

# Fuzzy Clustering Methods for the Segmentation of Multivariate Medical Images

F. Masulli<sup>(1)</sup>, M. Artuso<sup>(2)</sup>, P. Boguś<sup>(3)</sup>, and A. Schenone<sup>(4)</sup>

(1) Istituto Nazionale per la Fisica della Materia  
Dipartimento di Fisica - Università di Genova  
Via Dodecaneso 33 - 16146 Genova (Italy)

(2) DSI - Dipartimento di Scienze dell'Informazione  
Università di Milano - Via Comelico 39 - 20135 Milano (Italy)

(3) Akademia Medyczna w Gdańsku  
Katedra i Zakład Fizyki i Biofizyki  
ul. Debinki 1 80-211 Gdańsk (Poland)

(4) Istituto Nazionale Ricerca sul Cancro  
Largo R. Benzi 10 16132 Genova (Italy)

email : {masulli|artus|bogus}@ge.infm.it, schenone@cisi.unige.it

**Abstract**—In this article, we present some results on the application of fuzzy methods to the segmentation of multivariate medical images. We report the results obtained by using the Fuzzy C-mean (FCM) algorithm by J. Bezdek and the method by K. Rose, E. Gurewitz and G. Fox (RGF) based on the Maximum Entropy Principle (MEP) that avoids any a priori assumption on the number of classes. In particular we study the effect of using, for each new epoch of the algorithm, a Reduced Data Base (RDB) obtained through an uniform random sampling of the original data base. From our experiments the RGF method shows the best efficiency in term of reliability of the solutions while the FCM results faster for big RDBs.

## I. INTRODUCTION

In the clinical field, an increasing number of different diagnostic imaging methodologies have been introduced in the last few years. Nowadays medical images are obtained from different acquisition modalities, including X-ray tomography (CT), magnetic resonance imaging (MRI), single photon emission tomography (SPECT), and positron emission tomography (PET), each of them carrying complementary information (both structural and functional) on biological tissues. The visual inspection of a large set of images, as performed by a physician, permits to exploit only partially the global information.

It is necessary to build Medical Imaging Support Decision Systems (MISSD) able to extract the salient information embedded in the multivariate medical image, removing redundancies and noise.

In [1], a MISSD devoted to the segmentation

of multivariate medical images was presented. It is made up by an interactive graphical system supporting the full analysis sequence: feature extraction, reduction of dimensionality, unsupervised clustering, voxel classification, and post-processing refinements. The core of the system was a segmentation technique based on an unsupervised clustering neural network named "capture effect" [2].

The limits of the Capture Effect Neural Network (CENN) are mainly concerned with the difficulty to obtain good results in problems with high dimensionality, while its main merits are the speed and the reliability in finding the correct solution of the clustering problem in feature spaces of low dimensionality. To overcome those limits we are studying potential benefits in applying fuzzy clustering methods to the segmentation of multivariate image.

In this article, we present some results obtained by using the Fuzzy C-mean (FCM) algorithm by J. Bezdek [3] and the method by K. Rose, E. Gurewitz and G. Fox (RGF) [4], [5] based on the Maximum Entropy Principle (MEP) approach [6]. Moreover, we study the effect of data base sampling obtained by using for each new epoch of the training algorithm a new Reduced Data Base (RDB) obtained through an uniform random sampling of the original data base.

In the next section, the problem of segmentation of multivariate images through segmentation in the feature space is discussed. In Sections III and IV we introduce the FCM and the RGF algorithms. In Section V we present the data set. The sampling technique is explained in Section VI. Section VII presents the obtained results. Conclusions are given in Section VIII.

## II. SEGMENTATION THROUGH CLUSTERING

Multivariate volumes can be built from a number of different diagnostic volumes with complementary information (both structural and functional) provided by medical imaging technology, for fully correlating information about the same patient.

