Comparison between Algorithms of MRI Image Segmentation

Hayfa Masghouni

Abstract: ***In this article, we present different algorithms of MRI image segmentation based on classification of pixels. First, we present FCM (Fuzzy C_Means) and its different extensions with a comparison between them, after we present GMM (Gaussian Mixture Model) and EM (Expectation Maximization) and its extensions with a comparison between them.***

Index Terms: FCM, EM, GMM, MRI image segmentation

I. INTRODUCTION

The image segmentation consists on grouping the pixels into regions according to predefined criteria. This operation is used in various sectors like indexation, medical sector, object recognition......In this article, we focalise on its use in MRI.EM (Expectation Maximization) and FCM (Fuzzy C_Means) are the most popular algorithms based on classification.[1]The Gaussian Mixture Model (GMM) is the most known segmentation method which use EM to estimate the model's parameters.[1]

II. FCM AND ITS EXTENSIONS

FCM is based on a fuzzy classification of pixels.

A. Fuzzy classification

In image processing, the fuzzy set theory is very used to enable the management of uncertainty (due to a superposition of different signals from different tissues) and inaccuracy (due to partial volume effects in MRI) also it's used in pattern recognition, segmentation, filtering, resetting and for more advanced studies such as fuzzy topology and fuzzy distance between objects.

Mr. Roulier Vincent [7] chooses the diagnosis assistance systems based on the theories of fuzzy sets for the following reasons :

- This theory is used to represent qualitative, imprecise and uncertain information(principle exploited in the quantification of liver steatosis on histological slides)
- The inference is done by logic rules that can consider various concepts for deductive reasoning
- Recent studies allow to consider the significant uncertainties and inaccuracies (often presented on MRI and histological images)

There are many applications of the theory of fuzzy sets in medical fields, among these applications we have :

- Methods of classification of multimodal imaging in tissues using fuzzy logic and modeling.
- The use of fuzzy sets in the decision for the diagnosis of prostate cancer.

The use in the imaging of nerve fibers in the brain, fuzzy set theory enables the segmentation and the tracking of myelin fibers in different imaging modalities (MRI, in polarized light microscopy and confocal microscopy).

B. Unsupervised Fuzzy classification

According to Mrs. Hakima Zouaoui and Mr Moussaoui Abdelouahab,[12] the principle of this classification is that an item to classify doesn't belong to a specific class as in classic methods, but it belongs to each class with a degree of membership in the interval [0,1] . The advantage of this method is to consider the smooth transitions between classes and the ambiguity or partial membership in multiple classes. The net classification is a special case where the membership's degree is equal to 1 for one class and is equal to zero for the others.

According to gentlemen Michel Ménard, Vincent Courboulay and Pierre-André Dardignac,[14] there exist two approaches of classifications :

- The probabilistic approach proposed by Bezdek in 1981 which ensures the convergence of the classification process
- The possibilistic approach proposed by Krishnapuram in 1993 which allows more accurate convergence of the classification process than the probabilistic approach

C. Probabilistic approach : algorithm FCM

According to gentlemen M.A. Balafar, A. R.Ramli and S.Mashohor,[4] this approach is based on the minimization of an objective function.

Dunn was the first to offer an objective function using fuzzy least squares in 1973 to solve this type of problem.

Bezdek extended the objective function of Dunn by adding a blur factor m, and then proposed the algorithm of fuzzy c-means (FCM: Fuzzy C-Means) in 1981.

- Let Y = {y_1, y_2,. . . , Y_n} the data partition where y_k is the k-th gray level of the image.
- Let c the desired number of classes.
- Let U = (u_{ik}), a matrix having dimension c × n and u_{ik} is the degree of membership of the point y_k in class i.
- Let V = {v_1, v_2,. . . , V_c} the c centroids which are results of the classification algorithm.
- Let m the blur factor, m ∈] 1, + ∞ [.

The FCM algorithm proposed by Bezdek is based on the minimization of the following objective function:

$$J(U,V;Y)=\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^{m}d^{2}(y_k,v_i), \quad (1)$$

Revised Version Manuscript Received on August 05, 2016.

Hayfa Masghouni, National School of Engineers in Carthage, Rades, Tunisia.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

where $d(y_k, v_i)$ is the distance function between the point y_k and the representative v_i of the class i.

