

An Image Segmentation Algorithm Based on Fuzzy Clustering and Genetic Algorithms with a New Distance

F.HACHOUF

Department of electronics, Automatic and Robotic laboratory, Mentouri Constantine University, Road of Ain El Bey, 25000 Constantine, Algeria, Tel : 031.61.42.06 poste 31, Fax : 031.61.42.07

A.ZEGGARI

Department of Exact Science and Technology, Academic Center of Tebessa, Algeria

Z. AHMED SEGHIR

Department of Exact Science and Technology, Academic Center of LARBI BEN M'HIDI-Oum el bouaghi, Algeria

Abstract

This paper describes a new GA-clustering algorithm for image segmentation. We combine the classical fuzzy c-means algorithm (FCM) with a genetic algorithm, and we modify the distance function in FCM for taking into account the spatial information and the color of a pixel. Image segmentation is treated as an unsupervised classification which is optimised by a genetic algorithm. The idea is to choose several configurations of initial centres and to code chromosomes by the membership degrees of pixels to the clusters. The new proposed distance yield uniform regions while respecting the quality of segmentation.

Key words: Genetic Algorithms, Classification, Fuzzy C means, Segmentation.

1. Introduction

Image segmentation is an important step in image analysis and pattern recognition. It is the first essential step in low level vision. Segmentation is a process of partitioning an image into some non-intersecting regions such that each region is homogeneous, but the union of any two adjacent regions is not [1] [2] [3] [4]. It is applied in a variety of domains. The choice of segmentation technique depends widely on images and domain application

Image segmentation can be seen as an unsupervised classification problem. It supposes that it is possible to group in one cluster pixels belonging to one same region. We suppose therefore that:

- All pixels belonging to a region are affected to only one cluster.
- Pixels belonging to a cluster can form several and no adjacent regions in the image.

Several algorithms have been proposed based on clustering methods such as hard c-means (HCM) and fuzzy c-means (FCM) [5]. In this paper, we propose a genetic algorithm whose initial population is the different choices of initial centres, which converges toward a better solution in spite of the initial choice of centres.

The application of GAs (genetic algorithms) to the problem resolution requires coding the potential solution under bit strings to form chromosomes, optimising a fitness function and finally defining genetic operators.

In this work, the problem to solve is to find the best classification of pixels. Since there is no direct path toward the solution, we use the GAs to avoid the locally optimal solution in favor to the global one.

Another problem comes from the fact that, in its standard use, the resulting regions are not spatially continuous, due to the fact that only the color or gray level uniformity is checked [6]. Therefore we propose a modified distance to ensure spatially continuous regions after segmentation.

This paper is organized as follows. In sections 2, we introduce the main ideas of fuzzy C-means algorithm. In section 3, the proposed method is described. Section 4 will be dedicated to the application of the proposed method to monochromatic and color images. Experimental results are discussed. Finally, Conclusion and perspectives of our work are suggested.

2. Classifications with FCM

Fuzzy C-Means algorithm [5] can be formulated as follows:

$$\text{Minimize } J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2. \quad (1)$$

d_{ik} is the Euclidian distance between pixel k and the cluster center i

Under the constraints:

$$\sum_{k=1}^c u_{ik} = 1, \quad 1 \leq i \leq n. \quad (2)$$

$$u_{ik} \geq 0, \quad 1 \leq i \leq n, \quad 1 \leq j \leq c. \quad (3)$$

n is the total number of feature vectors (pixels) and c is the number of cluster centers. $m > 1$ is used to adjust the weighting effect of the membership value; large m will increase the fuzziness of the function causing feature vectors with low membership to contribute less. U is the fuzzy partition matrix, which contains the membership of each feature vector to each cluster. The cluster center matrix is V , such that $V = \{v_1, v_2, v_3, \dots, v_c\}$, where each v_i is a cluster center of s features, $v_i \in \mathcal{R}^s$. X is the matrix of feature vectors, where $X = \{x_1, x_2, x_3, \dots, x_n\}$, $x_i \in \mathcal{R}^s$.

The cluster center v_i can be obtained from the following:

$$v_i = \sum_{k=1}^n (u_{ik})^m x_k / \sum_{k=1}^n (u_{ik})^m \quad \forall i. \quad (4)$$

The fuzzy membership matrix U can be obtained from the following:

$$u_{ik} = 1 / \sum_{j=1}^c (d_{ik} / d_{jk})^{2/(m-1)} \quad (5)$$

for $d_{ik} = \|x_k - v_i\| > 0, \quad \forall i \text{ et } k.$

if $d_{ik} = 0$ then $u_{ik} = 1$ and $u_{jk} = 0$

for $j \neq i.$
(6)

The cluster center matrix, V is randomly initialised and the fuzzy partition matrix, U is created by using (5) for all feature vectors. The stopping condition is set to be ϵ .