An efficient analysis of multivariate medical imaging volumes is an inherently complex task in which each component of the data structure, that is the spatial distribution of the values of a single feature, must be considered together with all the other components. Such an analysis may be helpful in the clinical oncological environment to delineate volumes to be treated in radiotherapy and surgery and to assess quantitatively (in terms of tumor mass or detection of metastases) the effect of oncological treatments.

All these applications involve the extraction of objects or other entities of interest from the imaging data, usually by defining sets of voxels with similar features within the entire multivariate volume. This task is a possible definition of image segmentation and is usually accomplished, either

- by methods of edge detection (e.g. gradient operators), or
- by methods of similarity detection (e.g. thresholding and region growing techniques).

Actually, volumes of interest in medical imaging are not strictly bounded and the application of similarity methods to multivariate data is complex and often very time consuming with complex geometries.

Let us consider a multivariate volume resulting from the spatial registration of a set of  $s$  different imaging volumes. We may notice that its voxels are associated to an array of  $s$  values, each one representing the intensity of a single feature in that voxel. In other words, the  $s$  different intensity values related to each voxel in such multivariate volumes can be viewed as the coordinates of the voxel within a  $s$ -dimensional feature space where multivariate analysis can be made.

Two different spaces have therefore to be considered for a more complete description of the segmentation problem:

- an image space (usually 3D) defined by the spatial coordinates of the data set, and
- a multidimensional feature space as described before.

The principal steps in segmenting of multivariate volumes is the definition of clusters within the  $s$ -dimensional feature space and the classification of all the voxels of the volume in the resulting classes.

These two goals can be attained both by supervised and unsupervised methods.

Supervised methods has been largely employed in medical imaging segmentation studies but provide for conditions hardly satisfied in the clinical environment. First of all, they require the labeling of prototypical samples needed by the generalization process to be applied. Even if the number of clusters is predefined, careful labeling of voxels in the training set belonging with certainty to the different clusters is not trivial especially when concerning multivariate data sets. Moreover, bias can be introduced by users due to the large inter-user variability generally observed when manual labeling is performed [1].

On the contrary, unsupervised approaches self-organize the implicit structure of data and make clustering of the feature space independent from the user definition of the training regions.

## III. THE FUZZY C-MEAN ALGORITHM

Let us assume as a fuzzy C-Means Functional

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

where

- $\Omega = \{\mathbf{x}_k | k \in [1, n]\}$  is a training set containing  $n$  unlabeled samples;
- $Y = \{\mathbf{y}_j | j \in [1, c]\}$  is the set of the centers of clusters;
- $E_j(\mathbf{x}_k)$  is a dissimilarity measure (distance or cost) between sample  $\mathbf{x}_k$  and the center  $\mathbf{y}_j$  of a specific cluster  $j$ ;
- $\mathbf{U} = [u_{j,k}]$  is the  $c \times n$  fuzzy c-partition matrix, containing the membership values of all samples to all clusters;
- $m \in (1, \infty)$  is control parameter of fuzziness.

The clustering problem can be formulated as the minimization of  $J_m$  with respect to  $Y$ , under the normalization

$$\sum_{j=1}^c u_{j,k} = 1. \quad (2)$$

The Fuzzy C-Mean (FCM) algorithm found by Bezdek [3] consists in the iteration of the following formula:

$$\mathbf{y}_j = \frac{\sum_{k=1}^n (u_{j,k})^m \mathbf{x}_k}{\sum_{k=1}^n (u_{j,k})^m} \quad \text{for all } j, \quad (3)$$

with the assumptions:  $E_j(\mathbf{x}_k) = \|\mathbf{x}_k - \mathbf{y}_j\|^2$  and