The minimization of the objective function is made by an iterative process described by Algorithm 1 presented below:

Algorithm 1: Minimizing the objective function by an iterative process

Data: Y

/ * Data to partition * /

Inputs: m; V_0; c; ε

/ * M: fuzzy factor * /

/ * C: desired number of classes * /

/ * V_0: random initialization centroid * /

/ * ε: stopping criterion * /

Outputs: U; V

/ * U: degree matrix of belonging * /

/ * V: centroid matrix * /

While $| V_{t+1} - V_t | > \varepsilon$ do

$U_t \leftarrow F(V_t)$

/ * F: Update Function degrees of membership * /

$V_{t+1} \leftarrow G(U_t)$

/ * G: Update function centroid * /

end

To minimize the function defined in Equation (1), the functions F and G are determined by using the Lagrange multipliers of the method as shown below.

By considering that y_k vectors (k = 1,..., n) are independent, minimizing J (U, V, Y) can be treated individually for each vector y_k (k = 1, ...,n).

For a fixed k (k = 1,.., n):

$$J_k(U,V;Y) = \sum_{i=1}^{c} u_{ik}^{m} d^2(y_k, v_i). \quad (2)$$

Minimization of J is therefore to minimize J_k

for k = 1,. . . , n : min J_k (3)

with inherent probabilistic constraint in FCM:

$$\sum_{i=1}^{c} u_{ik} = 1,\ \forall k = 1,\dots,n. \quad (4)$$

The optimization problem becomes by using the technique of Lagrange multipliers, an optimization problem without constraint and J_k becomes:

$$J_k(\lambda, u_k) = \sum_{i=1}^{c} u_{ik}^{m} d^2(y_k, v_i) - \lambda\left(\sum_{i=1}^{c} u_{ik} - 1\right), \quad (5)$$

where λ is the Lagrange multiplier.

The pair (λ, u_k) forms a stationary point of the optimized function if and only if:

$$\frac{\partial J_k}{\partial \lambda} = 0 \text{ and } \frac{\partial J_k}{\partial u_k} = 0$$

These two equations lead to the following relationships:

$$\frac{\partial J_k}{\partial \lambda} = -\left(\sum_{i=1}^{c} u_{ik} - 1\right) = 0, \quad (6)$$

$$\frac{\partial J_k}{\partial u_{sk}} = m\, u_{sk}^{m-1} d^2(y_k, v_s) - \lambda = 0, \text{ for } s = 1,2,\dots,c. \quad (7)$$

From the equation (Eq. (6)), the probabilistic constraint is found. The second equality (Eq (7)) can also be written:

$$u_{sk} = \left(\frac{\lambda}{m\, d^2(y_k, v_s)}\right)^{\frac{1}{m-1}} \quad (8)$$

From the probabilistic constraint, we have:

$$\begin{aligned} \sum_{i=1}^{c} u_{ik} &= 1 \\ \sum_{i=1}^{c}\left[\lambda\left(\frac{1}{m\, d^2(y_k, v_i)}\right)^{\frac{1}{m-1}}\right] &= 1 \\ \lambda \sum_{i=1}^{c}\left(\frac{1}{m\, d^2(y_k, v_i)}\right)^{\frac{1}{m-1}} &= 1 \end{aligned}$$

The expression of λ can be deducted:

$$\lambda = \left[\sum_{i=1}^{c}\left(\frac{1}{m\, d^2(y_k, v_i)}\right)^{\frac{1}{m-1}}\right]^{-1} \quad (9)$$

By replacing λ in Eq.(8) with its expression in Eq.(9):

$$\begin{aligned} u_{sk} &= \left[\sum_{i=1}^{c}\left(\frac{1}{m\, d^2(y_k, v_i)}\right)^{\frac{1}{m-1}}\right]^{-1} \times \left(\frac{1}{m\, d^2(y_k, v_s)}\right)^{\frac{1}{m-1}} \\ u_{sk} &= \frac{1}{\sum_{i=1}^{c}\left(\frac{d^2(y_k, v_s)}{d^2(y_k, v_i)}\right)^{\frac{1}{m-1}}} \end{aligned} \quad (10)$$

Equation (10) is the function F of Algorithm 1.

The function G is written as:

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m}\, y_k}{\sum_{k=1}^{n} u_{ik}^{m}} \qquad (11)$$

When m → 1, the membership function is similar to the binary function.

Criticism of FCM:

According to Mr. Roulier Vincent,[7] FCM represents several limitations that consist on the membership function which is not intuitively satisfying where we are away from the centers of classes. It's equal to zero at the center of the other class, then starts to increase. Therefore, it is not decreasing in terms of the distance from the center of the class, and this effect is amplified with increasing number of classes.