Fuzzy c-Means Algorithm

Step-1 Initialise the cluster center matrix, V by using a random number generator, and record the cluster centers, set $l=0$.

Step-2 Initialise the fuzzy partition matrix, $U^{(l=0)}$ with (5) and (6).

Step-3 Increment $l=l+1$. Compute V by using (4).

Step-4 Compute $U^{(l)}$ by using function (5) and (6).

Step-5 If $\|U^{(l)} - U^{(l-1)}\| < \epsilon$ then stop, else repeat

Step-3 to Step-5.

3. Fuzzy-Genetic Classification Algorithm

In order to avoid the local minimum, and then to lead the image toward a natural segmentation, we

propose a fuzzy-genetic classification algorithm. Before presenting our algorithm, we are going to define the initial population, the fitness function and the genetic operators used here.

Population

At the difference with [7] [8] [9] that proposed to code every individual by c cluster centers, we propose to code chromosomes by the membership degrees of pixels to clusters. That is, a chromosome is divided into n cases (n is the number of pixels), and every case will be divided into c cases (c is the number of clusters).

Because we work with membership degrees, a chromosome is coded with values $\in [0,1]$.

We choose to work with the membership degrees instead the cluster centres to avoid the exchange of all information at time. Using membership degrees the parents exchange between them a portion but not the totality of all centres. Adjusting each cluster centre every generation minimizes the number of generations needed.

Fitness function

The considered fitness function is:

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2.$$

Selection

The selection consists in choosing the best chromosomes that maximize the selective function. In our case we select the better between 2 randomly chosen individuals to construct the first parent; this process will be repeated with the 2 news chosen competitors among the remainder of the population to construct the other parent. This two selected individuals are used during the operation of crossover.

Crossover

- Let R_c be the crossover rate,
- Generate a random number N_c ,
- If $(R_c > N_c)$ then
 - choose randomly two crossover points $z_1 z_2$ with $(z_i \text{ modulo } c) = 0$. (c is the cluster number) to ensure the exchange of all membership degrees of a pixel and therefore

to respect the constraint $\sum_{j=1}^c u_{ik} = 1$

- apply crossover.

Mutation

- Let R_m be the mutation rate,
- Generate a random number N_m ,
- If $(R_m > N_m)$ then
 - choose randomly a mutation point z with $(z \text{ modulo } c) = 0$. (c is the clusters number),

- choose randomly $a \in [0-val(z) \dots 1-val(z)]$,
 $val(z)$ is the membership degree indexed by z ,
- put $val(z) = val(z) + a$,
- In order to respect the constraints (2) and (3), we propose to modify the remaining membership degrees as follows:

$$\begin{aligned} val(z+1) &= val(z+1) - a / (c-1), \\ val(z+2) &= val(z+2) - a / (c-1), \\ &\vdots \\ val(z+c-1) &= val(z+c-1) - a / (c-1). \end{aligned}$$

The modified distance

The modified distance used is given by:

$$d((x_k, y_k), (v_{xi}, v_{yi})) = \min \left\{ D(x_k, v_{xi}), |y_k - v_{yi}| \right\}$$

So that:

x_k, y_k are respectively the position and intensity of the pixel k .

v_{xi} is the coordinate of the cluster center i calculated as:

$$\begin{aligned} X_{vi} &= \sum_{k=1}^n (u_{ik})^m X_k / \sum_{k=1}^n (u_{ik})^m \\ Y_{vi} &= \sum_{k=1}^n (u_{ik})^m Y_k / \sum_{k=1}^n (u_{ik})^m \quad \forall i. \end{aligned}$$

v_{yi} is the intensity of the cluster center i calculated as (4).

D is the Euclidean distance between the pixel and the cluster center positions.

The proposed algorithm

- E0:** choose several initial center configurations;
- E1:** initialize the population of P size using (5) and (6) (Fuzzification);
- E2:** for $i = 0, i < G$ (G the number of generations)
- calculate the fitness function (J_2) for every individual,
 - select the ($P/2$) pair individuals,
 - apply crossover and mutation operators on this population.
 - create a new generation of P size composed of $P/2$ children lately engendered, as well as the best $P/2$ individuals of the previous generation.
- E3:** choose the best chromosome that maximize the fitness function of all generations,

- calculate clusters centers using (4),
- affect every pixel to the nearest center (defuzzification).