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

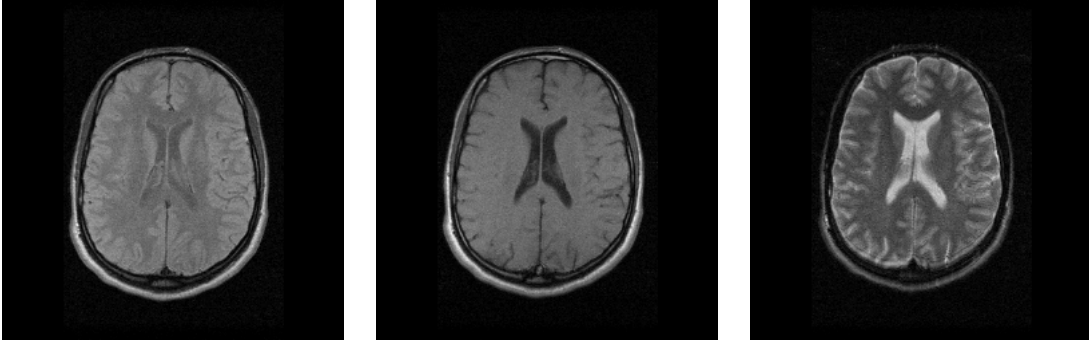


Fig. 1. NMR images: From the left to the right, T1-, T2- and proton density- weighted MRI data.

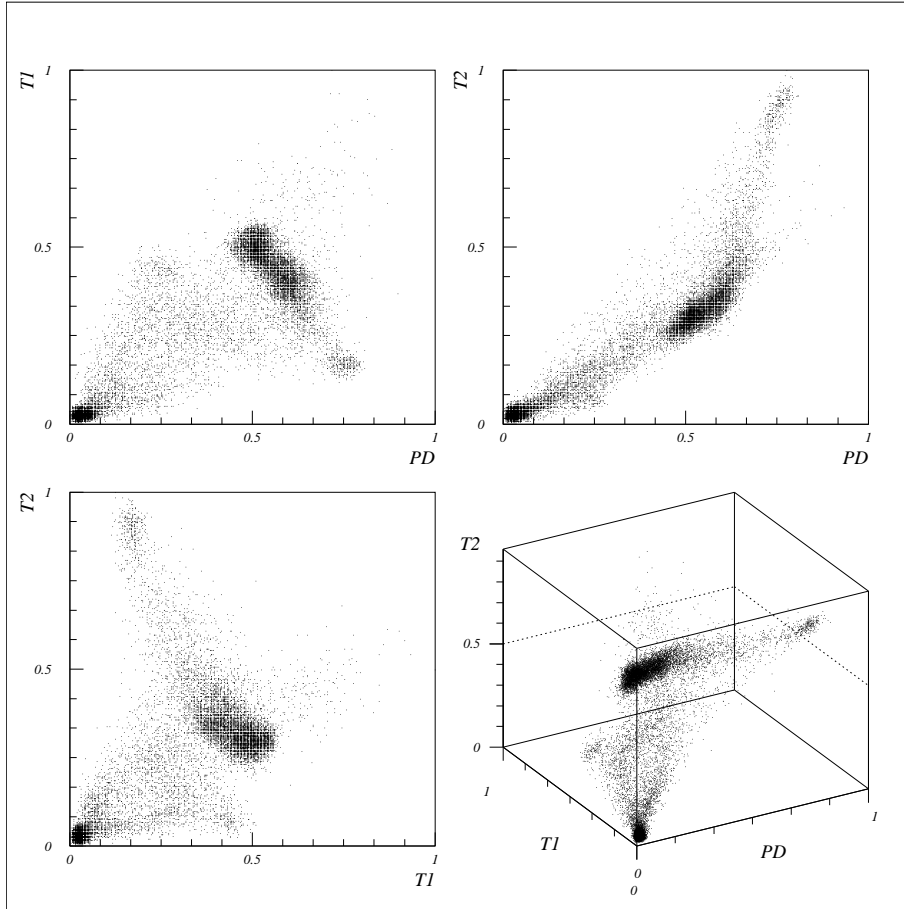


Fig. 2. Three dimensional feature space (down on the right) and its projections onto the three two dimensional spaces.