According to gentlemen Balafar MA, AR Ramli and S. Mashohor [4] FCM considers only the intensity of the image which causes problems with the presence of noise. Therefore, FCM does not give reliable results for low contrasted and noisy images.

We can conclude that FCM is an algorithm that is not efficient for a high number of classes and for noisy images.

D. The extensions of FCM

DI. .Integration of the possibilistic approach

Possibilistic approach

In 1993, Krishnapuram proposed to add a term in the objective function and abandon the normalization term to solve the problem with the probabilistic approach: [12]

$$J(U,V;Y) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^{m}\, d^2(y_k, v_i) + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{n} (1 - u_{ik})^m \qquad (12)$$

The minimization of the objective function is realized by the iterative process described by the algorithm 1. Due to the independence of the rows and columns of the matrix U, the minimization of the equation (12) is equivalent to minimizing a new equation:

$$J^{ik}(U,V;Y) = u_{ik}^{m} d^2(y_k, v_i) + \eta_i (1 - u_{ik})^m \qquad (13)$$

To find a local minimum, we must derive this equation relative to u_{ik} and solve the equation below:

$$\frac{\partial J^{ik}}{\partial u^{ik}} = m\, u_{ik}^{m-1}\, d^2(y_k, v_i) - m\, \eta_a (1 - u_{ik})^{m-1} = 0 \qquad (14)$$

The author concluded an expression of the function F of updating degree of the membership as follows:

$$u_{ik} = \frac{1}{1 + \left(\frac{d^2(y_k, v_i)}{\eta_i}\right)^{\frac{1}{m-1}}}, \qquad (15)$$

The G function of calculating centroid values remain the same (Eq 11).

The parameter η_i controls the decrease of the membership function in the class c_i by fixing the distance to the center cluster for which the degree of membership is 1/2. The m parameter controls blurring like the FCM method. This time, the membership functions have a more intuitive form because they decrease when the distance to the center of the class increases. This method is named PCM (Possibilistic C-Means). It is proposed by Mr. Krishnapuram in 1993

The expression of η_i conventionally used in the literature (and also proposed by Krishnapuram) is the mean of the fuzzy distance intraclass to centroid v_i is defined by:

$$\eta_i = \frac{\sum_{k=1}^{n} u_{ik}^{m}\, d^2(y_k, v_i)}{\sum_{k=1}^{n} u_{ik}^{m}} \qquad (16)$$

Criticism of the possibilistic approach:

According to [7] and [12], despite the advantage of having a more intuitive membership function, the possiblistic approach has the disadvantage of not approaching to the real values of centroids.

***Evaluation of the two approaches*:**

We can summarize the comparisons made by Mr. Roulier Vincent [7] and those made by Mrs. Hakima Zouaoui and Mr. Moussaoui Abdelouahab [12] by the following table:

	FCM	PCM
more accurate assessment of belonging		X
Stable algorithm (approaches to the real values of centroid)	X	

As shown in the table above, PCM has the advantage of giving a more accurate assessment of belonging and FCM has the advantage of having a stable algorithm by approaching to the real values of centroid.

DII. Generalized approach

In 2003, Menard et al proposed an extension of FCM named FGCM: Generalized Fuzzy C-Means [13].

The principle of this approach consists on adding to the objective function of FCM an additional information called information of Tsalis which consider relatively distant points from the centers of classes which are not taken by classical approaches. This approach is more efficient in case of incertitude.

The objective function proposed by Menard et al in 2003 in the case of the probabilistic approach is written (Eq. (17)).

$$J(U,V;Y) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} d^2(y_k, v_i) + \frac{1}{T.(m-1)} \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} - \frac{1}{T} \sum_{k=1}^{n} \lambda_k \left(\sum_{i=1}^{c} u_{ik} - 1\right) \qquad (17)$$

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The first term is the term of least squares of the objective function of Bezdek (Eq. (1)). The second term depends on the information of Tsalis promoted by T. The last term of the Equation (17) is the probabilistic constraint

The made experiments [7] show that unlike FCM, the FGCM does not annul the measure of belonging to another class at the centroid . This is achieved thanks to Tsalis information that in addition to the consideration of distant points from the centroids, it makes more optimal spacing of the centroids than the fuzzy classification proposed by Bezdek.

