Evolution Strategy for the C-Means Algorithm: Application to multimodal image segmentation

Francesco Masulli

DIBRIS - Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi - University of Genoa
Via Dodecaneso 35, 16146 Genoa - Italy
and SICRMM Temple University, Philadelphia - PA

francesco.masulli@unige.it

Anna Maria Massone

CNR - SPIN
via Dodecaneso 33 - I-16146 Genoa - Italy

annamaria.massone@spin.cnr.it

Andrea Schenone

DIBRIS - Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi - University of Genoa
Via Dodecaneso 35, 16146 Genoa - Italy

andrea.schenone@unige.it

February 24, 2013

Abstract

Evolutions Strategies (ES) are a class of Evolutionary Computation methods for continuous parameter optimization problems founded on the model of organic evolution. In this paper we present a novel clustering algorithm based on the application of an ES to the search for the global minimum of the C-Means (CM) objective functional. The new algorithm is then applied to the clustering step of an interactive

system for the segmentation of multimodal medical volumes obtained by different medical imaging diagnostic tools. In order to aggregate voxels with similar properties in the different diagnostic imaging volumes, clustering is performed in a multidimensional space where each independent dimension is a particular volumetric image. As a consequence, in this application clustering supports an inference process based on complementary information carried by each image (e.g. functional or anatomical) in oder to extract regions corresponding to the different anatomical and/or pathological tissues. A quantitative comparison of segmentation results obtained by the original CM and by the new algorithm is reported in the paper.

1 Introduction

C-Means (CM) [6] is a widely used clustering method based on a simple and efficient numerical approximation to the maximum likelihood technique for the estimation of probability mixtures parameters [6, 3].

The CM shows some intrinsic problems. In particular, it is subject to the problem of trapping in local optima of its objective function. In the clustering literature, many algorithms based on fuzzy set theory have been proposed in order to overcome this limit of CM, among them the *Fuzzy C-Means* algorithm [3], the Deterministic Annealing [20], and the Possibilistic C-Means [12, 13]. As shown by Miyamoto and Mukaidono in [18], all those methods are different kind of regularization [26] of the local optima problem of CM. Nevertheless, even with these methods we have no guarantee of finding the optimal solution of the problem of clustering.

In order to overcome this problem, in this paper we present a novel clustering algorithm based on the application of a global search technique based on an Evolution Strategy (ES) [19, 25, 1] to the minimization of the objective function of the C-Means Algorithm [6].

Evolution Strategies are a class of methods for continuous parameter optimization problems founded on the model of organic evolution. In this paper we present a novel clustering algorithm based on the application of a (μ, λ)-**ES** to the search for the global minimum of the classical C-Means (CM) objective function [6, 3]. The new Evolution Strategy based C- Means (ESCM) algorithm is applied to the clustering step of an interactive system for the segmentation of multimodal medical volumes [22].

This computer-based system supports the clinical oncologist in the tasks

of delineating the volumes to be treated by radiotherapy and surgery, and of quantitatively assessing (in terms of tumor mass or detection of metastases) the effect of oncological treatments. In order to aggregate voxels with similar properties in the different diagnostic imaging volumes, clustering is performed in a multidimensional space where each independent dimension is a particular volumetric image. Clustering algorithms can point out clusters of close voxels in that multidimensional feature space representing the probability distribution of intensities in the different modalities, and therefore sets of voxels with similar intensity values can be defined within the whole multimodal medical volume. These sets of voxels can then be used to delineate regions of interest, that is to make a segmentation of the multimodal volumetric image. In this application clustering supports an inference process based on complementary information carried by each image (e.g. functional or anatomical), each of them considered as an independent dimension of the input space, in order to extract regions corresponding to the different anatomical and/or pathological tissues. A quantitative comparison of segmentation results obtained by the original CM and by the new algorithm is reported in the paper.

The paper is organized as follows. The next section introduces the C-Means following the parametric learning framework. In Sect.s 3 and 4 we give some material on Evolution Strategies and we present a novel application of them to the clustering. In Sect. 5 we set clustering as the basic step of an inference process that, starting from raw data, mines region of interest in multimodal medical volumes. In Sect. 6, we present an experimental comparison of the application of the CM and of the new clustering algorithm to the segmentation of multimodal images. Conclusions are drawn in Sect. 7.

2 Parametric Learning Approach to Clustering

2.1 Maximum Likelihood estimation of cluster parameters

Let $X = \left\{\mathbf{x}_k \mid \mathbf{x}_k \in \mathbf{R}^d,\ k = 1, ..., n\right\}$ be a set of unlabeled random sampled vectors $\mathbf{x}_k = (x_{1k}, ..., x_{dk})$ or *training set*, and $Y = \{\mathbf{y}_j \mid \mathbf{y}_j \in \mathbf{R}^d,\ j = 1, ..., c\}$

be the set of centers of clusters (or classes) ω_j. Following a parametric learning approach, we make the following assumptions:

1. the samples come from a known number of c classes ω_j, $\quad j \in \{1, ..., c\}$;

2. the a priori probabilities $P(w_j)$ (i.e. the probability of drawing patterns of class ω_j from X) are known;

3. the form of class-conditional probabilities densities $p(\mathbf{x} \mid \omega_j, \Theta_j)$ (i.e. the probability density of sample $\mathbf{x}_k$ inside class ω_j) are known, while the vectors of parameters Θ_j are unknown.

Note that the third assumption reduces the clustering problem to the problem of estimation of the vectors Θ_j (*parametric learning*).

In this setting, we assume that samples are obtained by selecting a class ω_j and then selecting a pattern $\mathbf{x}$ according to the probability law $p(\mathbf{x} \mid \omega_j, \Theta_j)$, i.e.:

$$p(\mathbf{x} \mid \Theta) = \sum_{j=1}^{c} p(\mathbf{x} \mid \omega_j, \Theta_j) P(\omega_j) \tag{1}$$

where $\Theta = (\Theta_1, ..., \Theta_c)$. A density function of this form is called a *mixture density* [6], $p(\mathbf{x}_k \mid \omega_j, \Theta_j)$ are called the *component densities*, and $P(\omega_j)$ are called the *mixing parameters*.