4. Experimental results

The proposed method has been applied on grey level and color images of Fig.1, Fig.2 and Fig.3. Results are obtained with a crossover rate equal to 0.7, mutation rate is 0.05 and the degree of fuzziness m set to 2.

The MRI image of Fig. 1 has been segmented with a number of population $P=10$ and a total generations $G=40$. the number of clusters C is set to 3 which correspond to segmenting the image into three normal tissues. The color image of Fig.2 has been segmented with a number of clusters $C=4$ and a total of generations $G=30$. Because the image of Fig.3 is complex, including shadows and highlights, we use a number of clusters $c=8$ and a total generations $G=1000$.

We can notice a few change of values of all centres and not only one center because chromosomes interchange membership degrees instead of the values of the centres. Thus, the best chromosome of the previous generation and the one of the current generation differs by their membership degrees, what gives theoretically different centres according to (2) and (4).

The major drawback of the GAs and the classification methods is the execution time. In our case, the time is proportional to the size of population, the size of the image, the size of the feature vectors and the number of clusters. As it is mentioned in [7], this kind of algorithms consume more time to compute the fitness function.

5. Conclusion

Clustering techniques are widely used for segmentation. However, these techniques suffer from some limitations. In fact, there are a number of problems with these approaches such as cluster centre initialisation, existence of local minima, and the selection of several parameters. To overcome these shortcomings, We have presented a new image segmentation algorithm with classification pixels according to their intensity and position. A new distance is used in fuzzy C-means algorithm. To avoid the local minima, we start our segmentation process with several center configurations. These configurations represent the population that must make evolve a GAs. This last, if it is well coded, turns up near the path of the optimal solution instead to follow initially a path of the locally optimal solution. We have proposed to code chromosomes with membership degrees instead of clusters centers, so operators of crossover and mutation affect slightly the current solution toward an optimal solution.

The proposed method confirm that genetic guided clustering algorithms provide reliable image segmentation results. However, further studies are necessary for a compromise between computational time and best possible segmentation results. Future work include to study the cluster relations (clouds of points) for each segmentation problem, more relevant segmentation measures and image noise.

6. References

- [1] J. P. Cocquerez, S. Philipp, *Analyse d'images : filtrage et segmentation*, Paris, Masson, 1995.
 - [2] S.K. Pal et al. , A Review on image segmentation 29 (1993) 1277-1294.
 - [3] R. M. Haralick, L. G. Shapiro, “ *Image Segmentation techniques*”, *Computer Vision Graphics Image Processing*, (1985), 29, pp. 100-132.
 - [4] P. K. Sahoo et al., “ A survey of thresholding techniques ”, *Computer Vision Graphics Image Processing*, (1988), 41, pp.233-260.
 - [5] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective* J. C. Bezdek, *Pattern Recognition with*
 - [6] M. Gordan, C. Kotropoulos, A. Georgakis, I. Pitas. “A new fuzzy c-means based segmentation strategy. Applications to lip region identification”, A&QT-R 2002 (THETA 13) IEEE-TTTC International.
 - [7] L.O. Hall, B. Ozyurt, J.C. Bezdek, *the case for Genetic algorithms in Fuzzy Clustering* , *IEEE transactions on evolutionary computation*, vol. 3, no. 2, July 1999, pp.1-27.
 - [8] U. Maulik, S. Bandypadhyay, “Genetic algorithm based clustering technique”, *Pattern Recognition*, 33, 2000, pp. 1455-1465.
 - [9] L. Ballerini et al., “Image segmentation by a Genetic Fuzzy c-Means Algorithm using color and spatial information”, *Lecture Notes in Computer Science* , in *Proc. 6th European Workshop on Evolutionary Computation in Image Analysis and Signal Processing* , Coimbra, Portugal, 5-7 April, 2004.
- Fuzzy Objective function Algorithm*, plenum press, New York, 1981.