The belonging measures in FGCM tend toward those of FCM when T increases. [7]

The objective function in the case of the possibilistic approach (PGCM: Possibilistic Generalized C-Means), is written in the following manner (Eq. (18)):

$$J(U,V;Y) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} d^2(y_k, v_i) + \frac{1}{T.(m-1)}\sum_{i=1}^{c}\sum_{k=1}^{n}[u_{ik}^{m} - u_{ik}] - \frac{1}{T}\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}. \quad (18)$$

The minimization process of the objective functions is an iterative process identical to that presented by Algorithm 1, it determines U and V such that J (U, V, Y) is minimal. The function F used to update the membership degrees (U Matrix) is in the probabilistic case (Eq. (19)):

$$u_{ik} = \frac{1}{Z_m.(1+T.(m-1)d^2(y_k, v_i))^{\frac{1}{m-1}}} \quad (19)$$

where

$$Z_m = \sum_{j=1}^{c}[1 + T.(m-1)d^2(y_k, v_j)]^{-\frac{1}{m-1}}.$$

In both approaches, the update equation of the function G of the centroids is given (Eq. (11)).

DIII. Extensions based on the integration of spatial information

FCM considers only the intensity of the image which causes problems with the presence of noise. To solve this problem, many researchers incorporate spatial information of pixels at FCM because adjacent pixels generally belong to the same class.

The extensions based on the spatial information include:

- FCM_S [3]
- FCM_EN [4]
- FGFCM [3]
- FLICM [5]
- NonlocalFCM [6]
- Ghiduk and Zanaty Algorithm [15]

FCM_S:

Gentlemen Ahmed et al. proposed a modification of FCM by introducing a term which allows the labeling of the pixel (voxel in 3D) from the label of its neighborhood.[3]

The objective function is modified as follows:

$$J_m = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m}\|x_k - v_i\|^2 + \frac{\alpha}{N_R}\sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^{m}\sum_{r\in N_k}\|x_r - v_i\|^2, \quad (20)$$

where x_k is the gray value of the kth pixel, v_i represents the value of the ith cluster, u_{ik} represents the fuzzy membership of the kth pixel of cluster i, N_R is its cardinality, x_r represents the neighbor of x_k and N_k represents all the neighbors around x_k.

The parameter m represents the weight of each fuzzy membership. The α parameter is used for controlling the effect of neighborhoods.

FCM_EN:

Mr. Szilágyi proposed a new extension of FCM named FCM_EN [4]. In this extension, the sum image is obtained from the original image and its average is used at the input for clustering:

$$S_i = \frac{1}{1+\alpha}(X_i + \alpha\overline{X}_i) \quad (21)$$

where α plays the same role as before.

FGFCM:

The steps of this algorithm are the following:

1) Select the number of prototypes clusters c (can be preset or selected according to criteria or knowledge).
2) Initialize the prototypes and randomly choose a very low value of

$$\varepsilon > 0$$

3) Calculate the local similarity measures S_{ij}.
4) Calculate the sum image
5) Update the partition matrix
6) Update the prototypes
7) Repeat steps 5 and 6 until the satisfaction of this criterion:

$$|V_{new} - V_{old}| < \varepsilon,$$

where V is the set of vectors of prototypes clusters.

FLICM:

Gentlemen Krinidis and Chatzis introduced the FLICM (Fuzzy Local Information C-means) and it was proved that it has very good properties. After gentlemen S. Krinidis and M. Krinidis generalize this algorithm in order to be applied for any type of input data. It is effective and robust against noise. Also, it can handle any kind of empirically adjusted parameters.[5]

Nonlocal FCM:

This extension of FCM is based on the introduction of the regulation and the terms of data-driven.

By considering the data driven suitable neighborhood, the use of non-local framework is more effective against noise and intensity inhomogeneities.[6]

Ghiduk and Zanaty algorithm:

Recently in 2016, gentlemen Ghiduk and Zanaty developed a new extension of FCM. The principle of this method is to change the objective function of FCM to allow pixel labeling by the influence of the others pixel and to be robust against noise.[15]

The new objective function is:

$$J_m = \sum_{k=1}^{C}\sum_{i=1}^{N} u_{ki}^m \left\| x_i - c_k \right\|^2 + \gamma \sum_{k=1}^{C}\sum_{i=1}^{N} u_{ki}^m \sum_{x_i \in N_k}^{N_k} \left\| x_i - c_k \right\|^2 \quad (22)$$

where C is the number of intensity regions

N_K is the number of points in a region R_k

x_i is the point belonging to k and c_k is the center of R_k

m is the weight

$$\gamma = \frac{1}{m * N_k}$$

u_{ik} represents the fuzzy membership of *kth* pixel of cluster *i*

E. Comparison between FCM and its various extensions

In the case of a low-contrasted image, the representative peaks of the different tissues can be hardly discerned which causes uncertainty problem in the pixels classification due to the similarity of classes.