A well known parametric statistics method for estimating the parameter vector Θ is based on *maximum likelihood* [6]. It assumes that the parameter vector Θ is fixed but unknown. The likelihood of the training set X is the joint density

$$p(X \mid \Theta) = \prod_{k=1}^{n} p(\mathbf{x}_k \mid \Theta). \tag{2}$$

Then the maximum likelihood estimate $\hat{\Theta}$ is that value of Θ that maximizes the likelihood of the observed training set X.

If $p(X \mid \Theta)$ is a differentiable function of Θ, maximizing the logarithm of the likelihood, we can obtain the following conditions for the maximum-likelihood estimate $\hat{\Theta}_j$:

$$\sum_{k=1}^{n} P\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}\right) \nabla_{\hat{\Theta}_j} log\left(p\left(\mathbf{x}_k \mid \omega_i, \hat{\Theta}_j\right)\right) = 0 \quad \forall j. \tag{3}$$

Moreover, if the a priori class probabilities $P(\omega_j)$ are also unknown, the clustering problem can be faced as the constrained maximization of the likelihood $p(X \mid \Theta)$ over Θ and $P(\omega_j)$ subject to the constraints:

$$P(\omega_j) \geq 0 \qquad \text{and} \qquad \sum_{j=1}^{c} P(\omega_j) = 1. \tag{4}$$

If $p(X \mid \Theta)$ is differentiable and the a priori probabilities estimate $\hat{P}(\omega_j) \neq 0$ for any j, then $\hat{P}(\omega_j)$ and $\hat{\Theta}_j$ must satisfy:

$$\hat{P}(\omega_j) = \frac{1}{n} \sum_{k=1}^{n} \hat{P}\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}\right) \tag{5}$$

and

$$\sum_{k=1}^{n} \hat{P}\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}\right) \nabla_{\hat{\Theta}_j} log\left(p\left(\mathbf{x}_k \mid \omega_j, \hat{\Theta}_j\right)\right) = 0 \tag{6}$$

where

$$\hat{P}\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}\right) = \frac{p\left(\mathbf{x}_k \mid \omega_i, \hat{\Theta}_j\right) \hat{P}(\omega_j)}{\sum_{h=1}^{c} p\left(\mathbf{x}_k \mid \omega_h, \hat{\Theta}_h\right) \hat{P}(\omega_h)}. \tag{7}$$

Let we assume now that the component densities are multivariate normal, i.e.:

$$p\left(\mathbf{x}_k \mid \omega_i, \hat{\Theta}_j\right) = \frac{1}{(2\pi)^{\frac{d}{2}} \mid \Sigma_j \mid^{\frac{1}{2}}} \; exp[-\frac{1}{2}(\mathbf{x}_k - \mathbf{y}_j)^t \Sigma_j^{-1} (\mathbf{x}_k - \mathbf{y}_j)] \tag{8}$$

where d is the dimensionality of the feature space, $\mathbf{y}_j$ is the *mean vector*, Σ_j is the *covariance matrix*, $(\mathbf{x}_k - \mathbf{y}_j)^t$ is the transpose of $(\mathbf{x}_k - \mathbf{y}_j)$, Σ_j^{-1} the inverse of Σ_j, and $\mid \Sigma_j \mid$ the determinant of Σ_j.

In the general case (i.e. $\mathbf{y}_j$, Σ_j, and $P(\omega_j)$ are all unknown) the maximum likelihood principle yields useless singular solutions. As shown by Duda and Hart [6], we can obtain meaningful solutions by considering the largest of the finite local maxima of the likelihood function.

The local-maximum-likelihood estimate for $P(\omega_j)$ is the same as Eq. 5, while

$$\hat{\mathbf{y}}_j = \frac{\sum_{k=1}^{n} \hat{P}\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}_j\right) \mathbf{x}_k}{\sum_{k=1}^{n} \hat{P}\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}_j\right)} \tag{9}$$

Table 1: C-Means (CM) Algorithm.

1. assign the number of clusters and the tolerance ϵ_1 for the stop criterion;
2. initialize the centers of clusters;
3. **do until** any center changes less than ϵ_1;
   (a) assign the samples to the clusters with smaller Euclidean distance using Eq.s 12 and 14;
   (b) recalculate the centers using Eq. 9;
4. **end do.**

$$\hat{\Sigma}_j = \frac{\sum_{k=1}^{n} \hat{P}\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}_j\right)(\mathbf{x}_k - \hat{\mathbf{y}}_j)(\mathbf{x}_k - \hat{\mathbf{y}}_j)^t}{\sum_{k=1}^{n} \hat{P}\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}_j\right)} \tag{10}$$

where (from Eq.s 7, and 8)

$$\hat{P}\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}_j\right) = \frac{\mid \hat{\Sigma}_j \mid^{-\frac{1}{2}} exp[-\frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{y}}_j)^t \hat{\Sigma}_j^{-1}(\mathbf{x}_k - \hat{\mathbf{y}}_j)]\ \hat{P}(\omega_j)}{\sum_{h=1}^{c} \mid \hat{\Sigma}_h \mid^{-\frac{1}{2}} exp[-\frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{y}}_h)^t \hat{\Sigma}_h^{-1}(\mathbf{x}_k - \hat{\mathbf{y}}_h)]\ \hat{P}(\omega_h)}. \tag{11}$$

The set of Eq.s 5, 9, 10, and 11 can be interpreted as a gradient ascent or hill-climbing procedure for maximizing the likelihood procedure. A Lloyd-Picard iteration can start with Eq. 11 using initial estimates to evaluate Eq. 11 for $\hat{P}\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}_j\right)$ and then using Eq.s 5, 9, and 10 to update the estimates.

Like all hill-climbing procedures the results of this iteration do depend upon the starting point, and, moreover, the inversion of $\hat{\Sigma}_j$ is quite time consuming, and there is the possibility of multiple solutions.