It is worth to underline that if one chooses  $m = 1$  the fuzzy C-Means Functional  $J_m$  (Eq. (1)) reduces to the expectation of the global error (that we shall denote as  $\langle E \rangle$ ):

$$\langle E \rangle = \sum_{k=1}^n \sum_{j=1}^c u_{j,k} E_j(\mathbf{x}_k), \quad (4)$$

and the FCM becomes the classic crisp C-Means algorithm [7].

#### IV. THE RGF CLUSTERING ALGORITHM

Let be  $p(\mathbf{x}_k)$  the probability distribution of patterns of the training set  $\Omega$ , and

$$H(p(\mathbf{x}_1), \dots, p(\mathbf{x}_n)) = -K \sum_{k=1}^n p(\mathbf{x}_k) \ln p(\mathbf{x}_k)$$

be the Shannon entropy [8] of  $\Omega$ . If we maximize  $H$  under the constraints of Eq.s (4) and (2), the mem-

bership functions of samples to clusters are Gibbs distributions[6]:

$$p(\mathbf{x}_k) = \frac{e^{-\beta E_j(\mathbf{x}_k)}}{Z_k},$$

with

$$Z_k = \sum_{l=1}^c e^{-\beta E_l(\mathbf{x}_k)}$$

being a normalization factor named Partition Function. From a Statistical Mechanics point of view, the Lagrange multiplier  $\beta$  is interpreted as the inverse of temperature  $T$  ( $\beta = 1/T$ ). Moreover, it can be interpreted as a control parameter of fuzziness. In fact when  $\beta$  increases, the association of samples to a clusters become crisper.

In the following we shall use  $u_{j,k}$  instead of  $p(\mathbf{x}_k)$ , in order to make explicit the interpretation of  $\beta$  as a control parameter of fuzziness.

The limit cases are:

- for  $\beta \rightarrow 0^+$  we have  $u_{j,k} = 1/c$  for all  $j, k$ , i.e. each sample is equally associated to each cluster;
- for  $\beta \rightarrow +\infty$  we have  $u_{j,k} = 1$  if  $x_k$  belongs to the cluster  $j$ , and  $u_{i,k} = 0$  for all  $i \neq j$ ,  $i \in [1, c]$ , i.e. each sample is associated to only one cluster (hard-limit).

Let we define the Effective Error (also named the Free Energy, in analogy with Statistical Mechanics)

$$F = -\frac{1}{\beta} \ln Z, \quad (5)$$

where  $Z = \prod_k Z_k$  is named the Total Partition Function. One can demonstrate that

$$\lim_{\beta \rightarrow \infty} F = \langle E \rangle \quad (6)$$

This limit allows us to find the solution of the constrained minimization of  $\langle E \rangle$  by performing a so-called *Deterministic Annealing* on  $F$ , as proposed by Rose, Gurewitz and Fox [4], [5] (see also [9]). This method (that we name RGF) starts by minimizing  $F$  with a high  $T$ , for which there is one unique solution ( $u_{j,k} = 1/c$  for all  $j, k$ ), and then reduces  $T$ , until the hard-limit is reached.

In the algorithm used here, we assume, as proposed in [4], [5]  $E_j(\mathbf{x}_k) = \|\mathbf{x}_k - \mathbf{y}_j\|^2$ , and for each value of  $\beta$  we performed a minimization of  $F$  with respect to  $Y$ , by iterating the following formula:

$$\mathbf{y}_j = \frac{\sum_{k=1}^n u_{j,k} \mathbf{x}_k}{\sum_{k=1}^n u_{j,k}} \quad \text{for all } j, \quad (7)$$

It worth noting that, while standard clustering algorithms (included FCM) need to specify the number of clusters, the RGF can start with a over-dimensioned number of clusters. For high temperatures all centers collapse in a unique point, and then, during annealing, the "natural" clusters differentiate.