Mr. Vincent Roulier [7] made a comparison of robustness of this similarity between FCM + PCM and FGCM + PGCM. For this, he created a dataset with two sets of points C1 and C2.

The results are evaluated by calculating for each approach the percentage of well classified points according to the displacement of C1 to C2.

A low percentage of well ranked points means that the method does not distinguish the two classes. So, the centroids are close and the degrees of belonging are few discriminating.

According to the obtained results, [7] the FGCM + PGCM method is more robust face increasing simlarity of classes. In effect, it optimizes the distance between the centers of classes and does not merge them quickly as in FCM + PCM.

Then, Mr. Vincent Roulier [7] made a comparison between the robustness of methods to noise in images and it proved that FGCM + PGCM is slightly less sensitive to noise than FCM + PCM.

Gentlemen Weiling Cai, Songcan Chen and Zhang Daoqiang [3] compared FGFCM with FCM_S and FCM_EN. The comparisons show that FGFCM is more robust to noise than the other two algorithms.

Also, these researchers compared the running time between FGFCM and FCM and the results show that FGFCM is faster than FCM.

Mr. Balafar [1] compared between FCM-S, FCM-EN, FGFCM, NonLocalFCM and FLICM The results show that the algorithms give close results in noisy images except FLICM which is less robust to noise.

Gentlemen Ghiduk and Zanaty [15] compared their algorithm with FCM_S and FCM. The results showed that their algorithm is more robust and give more accurate results than the other algorithms. Unfortunately, it is expensive that leads to a limitation of its use in 3D.

III. GMM AND ITS EXTENSIONS

A. GMM

The Gaussian Mixture Model GMM considered for each pixel (voxel in 3D) M density mixed components (Gaussian distribution) and M mixing coefficients.

The probability distribution for the jth component is $p_j(x_i \mid \theta_j)$ where x_i is the pixel i and θ_j is the parameter (mean μ_j and covariance matrix Σ_j).

The probability distribution of each pixel (voxel) can be considered as a mixture of the distribution probabilities:

$$p(x_i \mid \theta) = \sum_{j=1}^{M} \alpha_j p_j(x_i \mid \theta_j) \quad (23)$$

$$= \frac{1}{\sqrt{\det(2\pi\Sigma_j)}} e^{-(x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)/2}$$

where α_j is the mixing coefficient with the following constraint:

$$\sum_{j=1}^{M} \alpha_j = 1$$

The component j 's probability distribution is modeled by a Gaussian distribution having as mean μ_j and as covariance matrix Σ_j like

$$p_j(x_i \mid \theta_j) = p_j(x_i \mid \mu_j, \Sigma_j) \quad (24)$$

Generally the estimation ML (maximum likelihood) [1] is used to find the parameters μ_j and Σ_j.The log-likelihood expression of the parameter θ and the image X is:

$$\log(L(\theta \mid X)) = \log \prod_{i=1}^{N} p(x_i \mid \theta) \quad (25)$$

$$= \sum_{i=1}^{N} \log\left(\sum_{j=1}^{M} \alpha_j^t p_j(x_i \mid \theta_j^t)\right)$$

- EM:

It is very hard to find the solution of this ML equation. So,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

the researchers used the EM algorithm (Exceptation Maximization) to obtain theses parameters. The steps are as follows:[1]

Step E: In this step we try to find the probability of x_i belonging to the class θj.

$$p(j \mid x_i, \theta^t) = \frac{\alpha_j^t p_j(x_i \mid \theta_j^t)}{\sum_{j=1}^{M} \alpha_j^t p_j(x_i \mid \theta_j^t)} \qquad (26)$$

Step M: From the probability obtained in step E, the mixing coefficient α_j, the mean μ_j and the covariance matrix Σ_j are calculated.

$$\alpha_j^{t+1} = \frac{1}{N}\sum_{i=1}^{N} p(j \mid x_i, \theta^t) \qquad (27)$$

$$\mu_j^{t+1} = \frac{\sum_{i=1}^{N} x_i p(j \mid x_i, \theta^t)}{\sum_{i=1}^{N} p(j \mid x_i, \theta^t)} \qquad (28)$$

$$\Sigma_j^{t+1} = \frac{\sum_{i=1}^{N} p(j \mid x_i, \theta^t).(x_i - \mu_j^{t+1})(x_i - \mu_j^{t+1})^T}{\sum_{i=1}^{N} p(j \mid x_i, \theta^t)} \qquad (29)$$

These steps are repeated until convergence.