2.2 C-Means (CM) Algorithm

An efficient implementation of the previous procedure is based on the following approximation of Eq. 11:

$$P\left(\omega_j \mid \mathbf{x}_k, \hat{\Theta}_j\right) = \begin{cases} 1 & if\ D_j\left(\mathbf{x}_k\right) = \min_{1 \le j \le C} D_j\left(\mathbf{x}_k\right) \\ 0 & otherwise \end{cases} \tag{12}$$

where $D_j\left(\mathbf{x}_i\right)$ is a *local cost function* or *distortion* measure and in many cases can be assumed as the *scaled Mahalanobis distance* $\mathcal{M}_j(\mathbf{x}_k)$,

$$\mathcal{M}_j^2(\mathbf{x}_k) \equiv \mid \Sigma_j \mid^{1/d} (\mathbf{x}_k - \hat{\mathbf{y}}_j)^t \Sigma_j^{-1} (\mathbf{x}_k - \hat{\mathbf{y}}_j). \tag{13}$$

This observation is the rationale of the *C-Means* (CM), also named *Basic Isodata* algorithm [6] and *Hard C-Means* [3]. It is worth noting that the usage of the Mahalanobis distance still involves a heavy computational overhead. In many implementations of CM a strong approximation of $D_j\left(\mathbf{x}_k\right)$ is adopted, using the *Euclidean distance* $\mathcal{E}_j(\mathbf{x}_k)$

$$\mathcal{E}_j(\mathbf{x}_k) \equiv \mid\mid \mathbf{x}_k - \hat{\mathbf{y}}_j \mid\mid . \tag{14}$$

The resulting CM algorithm is an efficient approximate way to obtain the maximum likelihood estimate of the centers of clusters [6].

One implementation of the CM using the Euclidean distance is illustrated in Tab. 1. In this algorithm the initialization of the number of clusters (Step 1) is performed by using the a-priori knowledge on the problem. At Step 2, the position of centers of clusters can be initialized either using a-priori knowledge or at random in the d-dimensional hyperbox $\mathbf{I}$:

$$\mathbf{I} = \Pi_{i=1}^{d} \left[min_k(x_{ik}), max_k(x_{ik})\right], \quad \mathbf{I} \subset \mathbf{R}^d \tag{15}$$

As demonstrated by Bezdek [3], the CM, while maximizes the likelihood of the training set, minimizes at the same time a *global error function* J_w defined as the expectation of the squared local cost function:

$$J_w \equiv < D^2 > = \sum_{k=1}^{n} \sum_{j=1}^{c} u_{jk} D_j^2(\mathbf{x}_k) \tag{16}$$

where $u_{jk} \equiv P\left(\omega_j \mid \mathbf{x}_k\right)$ or, in general, a membership value of pattern $\mathbf{x}_k$ $(k = \{1, ..., n\})$ to cluster ω_j $(j = \{1, ..., c\})$.

The CM, while is an efficient approximation of the maximum likelihood procedure for estimating the centers of clusters, shows some intrinsic problems. In particular, it is subject to the problem of trapping in local minima of J_w (i.e. on the local maxima of the likelihood).

This locality in searching for minima is its main limitation, in particular when we try to apply this algorithm as the basis for inference procedures.

In order to overcome these problems, many attempt, based on different fuzzy clustering paradigms, have been proposed in the literature. The most popular fuzzy clustering method is the Fuzzy C-Means algorithm by Bezdek [3] that is based on the constrained minimization of a generalization of the CM global error expectation. We cite also the technique proposed by Rose et al [20] based on the maximum entropy principle [9] and using a Deterministic Annealing technique, and the Possibilistic C-Means algorithm by Krishnapuram and Keller [12, 13].

In [18], Miyamoto and Mukaidono showed that the Fuzzy C-Means [3], and the maximum entropy methods correspond to different types of application of the regularization theory to the CM in order to reduce the problem of local minima.

An alternative approach to the solution of the local minima problem of CM can be based on the application of global search techniques. In [5] we propose a global search method for the minimization of J_w based on the Simulated Annealing technique [11]. In next sections we shall present some search techniques based on Evolution Strategies, that will be applied to clustering problem.

3 Evolution Strategies

Evolutions Strategies (ES) [19, 25, 1] are a class of Evolutionary Computation methods for continuous parameter optimization problems founded on the model of organic evolution. During each generation (iteration of the ES algorithm) a population of individuals (potential solutions) is evolved to produce new solutions. Only the highest-fit solutions survive to become parents for the next generation.

In biological terms, the genetic encoding for an individual is called *genotype*. New genotypes are created from existing ones by modifying the genetic material. The interaction of a genotype with its *environment* induces an observed response called *phenotype*.

Reproduction takes place at the genotype level, while survival is determined at the phenotype level. Only highly fit individual survive and reproduce in future generations.

Individuals in the population are composed by *object variables* and *strategy parameters*. In basic ES, an individual is represented as a vector

$$\mathbf{a} = (x_1, ..., x_n, \sigma_1, ..., \sigma_n) \in \Re^{2n} \tag{17}$$

consisting of n object variables and their corresponding n standard deviations for individual mutations.

There are two variants of an ES. The *multi-membered ES plus strategies* (denoted as $(\mu + \lambda)$**-ES**) and the *multi-membered ES comma strategies* (denoted as (μ, λ)**-ES**). In $(\mu + \lambda)$**-ES** μ parents create $\lambda \geq \mathbf{1}$ offspring individuals by means of recombination and mutation. The μ best parents and offspring are selected to form the next population. For a (μ, λ)**-ES**, with $\lambda > \mu \geq \mathbf{1}$, the μ best individuals are selected from offspring only. We shall discuss now the ES operators, i.e. recombination, mutation, and selection.

3.1 Recombination

Recombination (or crossover) in ES is performed on individuals of the population. The most used recombination rules are:

1. *no recombination*;
2. *discrete recombination*: the components of two parents are selected at random from either the first or the second parent to form an offspring individual;
3. *intermediate recombination*: offspring components are somewhere between the corresponding components of the parents;
4. *global and discrete recombination*: one parent is selected and fixed and for each component a second parent is selected anew from the population to determine the component values using discrete recombination;
5. *global and intermediate recombination*: one parent is selected and fixed and for each component a second parent is selected anew from the population to determine the component values using intermediate recombination.

The recombination operator may be different for object variables and strategy parameters.