## V. EXPERIMENTAL DATA SET

In this paper, the FCM and the RGF fuzzy algorithms were applied to the segmentation of white matter, gray matter, cerebrospinal fluid, and skull, using MRI images of the head. In particular, three volumetric data sets representing T1-, T2-, and proton density- weighted MRI data of a healthy volunteer have been used (see Fig.1). No corrections have been made to reduce the inter-slice variability of image intensity.

As shown in Fig.2, the fusion of data sets produces a three-variate volume that defines a three-dimensional feature space. Each triplet of voxel intensity in the volume is represented by a point in a 3D feature space, whose coordinates represent the intensity values in that voxel of each volume belonging to the multivariate volume.

## VI. SAMPLING OF DATA BASE

Our aim is to detect clusters in the feature space and to use them for segmenting the input images.

By using the full training set originated from our multivariate image, we obtained a training set consisting of 37620 samples. On a UltraSPARC workstation the obtained convergence times are about 158 sec for the RGF, and 35 sec for the FCM.

In order to speed-up those algorithms we borrowed from the Capture Effect Neural Network (CENN) [2] the idea of sampling the training set, but instead of using the same subset of  $\Omega$  (Reduced Data Base (RDB)) for the whole algorithm, we randomly re-sample a new RDB for each epoch of training.

## VII. RESULTS AND DISCUSSION

The projections into the two dimensional feature space (T1 versus PD) of some results of segmentation performed in the 3D space, and the corresponding image spaces are shown in Fig.3 and Fig.4.

Our implementation of FCM used  $m = 2$ , while in the RGF we performed the deterministic annealing with  $\beta$  ranging from 5 to 245, increasing by steps of 15 in the range [5, 95], and steps of 25 in the range (95, 245].

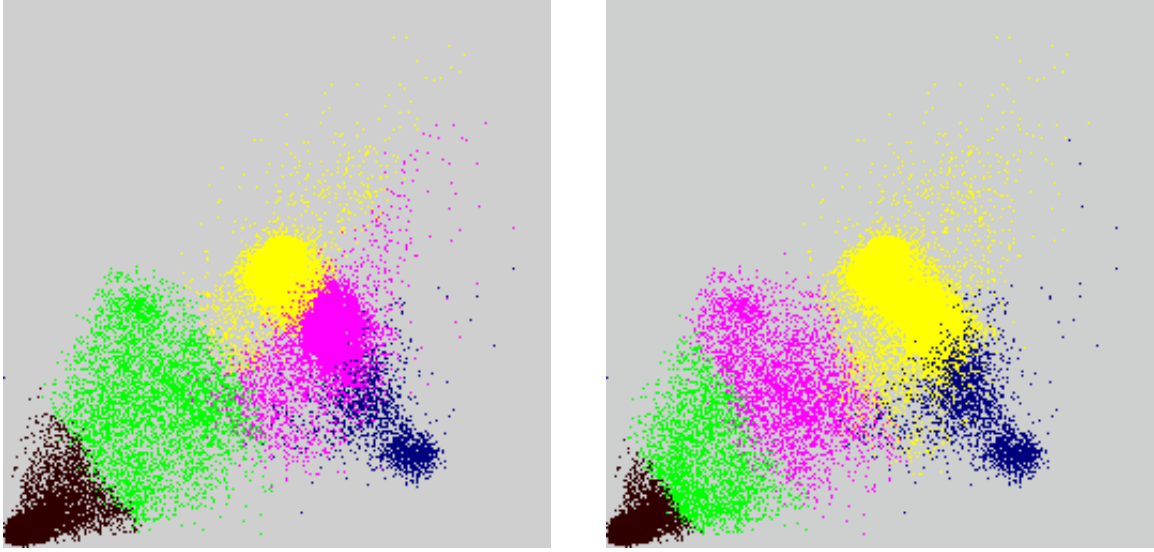


Fig. 3. Segmented T1 versus PD subspace for the optimal solution (on the left) and the second most frequent one in our trials using the FCM (on the right).