B. GMM with spatial data

The spatial information of the pixels is included, so that the GMM method is more robust to noise.

The average of neighboring pixels around the pixel xi denoted $\bar{x}_i$ is calculated before the GMM clustering to make it faster. Its distribution value is added to the distribution value of x_i in the probability function (equation 25)

$$\log(L(\theta \mid X)) = \log \prod_{i=1}^{N} p(x_i \mid \theta) = \sum_{i=1}^{N} \log\Big(\sum_{j=1}^{M} \alpha_j^t [(1-\beta) * p_j(x_i \mid \theta_j^t) + \beta * p_j(\bar{x}_i \mid \theta_j^t)]\Big) \qquad (30)$$

The parameter β refers to the weight information about the neighborhood.

The integration of neighborhood information has the effect of improving the segmentation results with a high noise level, but unfortunately this is not the case for low noise levels where performance is degraded because of the fuzzy effect .To solve this problem, the noise variance is used to specify the parameter β.

-NWEM:

An extension of EM using neighborhood information called NWEM is described as follows:[2]

Step E:

Equation 26 is modified for a component k as follows:

$$p(k|x_i, \Theta^t) = \frac{\alpha_k^t W_{ik}^t p_k(x_i|\Theta_k^t)}{\sum_{j=1}^{K} \alpha_j^t W_{ij}^t p_j(x_i|\Theta_j^t)} \qquad (31)$$

where

$$W_{ik} = \frac{\sum_{n=1}^{|N_i|} p(k|x_{ni}, \Theta^t)}{|N_i|}$$

is the weight of the neighborhood information, N_i is the number of neighboring voxels of x_i and x_{ni} is the intensity of the nth neighbor of the ith voxel

Step M:

At this stage the equations 27, 28 and 29 are modified as follows:

$$\alpha_k^{t+1} = \frac{1}{N}\sum_{i=1}^{N} p(k \mid x_i, \Theta^t) \qquad (32)$$

$$\mu_k^{t+1} = \frac{\sum_{i=1}^{N} x_i \cdot p(k \mid x_i, \Theta^t)}{\sum_{i=1}^{N} p(k \mid x_i, \Theta^t)} \qquad (33)$$

$$\Sigma_k^{t+1} = \frac{\sum_{i=1}^{N} p(k \mid x_i, \Theta^t) \cdot (x_i - \mu_k^{t+1})(x_i - \mu_k^{t+1})^T}{\sum_{i=1}^{N} p(k \mid x_i, \Theta^t)} \qquad (34)$$

- EM1

Another extension of EM noted EM1 [1] was proposed by Mr. Med Ali Balafar which is used to resolve the probability function, its principle is:

1) In the equation 26 in step E, the average value of the distribution of the neighboring pixels around the pixel x_i is added to the distribution value of x_i as a neighborhood information:

$$A = [(1-\beta) * p_l(x_t \mid \theta_j^t) + \beta * p_l(\bar{x}_t \mid \theta_j^t)]$$

$$p(j \mid x_i, \theta^t) = \frac{\alpha_j^t . A}{\sum_{j=1} \alpha_j^t . A} \quad (35)$$

2) In the equation 28 in step M, the value of the average of neighboring pixels of x_i is added to the x_i value

$$A = [(1-\beta) * p_l(x_t \mid \theta_j^t) + \beta * p_l(\bar{x}_t \mid \theta_j^t)]$$

$$p(j \mid x_i, \theta^t) = \frac{\alpha_j^t . A}{\sum_{j=1} \alpha_j^t . A} \quad (36)$$

3) In the equation 29 in Step M, the distance between the average of the neighboring pixels of x_i and the component 's center is added to the distance between x_i and the component 's center.

$$d(x) = (x - \mu_j^{t+1})(x - \mu_j^{t+1})^T$$

$$\Sigma_j^{t+1} = \frac{\sum_{i=1}^{N} p(j \mid x_i, \theta^t).(d(x_i) + \beta . d(\bar{x}_i))}{\sum_{i=1}^{N} p(j \mid x_i, \theta^t)} \quad (37)$$

In MRI, the noise behaves like Rician distributed noise.The Rician noise approaches the Gaussian distribution with high SNR (Signal to Noise Ratio) to the Rayleigh distribution with low SNR.