3.2 Mutation

For mutations each x_j is mutated by adding an individual, $(0, \sigma_j)$-normally distributed random number. The σ_j themselves are also subject to mutation and recombination (*self-adaptation* of strategy parameters [24]), and a complete mutation step $m(\mathbf{a}) = \mathbf{a}'$ is obtained by the following equations:

$$s = exp(N(0, \tau)) \tag{18}$$

$$\sigma_j' = \sigma_j \cdot \exp(N_j(0, \tau')) \cdot s \tag{19}$$

$$x_j' = x_j + N_j(0, \sigma_j') \tag{20}$$

Mutation is performed on the σ_j by multiplication with two log-normally distributed factors, one individual factor, sampled for each σ_j ($\tau' = 1/\sqrt{2\sqrt{n}}$), and one common factor s ($\tau = 1/\sqrt{2n}$), sampled once per individual. This way, a scaling of mutations along the coordinate axes can be learned by the algorithm itself, without an exogenous control of the σ_j.

More sophisticated ES using so-called *correlated mutation* are presented in [1].

3.3 Selection

Selection for survival is completely deterministic, as it is only based on the rank of fitness. It is called also an *extinctive selection*, as $\lambda - \mu$ worst individuals are definitively excluded from contribution offspring to the next generation.

It is worth noting that $(\mu + \lambda)$**-ES** is elitist and therefore, while performance is monotonously improved, the implemented search is local and unable to deal with changing environment.

On the contrary, (μ, λ)**-ES** enables the search algorithm to escape from local optima, to follow a moving optimum, to deal with noisy objective function, and to self adapt strategy parameters effectively. The ratio μ/λ is named the *degree of extinctiveness* and is linked to the probability to locate the global optimum. If it is large there is a high convergence reliability, whereas if it is small there is a high convergence velocity. Investigations presented in [24] suggest an optimal ratio of $\mu/\lambda = \mathbf{1/7}$.

Table 2: Evolution Strategy based C-Means (ESCM) algorithm.

1. assign μ, λ, the number of clusters, and the threshold ϵ_2;
2. initialize the population;
3. evaluate J_w for each individual (Eq. 16);
4. **do until** $\Delta J_w^{best}/J_w^{best}$ is greater than ϵ_2;
5. **count1**=0;
   (a) **while count1** less then μ;
      i. **count1**++;
      ii. select by rank two individuals for mating;
      iii. order consistently the centers of clusters in both selected individuals using algorithm RI (Tab. 3);
      iv. crossover object variables (discrete recombination);
      v. crossover strategy parameters (intermediate recombination);
      vi. mutate individual as shown in Sect. 3.2;
   (b) **end do**;
   (c) evaluate J_w for each individual (Eq. 16);
   (d) select the μ fittest individuals for next population;
6. **end do**.

4 Evolution Strategy based C-Means (ESCM) algorithm

In order to overcome the limits of C-Means, a (μ, λ)-**ES** can be used to find the global optimum of J_w (Eq. 16).

Tab. 2 illustrates the Evolution Strategy based C-Means (ESCM) algorithm. Each genotype $\mathbf{a}$ is a list containing the object variables (i.e. the centers of clusters) and the strategy parameters:

$$\mathbf{a} = (\mathbf{y}_1, ..., \mathbf{y}_c, \sigma_1, ..., \sigma_c) \tag{21}$$

where c is the number of clusters. ESCM works in a $(c \times (d+1))$-dimensional space, where d is the dimension of the pattern space.

After the initialization of parameters (step 1), the population is initialized (step 2) in the following way: Centers of clusters (i.e. object variables) are initialized at random in the hyperbox $\mathbf{I}$ (Eq. 15), while strategy parameters are initialized at random in the range $[0, \alpha]$, where α is order of $1/10$ the side of $\mathbf{I}$.

The remaining steps are quite standard for an (μ, λ)-**ES**, with the exception of Step 5(A)iii. In fact we must note that, before mixing object variables of parents (centers of clusters) using discrete recombination crossover, they must be re-indexed, in such a way centers with same index are likely to correspond to the same cluster. The re-indexing algorithm is described in Tab. 3 and is modified by the RL algorithm proposed in [27]. Besides, the stop condition (Step 4)

$$\frac{\Delta J_w^{best}}{J_w^{best}} < \epsilon_2 \tag{22}$$

is based on the ratio of normalized difference of objective function J_w evaluated on the fittest individual of two successive generations.

In principle, ESCM allows us to avoid local minima of J_w and to find the global optimum, improving in this way the reliability of inferential tasks associated to the clustering procedure.

Moreover it is simple to create variants of the basic ESCM. For instance, if we want to reduce the interference of big blobs to the localization of the centers of small clusters, it is straightforward to change in the algorithm J_w with the following *scaled global error function* J_s:

Table 3: Re-indexing (RI) algorithm.

1. compile the matrix of distances M among centers of clusters of the two individuals;
2. **count2**=0;
3. **while count2** less than c;
   (a) **count2**++;
   (b) find the minimal item of the matrix;
   (c) assign the same index to both centers of clusters in the two individuals;
   (d) delete the corresponding row and column in the matrix of distances M;
4. **end do**.

$$J_s \equiv \sum_{j=1}^{c} \frac{1}{C_j} \sum_{k=1}^{n} u_{jk} D_j^2(\mathbf{x}_k), \tag{23}$$

where C_j is the cardinality of cluster w_j.

5 Segmentation of multimodal medical volumes

5.1 Multimodal medical volumes (MMV)

Medical images are obtained by different acquisition modalities, including X-ray tomography (CT), magnetic resonance imaging (MRI), single photon emission tomography (SPECT), and positron emission tomography (PET), ultrasounds (US), etc. [15]. Multimodal volumes can be derived from sets of such different diagnostic volumes by spatial coregistration of volumes in order to fully correlate complementary information (e.g., structural and functional) about the same patient.

The visual inspection of a large set of such volumetric images permits only partially to the physician to exploit the available information. Therefore, computer-assisted approaches may be helpful in the clinical oncological environment as support to diagnosis in order to delineate volumes to be treated by radiotherapy and surgery, and to assess quantitatively (in terms of tumor mass or detection of metastases) the effect of oncological treatments.