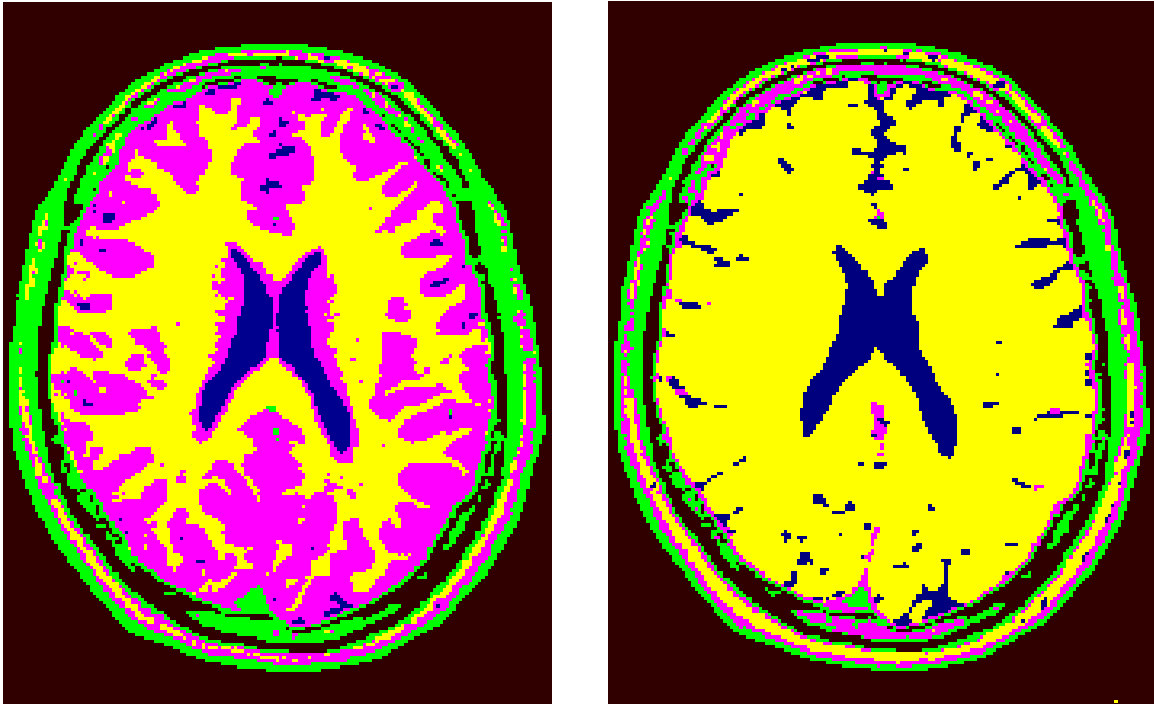


Fig. 4. Optimal segmentation (on the left) and second most frequent one in our trials using the FCM (on the right).

Note that in both method the centers of clusters  $Y$  were initialized at random in the feature space, while the stop criteria was:

$$\bigvee_j (\|y_j(t+1) - y_j(t)\|) \leq .01 \quad j \in [1, c],$$

where  $t$  is the iteration index.

In Figure 5 we report some results obtained with epochs using RDB ranging from 5% to 100% of the original DB. Each point in the graphs is an averaged value obtained from 10 experiments.