The Rician distribution and the Rayleigh distribution:

The Rayleigh distribution is a density probability distribution . It appears as the norm of a two-dimensional Gaussian vector whose coordinates are independent, focused and have the same variance. This probability distribution is named by Lord Rayleigh.

Rician's law, named by Stephen O. Rice (1907-1986),[8] is a density probability distribution. It is a generalization of Rayleigh law which is used to describe the behavior of a radio signal that travels along several paths (multipath) before being received by an antenna,

In the background, the Rician distribution is the same as the Rayleigh distribution because there is no signal in it,

The Rayleigh PDF of the independent statistical observations is defined by:

$$p(\{O_i\}) = \prod_{i=1}^{n} \frac{O_i}{\sigma^2} e^{-(O_i^2)/(2\sigma^2)} \quad (38)$$

where O designates the observations and σ^2 is the noise variance

The noise variance is obtained from the PDF's log-likelihood maximization

$$\sigma_{Noise}^2 = \frac{1}{2n} \sum_{i=1}^{n} O_i^2 \quad (39)$$

The background pixels are considered as observations O and the noise 's variance is obtained by applying Eq 35 on the pixels in the background. Half the average of the background pixels's powers is considered as noise variance.[1]

EM1 with the user's interaction :-

The steps are as follows:[1]

1) The volume is divided into n clusters where n is the number of target classes (tissues) .

2) <u>Before segmentation:</u> If clusters contain more than one target class, the user selects the clusters to partition and the process is repeated until the user's satisfaction.

3) <u>After segmentation:</u> If multiple clusters correspond to a target class, the user selects the clusters for each target class.

Experiments have shown that EM1 with the user's interaction offer better segmentation results than EM1.[1]

- Other EM extensions:

A. R. F. d. Silva [9] proposed two Bayesian algorithms (DPM, rJMCMC) using the Markov chain models to find the normal mixing models with an unknown number of components. Both algorithms are used in the segmentation of MRI images .

There are also two other Bayesian algorithms based on the segmentation of brain MRI images (KVL [10] and MPMMAP [11]).

Recently in 2015, a generative probabilistic model for segmentation of brain lesions in multi-dimensional images was invented. It uses the Gaussian mixture model for modeling brain images and the probabilistic tissue atlas that employs expectation-maximization (EM) to estimate the label map for a new image. It shares information about the spatial location of the lesion among channels.[17]

IV. COMPARISON BETWEEN THE DIFFERENT EXTENSIONS OF EM

The inventors of the generative probabilistic model [17] showed that their algorithm is more robust and accurate for delineating lesion structures than classic EM.

Gentlemen R. Adelino and Ferreira da Silva [16] made a comparison between DPM and rjMCMC. The results showed that DPM gives better segmentation results than rJMCMC with a less expensive cost of implementation .

Mr. Mohammad Ali Balafar [1] compared his method to other extensions of EM and FCM:

The first experiments were realised on brain MRI images with a Rician noise percentages equal to 9% and 7% by applying algorithms EM1 and NWEM.The results showed that EM1 is more robust to noise than NWEM.

The following experiments compare the similarity index, the rfp (Ratio of False Positive) and the rfn (Ratio of False Negative) between EM1 and NWEM. The results showed that EM1 has higher values of similarity indexes and lower values of rfp and rfn, which means that this algorithm produces more accurate segmentation results.

EM1 and NWEM are identical for a percentage of noise less than 5% . For a percentage higher than 5%, EM1 offer better segmentation results.

Mr. Mohammad Ali Balafar [1] also showed that for a noise percentage equal to 9%, when the neighborhood size increases, the similarity index decreases which means that the fuzzy effect depends on the neighborhood size. He also showed that EM1 is faster than NWEM.

Then he compared EM1 to other extensions of EM (DPM RJMCMC, KVL and MPM-MAP).The results showed that EM1 has the highest values of the average similarity indices, which induces a better segmentation for EM1.