The extraction of such volumes or other entities of interest from imaging data is named *segmentation* and is usually performed, in the image space, by defining sets of voxels with similar features within a whole multimodal volume.

5.2 Clustering-based inference approach to MMV segmentation

It is worth noting that it is very difficult or impossible to settle the solution of the multimodal volumes segmentation problem in a reliable rule based systems framework, as physicians are hardly able, at least for low level steps in image analysis, to describe the rationale of their decisions. Moreover, for higher level in image analysis, rationales of physicians, even if more precise, strongly depend on many factors, such as different clinical frameworks, different anatomical areas, different theoretical approaches, etc.

Inference procedures based on learning from data must be then employed for design a computer-assisted systems for segmenting multimodal medical volumes.

Actually, in such data based systems, a possible supervised approach has two major drawbacks:

- it is very time-consuming (especially for large volumes), as it requires the labeling of prototypical samples needed for applying the generalization process. Even if the number of clusters is predefined, a careful manual labeling of voxels in the training set belonging with certainty to the different clusters is not trivial, especially when it concerns multimodal data sets and

- heavy biases may be introduced by physicians unskilled or fatigued due to the large inter-user and intra-user variability generally observed when manual labeling is performed.

On the contrary, unsupervised methods may fully exploit the implicit multidimensional structure of data and make clustering of the feature space independent from the user's definition of training regions [2, 8] due to their self-organizing approach.

A multimodal volume may be defined by the spatial registration of a set of d different imaging volumes. As a consequence, its voxels are associated with an array of d values, each representing the intensity of a single modality in a voxel. From another point of view, the d different intensity values related to the voxel in such multimodal volume can be viewed as the coordinates of the voxel within a d-dimensional feature space where multimodal analysis can be made.

An image space (usually 3D) defined by the spatial coordinates of the data set, and a multidimensional feature space, as described before, must be considered for a more complete description of the segmentation problem. The interplay between these two spaces turns out to be very important in the task of understanding the data structure.

Actually, the definition of clusters within the above described d-dimensional feature space and the classification of all the voxels of the volumes to the resulting classes are the main steps in segmenting multimodal volumes.

This approach, where an inference process based on clustering constitutes the principal procedure for the MMV segmentation, has been followed in many recent papers [4, 22, 17, 10, 14], and it has been shown to be more robust to noise in discrimination of different tissues than techniques based on edge detection [4].

Nevertheless, the used clustering method itself must be well founded in statistics and must be not limited by intrinsic problems, such as the problem of local optima in CM.

Moreover, many bias effects must be taken into account in considering clustering for the segmentation of medical images. Actually, very heterogeneous clusters may be found in the feature space, with very different probability densities, and considering the cardinality of clusters may be necessary in order to include in the analysis the statistical nature of the data set. Furthermore, the partial volume effect during acquisition may produce a really intrinsic ambiguity of borders between regions of interest. As a consequence, unsupervised clustering based segmentation of medical images emerges as a very difficult task, whose usefulness is related to the balance of two conflicting actions, namely, the elimination of noise and redundancy from original images and the preservation of significant information in the segmented im-

age. These constraints may force users to introduce their knowledge in the sequence of analysis and further refinements are often needed in order to obtain meaningful and affordable results.

5.3 Interactive segmentation system

From all these considerations a correct architecture for a computer based system for multimodal medical volumes segmentation should include a computational core grounded on unsupervised clustering together with powerful interactive tools for knowledge based refinements that physicians could tune and organize to specific diagnostic tasks to be performed. This way, as requested in the clinical practice, physicians can stay in control both of the sequence of choices and of the results in the analysis process in order to introduce in the segmentation process their theoretical and heuristic knowledge.

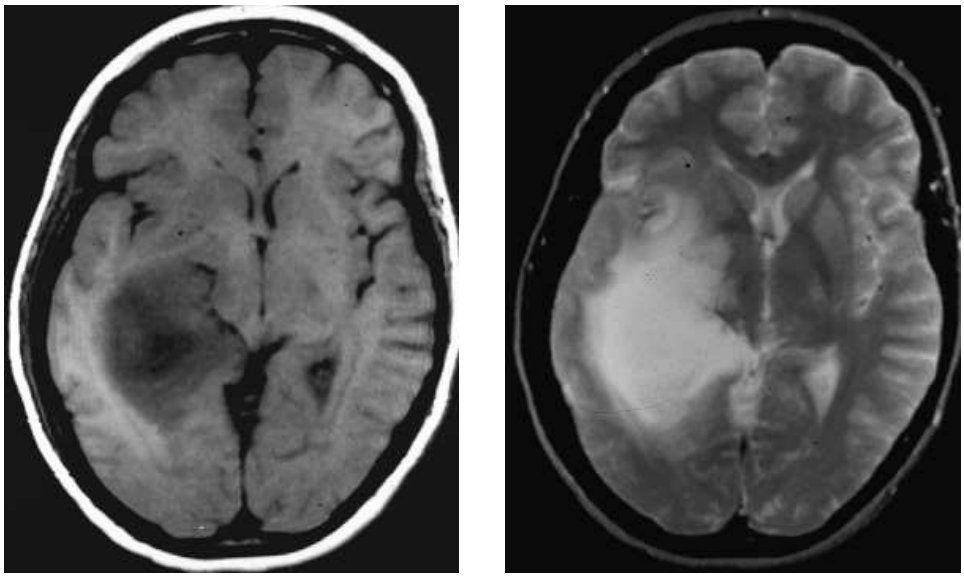
A system based on those assumptions has been developed by our group and is described in [22]. It is an interactive system with a friendly Graphics User Interface, and supporting a full sequence of analysis of multimodal medical images. The main functions performed by this system are: Feature extraction, dimensionality reduction, unsupervised clustering, voxel classification, and intra- and post-processing refinements.

The main component of this system is the clustering subsystem that make possible to run in the feature space alternative clustering algorithms, including the C-Means [6], the Capture Effect Neural Network [7], Fuzzy C-Means [3], the Deterministic Annealing [20, 21], and the Possibilistic C-Means [12, 13]. In [16, 17] we report some comparisons of application of such algorithms on clinical images.

