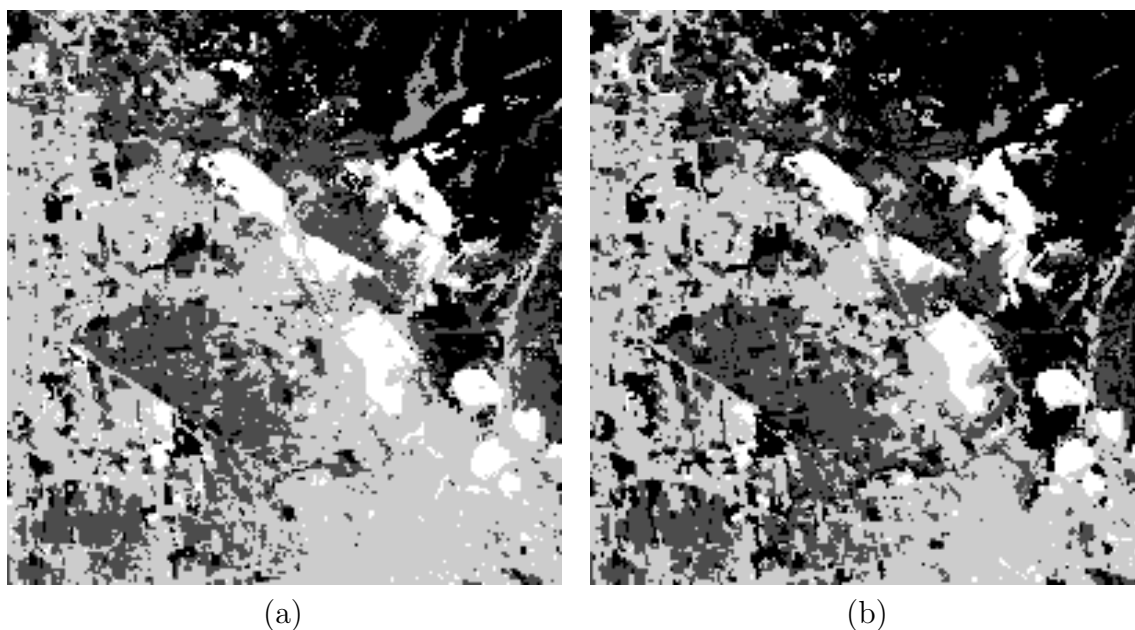Figure 4: Segmentations obtained using HCM (a) FCM (b), PCM (c), and Deterministic Annealing (d). March 1997.

Figure 5: The Maximum Likelihood (a) and K-Nearest Neighbour (b) classification results over the set of bands 5-4-1 of the Landsat images. March 1997.

In addition, while spectral knowledge plays an important role in the interpretation of Landsat images, spatial domain knowledge can be efficiently used to adjust image interpretation on the basis of the expected relationships (such as contiguity) among different land structures. Methods for integrating different forms of knowledge and knowledge based methods are therefore needed both to manage symbolic and numerical information.

Acknowledgments

References

[1] A. Baraldi *et al.* "Model Transitions in Descending FLVQ". *IEEE Transactions on Neural Networks*, vol.9, no.5, pp. 724-738, 1998.

[2] J.C. Bezdek and N.R. Pal. "Two soft relative of learning vector quantization". *Neural Networks*, vol.8, no.5, pp. 729-743, 1995.

[3] T. Kohonen. "The self-organizing map". *Proc. IEEE*, vol.78, no.9, pp. 1464-1480, 1990.

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

[5] J.C. Bezdek. "Pattern Recognition with Fuzzy Objective Function Algorithms". Plenum Press, New York, 1981.

[6] R. Krishnapuram and J.M. Keller. "A possibilistic approach to clustering". *IEEE Transactions on Fuzzy Systems*, 1:98–110, 1993.

[7] R. Krishnapuram and J.M. Keller. "The Possibilistic C-Means algorithm: Insights and recommendations". *IEEE Transactions on Fuzzy Systems*, 4:385–393, 1996.

[8] K. Rose, E. Gurewitz, G. Fox. "A deterministic approach to clustering". *Pattern Recognition Letters*, vol.11, pp. 589-594, 1990.

[9] S. Miyamoto, M. Mukaidono. "Fuzzy C-Means as a Regularization and Maximum Entropy Approach". *Proceedings of the Seventh IFSA World Congress*, pp. 86-91, 1997.

[10] E.C.K. Tsao, J.C. Bezdek and N.R. Pal. "Fuzzy Kohonen Clustering Networks". *Pattern Recognition*, vol.27, pp. 757-764, 1994.

[11] K. Rose. "Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems". *Proceedings of the IEEE*, vol.86, No. 11, pp. 2210-2239, 1998.