After he compared EM1 to other extensions of FCM based on neighborhood information((FCM_S [3], FCM_EN [4], FGFCM [3], FLICM [5] and NonlocalFCM [6])). The experiments have shown that EM1 has the highest similarity index 's mean value so it is the best in MRI segmentation.

V.CONCLUSION

FCM is an algorithm used in MRI image segmentation based on the fuzzy clustering. Many extensions were developed to boost the performance of this algorithm: It was integrated with PCM to obtain more accurate segmentation results .An extension of FCM called FGCM was invented to be more robust against the similarity of classes.To be more robust against noise and provide better segmentation results , the spatial information has been integrated. Hence the appearance of several extensions of FCM such as FCM_S, FCM_EN, FGFCM, NonLocalFCM, FLICM and Ghiduk and Zanaty algorithm.

GMM is an another algorithm used in MRI image segmentation which used EM to estimate the model's parameters. To be more robust against noise, the spatial information was integrated. Hence the appearance of several extensions of EM such as EM1, NWEM, DPM RJMCMC, KVL, MPM-MAP and the generative probabilistic model.

REFERENCES

1. Balafar, "Spatial based Expectation Maximizing (EM)," Diagnostic Pathology 2011 6:103.
2. Tanga H, Dillensegerb J, Baoa XD, Luoa LM, "A Vectorial Image Soft Segmentation Method Based on Neighborhood Weighted Gaussian Mixture Model," Computerized Medical Imaging Graphics 2009, 33:644-650
3. Weiling Cai, Songcan Chen, Daoqiang Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation,"Nanjing 210016, PR China27, July 2006
4. M.A. Balafar, A. R. Ramli and S. Mashohor,"Edge-preserving Clustering Algorithms and Their Application for MRI Image Segmentation," Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 Vol I, IMECS 2010,March 17-19.2010.Hong Kong
5. Krinidis S, Chatzis V, "A Robust Fuzzy Local Information C-Means Clustering Algorithm,"IEEE Transactions on Image Processing 2010,19:1328-1337.
6. Wang J, Kong J, Lub Y, Qi M, Zhang B,"A modified FCM algorithm for MRI brain image segmentation using both local and non-local spatial constraints,"Computerized Medical Imaging and Graphics 2008, 32:685-98.
7. Vincent Roulier,"Fuzzy classification and MRI modeling: Application to the quantification of fat for optimal evaluation of health hazards associated with obesity," Ph.D. Thesis, Doctoral school of ANGERS, 2008
8. Stephen O. Rice," Mathematical Analysis of Random Noise,"Bell System Technical Journal, vol. 24, 1945, p. 46–156
9. Silva ARFD, "Bayesian mixture models of variable dimension for image segmentation,"Computer methods and programs in biomedicine 2009,94:1-14.
10. Leemput FMKV, Vandermeulen D, Suetens P,"Automated model-based tissue classification of MR images of the brain," IEEE Transactions on Medical Imaging 1999, 18:897-908.
11. Marroquin BCVJL, Botello S, Calderon F, Fernandez-Bouzas A,"An accurate and efficient Bayesian method for automatic segmentation of brain MRI," IEEE Transactions on Medical Imaging 2002, 21:934-945.
12. Zouaoui Hakima, Moussaoui Abdelouahab, "Clustering fuzzy data fusion applied to the segmentation of brain MRI images," CIIA, 2009
13. Weiling Cai, Songcan Chen, Daoqiang Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," Pattern Recognition 40(2007) 825 – 838
14. Michel Menard, Vincent Courboulay, Pierre-Andrée Dardignac, "Possibilistic and probabilistic fuzzy clustering unification within the framework of the non-extensive thermostatistics,"Pattern Recognition 36 (2003) 1325 – 1342
15. Ahmed S. Ghiduk,E.A.Zanaty, "Modified Fuzzy C-Means for Segmenting Magnetic Resonance Images (MRIs)," International Journal of informatics and medical data processing (JIMDP) vol.1, no.2, pp. 48-58, 2016.
16. Adelino R. Ferreira da Silva, "Bayesian mixture models of variable dimension for image
17. segmentation," Rua Dr. Bastos Goncalves, n. 5, 10A, 1600-898 Lisboa, Portugal,2008
18. Bjoern H Menze, Koen Van Leemput, Danial Lashkari, Tammy Riklin-Raviv, Ezequiel Geremia, al, "A generative probabilistic model and discriminative extensions for brain lesion segmentation - with application to tumor and stroke,"IEEE Transactions on Medical Imaging, 2015. <hal-01230846>