6 Experimental analysis

6.1 Data set

We have implemented the Evolution Strategy based C-Means (ESCM) algorithm as a clustering module of the previously described graphical interactive system supporting the full sequence of analysis of multimodal medical volumes.



(a) (b)

Figure 1: T1-weighted (a) and T2-weighted (b) MRI images of a patient with glioblastoma multiforme in the right temporal lobe.

In order to illustrate in a specific case the inference task of MMV segmentation based on clustering, and to show the gain in precision and reliability obtained in this task using the ESCM instead of the original CM, let we consider now a simple data set consisting of a multimodal transverse slice of the head (Fig. 1) composed by spatially correlated T1-weighted and T2-weighted MRI images from an head acquisition volume of an individual with glioblastoma multiforme.

The images are 288 x 362 with 256 gray levels. The tumor is located in the right temporal lobe and appears bright on the T2-weighted image and dark on the T1-weighted image. A large amount of edema is surrounding the tumor and appears very bright on the T2-weighted image. The lower signal area within the mass suggests tissue necrosis. Each pixel in the above defined two-modal slice is associated to an array of two intensity values (T1 and T2). Therefore, each of these couples of pixel intensity is represented by a point in a 2D feature space (Fig. 2), whose coordinates represent the intensity values in that pixel of each modality belonging to the multimodal set. The segmentation task consists in finding the main classes in this feature space and in associating each pixel in image to one of this classes. The main classes

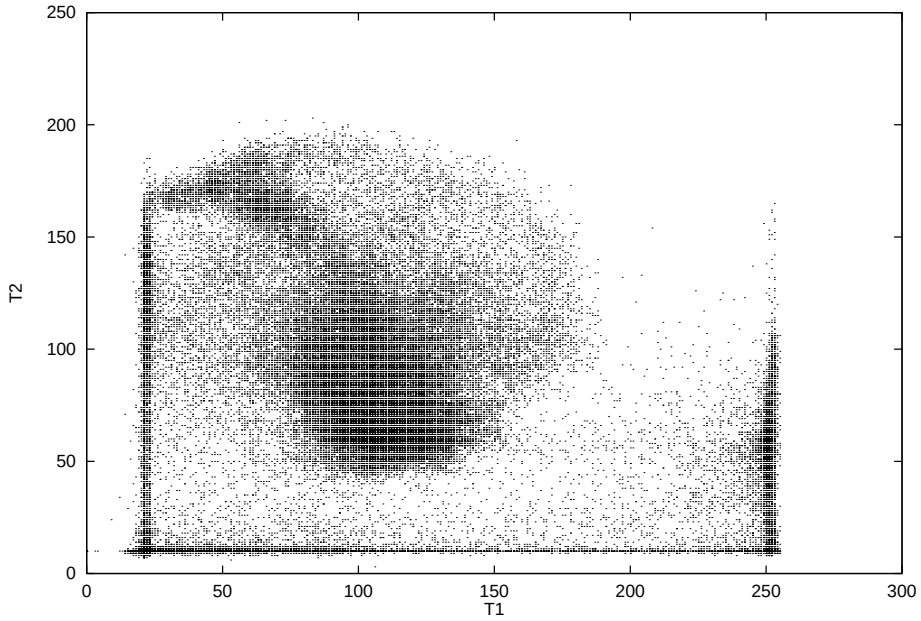


Figure 2: Feature space (T2 versus T1) obtained from the MRI images in Fig. 1.

in the data set are: white matter, gray matter, cerebro spinal fluid (CSF), tumor, edema, necrosis, scalp. A slight mis-registration between images may be responsible of some mis-classification errors in final results.

6.2 Methods

We give here some information on the implementation of clustering algorithms used in the experimental analysis.

- The CM uses 7 clusters and a tolerance for the stop criterion $\epsilon_1 = .01$, centers of clusters are initialized at random, and convergence is noticed in 10-15 fast iterations.
- For the ESCM using J_w, according to the $\mu/\lambda = \mathbf{1/7}$ rule proposed by Schwefel [24], we selected $\mu = 10$ and $\lambda = 70$. Moreover, we initialized $c = 7$, $\epsilon_2 = .005$, and the centers of clusters at random. We implemented the selection by rank using a linear probability distri-

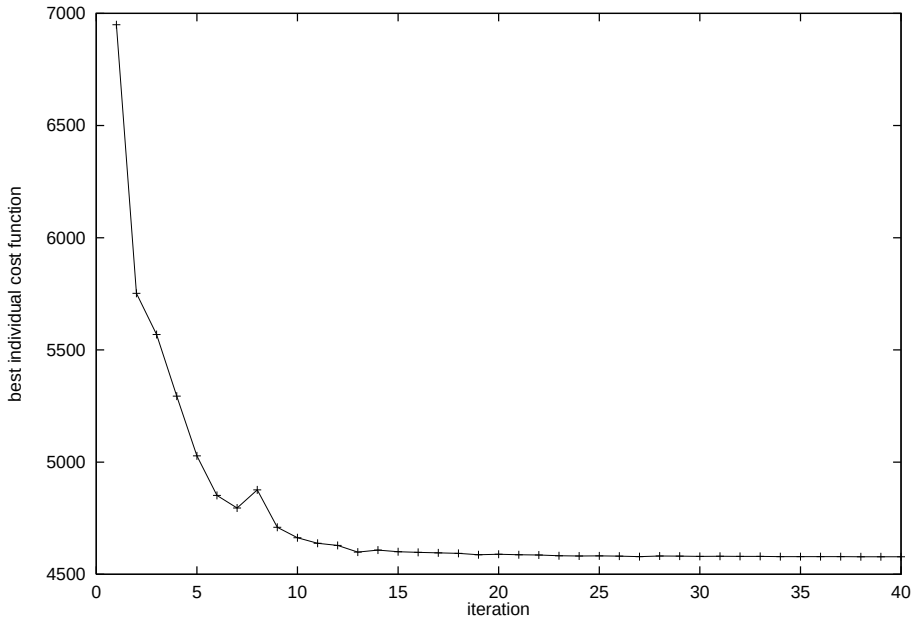


Figure 3: Cost function of best individuals versus iteration of ESCM.