As shown, the data base re-sampling method

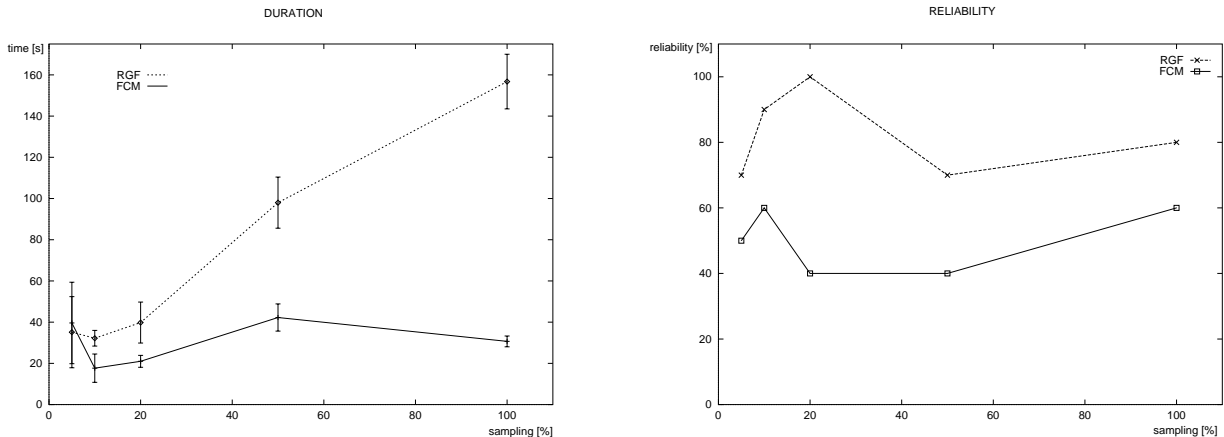


Fig. 5. Duration and reliability versus size of the RDB. The experiments were performed on an UltraSPARC workstation.

reduces the execution times of the training algorithms (durations), maintaining at the same time a high segmentation quality of multivariate images.

Moreover, the uncorrelation of data sets used in each training epoch, produces a regularizing effect leading to more reliable solutions. In particular we can notice that, by using an RDB size of 20%, the RGF shows the optimal segmentation reliability (100%), and its durations are acceptable for real applications.

## VIII. CONCLUSIONS

In this article, some results on the application of fuzzy methods to the segmentation of medical multivariate images are presented.

The results obtained by using the Fuzzy C-mean (FCM) algorithm by J. Bezdek and the method by K. Rose, E. Gurewitz and G. Fox (RGF) are compared. In particular the effect of sampling of data base is studied.

As shown by our results, this approach reduces the execution times of the training algorithms, maintaining in the same time a high segmentation reliability of multivariate images.

## IX. ACKNOWLEDGMENTS

This work was supported by grants from INFM, MURST, and GNCB-CNR. Piotr Boguś undertook this work with the support of the ICTP Programme for Training and Research in Italian Laboratories, Trieste, Italy. We thank Leonard Studer for useful discussions.

## REFERENCES

- [1] A. Schenone, F. Firenze, F. Acquarone, M. Gambaro, F. Masulli, and L. Andreucci, "Segmentation of multivariate medical images via unsupervised clustering with adaptive resolution", *Computer Aided Medical Imaging and Graphycs*, vol. 20, 1996, (*in press*).
- [2] F. Firenze, A. Schenone, F. Acquarone, and P. Morasso, "An interactive neural network based approach to the segmentation of multimodal medical images", in M. Marinaro and R. Tagliaferri, editors, *Neural Nets WIRN Vietri-95*, pp. 251–259, Singapore, 1996. World Scientific.
- [3] J.C Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- [4] K. Rose, E. Gurewitz, and G. Fox, "A deterministic annealing approach to clustering", *Pattern Recognition Letters*, vol. 11, pp. 589–594, 1990.
- [5] K. Rose, E. Gurewitz, and G. Fox, "Constrained clustering as an optimization method", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, pp. 785–794, 1993.
- [6] E. T. Jaynes, "Information theory and statistical mechanics", *Physical Review*, vol. 106, pp. 620–630, 1957.
- [7] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
- [8] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.
- [9] G. Beni and X. Liu, "A least biased fuzzy clustering method", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, pp. 954–960, 1994.