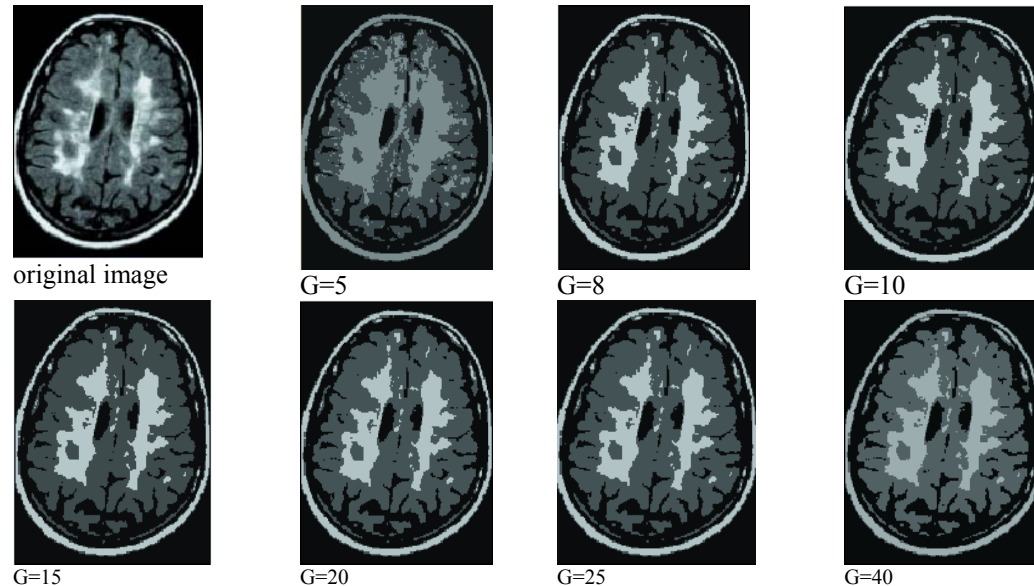


Fig 1. Results of MRI image C = 3

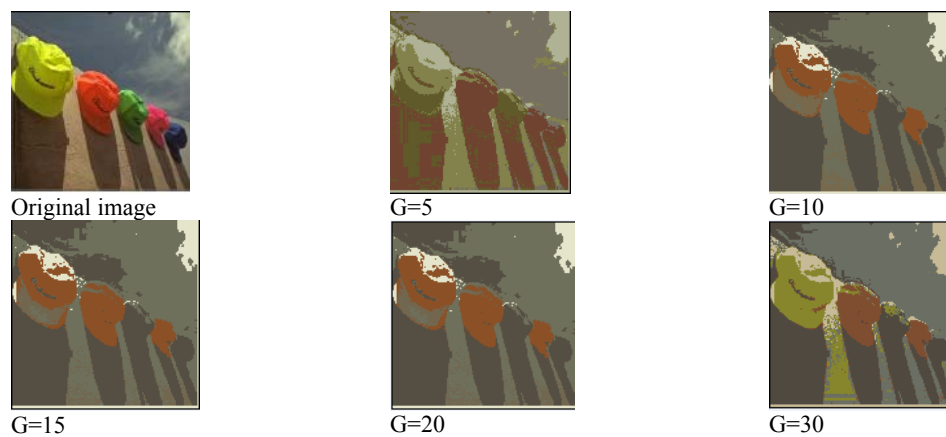


Fig 2. Results with C=4

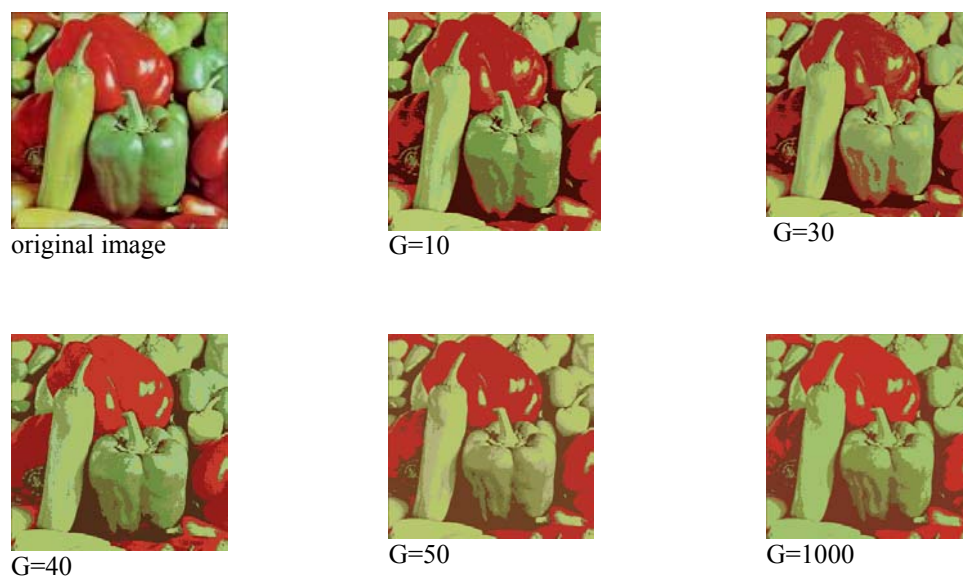


Fig 3. Results with C=8