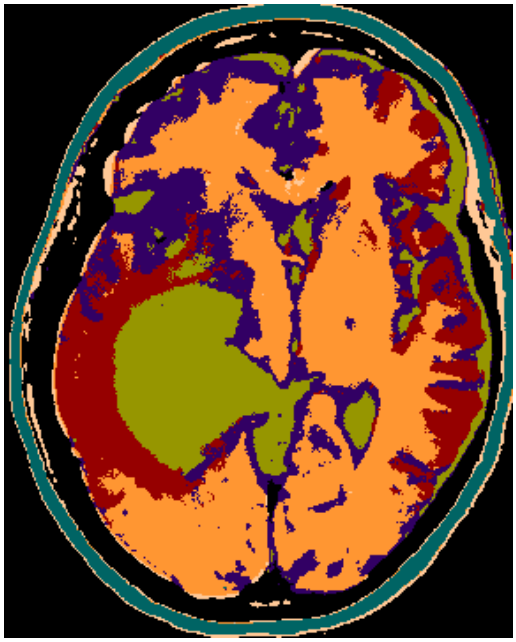


Figure 4: Segmentation obtained by the CM algorithm with 7 clusters.

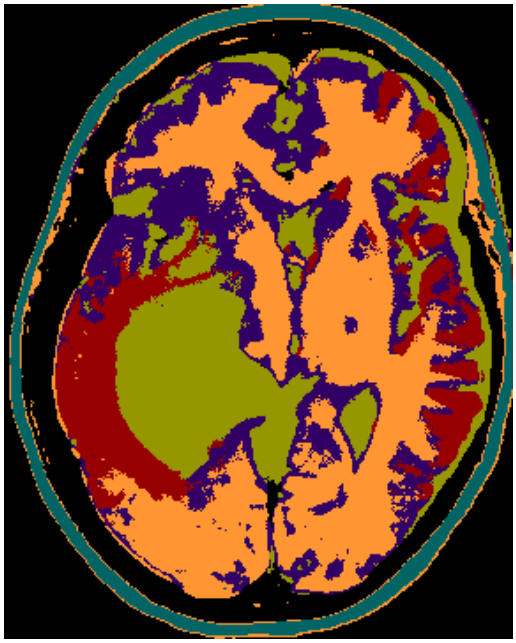


Figure 5: Segmentation obtained by the ESCM algorithm using J_w and with 7 clusters.

bution with negative slope, while the intermediate recombination is implemented as the average of components of parents.

- The implementation of ESCM using J_s is identical to the previous one, with the obvious exception of the objective function. A typical plot of J_s^{best} is presented in Fig. 3. Using $\Delta J_s^{best}/J_s^{best} \leq \epsilon_2$ as the stop condition, the ESCM ends in 15 iteration.

6.3 Results and Discussion

Let us compare the results produced by the ESCM clustering algorithm and by the standard C-Means (CM) algorithm.

In Fig. 4 the results of the unsupervised segmentation with the CM algorithm are shown. CM almost correctly defines scalp and white matter. Nevertheless it produces mistakes in classification of gray matter and edema in the left side of brain, and especially is not able to separate tumor, necrosis and CSF. Similar results are obtained by the basic ESCM with the standard cost function J_w (Fig. 5). Nevertheless, as an important difference, from

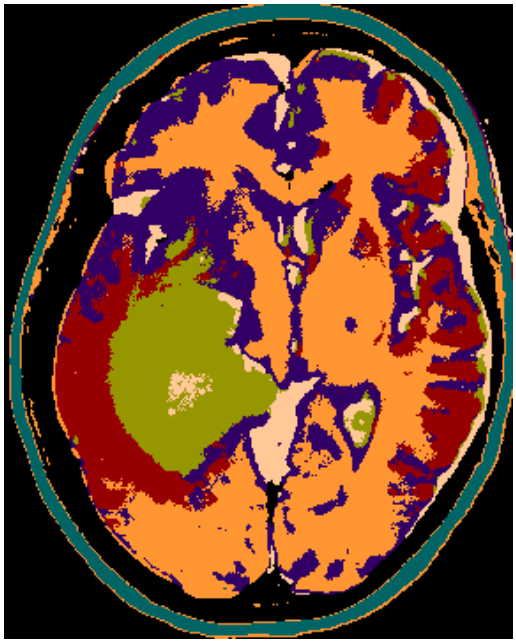


Figure 6: Segmentation obtained by the ESCM algorithm using J_s and with 7 clusters.

a large number of tests, ESCM results to be largely more stable than CM with respect to the positions of centroids and to the extension of clusters in the feature space. Eventually, by using the newly defined *scaled global error function* J_s to take into account the cardinality of clusters, the results of ESCM (Fig. 6) dramatically improve. Actually, we may notice that, in comparison with CM, and with the basic version of ESCM, the final version of ESCM correctly distinguishes between tumor and CSF, and within the tumor region is able to find the necrosis region. Correct definition of scalp and white matter and misclassification in the left side of the brain remains as from CM.

7 Conclusions

The C-Means (CM) [6], while is an efficient approximation of the maximum likelihood procedure for estimating the centers of clusters, shows some intrinsic problems. In particular, it is subject to the problem of trapping in local minima of its objective function J_w (Eq. 16). This locality in search-

ing for minima is a main limitation, in particular when we try to apply this algorithm as the basis for inference procedures.

In order to overcome the limits of C-Means, we have proposed in this paper a novel clustering algorithm based on the application of an Evolution Strategy (ES) [19, 25, 1] to the search for the global minimum (Evolution Strategy based C-Means or ESCM algorithm).

The ESCM is based on a (μ, λ)-**ES** strategy where the object variables of genotypes are the centers of clusters. The implementation of the (μ, λ)-**ES** strategy is quite standard, but before mixing object variables of parents using discrete recombination crossover, they are re-indexed, in such a way centers with same index are likely to correspond to the same cluster.

It is worth noting that it is easy to make variants to the basic ESCM. For instance, with the straightforward change of J_w with the scaled global error function J_s (Eq. 23) it is possible to reduce the interference of big blobs to the localization of the centers of small clusters.

In this paper we considered a complex inference processes based on clustering consisting in multimodal medical volumes (MMV) segmentation. This approach has been shown to be very robust to noise and able to process complementary information carried by each image (e.g. functional or anatomical) [4]. In this inference task, devoted to aggregate voxels with similar properties (corresponding to the different anatomical and/or pathological tissues) in the different diagnostic imaging volumes, clustering is performed in a multidimensional space where each independent dimension is a particular volumetric image. Nevertheless, the used clustering method itself must be well founded in statistics and must be not limited by intrinsic problems, such as the problem of local optima in CM. Moreover, many bias effects (due, e.g., to heterogeneous clusters and to partial volume effect during acquisition) must be taken into account in considering clustering for the segmentation of medical images.

We have implemented the ESCM algorithm as a clustering module of the previously described graphical interactive system supporting the physician for the full sequence of analysis of multimodal medical volumes.

In the experimental results presented in the paper, we have compared the segmentation obtained by the application of CM, ESCM using J_w and ESCM using J_s to a simple data set consisting of a multimodal transverse slice of the head (Fig. 1) composed by spatially correlated T1-weighted and T2-weighted MRI images from an head acquisition volume of an individual with glioblastoma multiforme.

The two implementations of ESCM give more stable solutions than CM with respect to the positions of centroids and the extension of clusters in the feature space. In particular, the ESCM using J_s, as is able to take into account the cardinality of clusters, dramatically improves the quality of segmentation results.

Acknowledgments

The images are from the BrighamRAD Teaching Case Database of the Department of Radiology at Brigham and Women's Hospital in Boston.

References

[1] T. Baeck. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithm.* Oxford University Press, 1996.

[2] A.M. Bensaid, L.O. Hall, L.P. Clarke, and R.P. Velthuizen. MRI segmentation using supervised and unsupervised methods. In *Proc. 13th IEEE Eng. Med. Biol. Conf.*, pages 483–489, Orlando, 1991. IEEE.

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

[4] J.C. Bezdek, L.O. Hall, and L.P. Clarke. Review of MR image segmentation techniques using pattern recognition. *Med. Phys.*, 20:1033–1048, 1993.

[5] P. Bogus, A. Massone, and F. Masulli. A Simulated Annealing C-Means Clustering Algorithm. In F. Masulli and R. Parenti, editors, *Proceeding of SOCO'99 ICSC Symposium on Soft Computing - Genova*, pages 534–540, Millet, Canada, 1999. ICSC.

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

[7] F. Firenze and P. Morasso. The capture effect model: a new approach to self-organized clustering. In *Sixth International Conference. Neural Networks and their Industrial and Cognitive Applications. NEURO-NIMES*

93 Conference Proceedings and Exhibition Catalog, pages 65–54, Nimes, France, 1993.

[8] G. Gerig, J. Martin, R. Kikinis, O. Kubler, M. Shenton, and F.A. Jolesz. Unsupervised tissue type segmentation of 3D dual-echo MR head data. *Im. Vis. Comput.*, 10:349–360, 1992.

[9] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.

[10] Z-X Ji, Q-S Sun, D-S Xia. A modied possibilistic fuzzy c-means clustering algorithm for bias field estimation and segmentation of brain MR image. *Computerized Medical Imaging and Graphics* 35 38–397, 2011.

[11] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:661–680, 1983.

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

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

[14] H. Mahmoud, F. Masulli, S. Rovetta. A Fuzzy Clustering Segmentation Approach for Feature-Based Medical Image Registration. In *Proc. CIBB 2012, 9-th International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics*, Houston, Tx, USA, CIBB Proceedings Series, ISBN 978-88-906437-1-2, 2012.

[15] M.N. Maisey and et al. Synergistic imaging. *Eur. J. Nucl. Med.*, 19:1002–1005, 1992.

[16] F. Masulli, P. Bogus, A. Schenone, and M. Artuso. Fuzzy clustering methods for the segmentation of multivariate images. In M. Mares, R. Mesia, V. Novak, J. Ramik, and A. Stupnanova, editors, *Proceedings of the 7th International Fuzzy Systems Association Word Congress IFSA'97*, volume III, pages 123–128, Prague, 1997. Academia.

[17] F. Masulli and A. Schenone. A fuzzy clustering based segmetation system as support to diagnosis in medical imaging. *Artificial Intelligence in Medicine* 16, 129–147, 1999.

[18] S. Miyamoto and M. Mukaidono. Fuzzy C-Means as a regularization and maximum entropy approach. In M. Mares, R. Mesia, V. Novak, J. Ramik, and A. Stupnanova, editors, *Proceedings of the 7th International Fuzzy Systems Association Word Congress IFSA'97*, volume III, pages 86–91, Prague, 1997. Academia.

[19] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog Verlag, Stuttgart, 1973.

[20] K. Rose, E. Gurewitz, and G. Fox. A deterministic annealing approach to clustering. *Pattern Recognition Letters*, 11:589–594, 1990.

[21] K. Rose, E. Gurewitz, and G. Fox. Constrained clustering as an optimization method. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 15:785–794, 1993.

[22] A. Schenone, F. Firenze, F. Acquarone, M. Gambaro, F. Masulli, and L. Andreucci. Segmentation of multivariate medical images via unsupervised clustering with adaptive resolution. *Computerized Medical Imaging and Graphics*, 20:119–129, 1996.

[23] K. E. Parsopoulos, M. N. Vrahatis. Recent approaches to global optimization problems through Particle Swarm Optimization. *Natural Computing* 1(2-3):235-306, 2002.

[24] H.P. Schwefel. Collective phenomena in evolutionary systems. In *Preprints of the 31-th annual meeting of the International Society for General Systems Research*, volume 2, pages 1025–1033, Budapest, 1988.

[25] H.P. Schwefel. *Evolution and Optimum Seeking*. Wiley, 1995.

[26] A. Tikhonov and V. Arsenin. *Solutions of ill-posed problems*. Winston and Sons, New York, 1997.

[27] E.C.K. Tsao, J.C. Bezdek, and N.R. Pal. Fuzzy Kohonen clustering networks. *Pattern Recognition*, 27(5):757–764, 1994.