A Weighted K-means Algorithm applied to Brain Tissue Classification

Guillermo N. Abras and Virginia L. Ballarin

*Signal Processing Laboratory, School of Engineering,
Universidad Nacional de Mar del Plata, vballari@fi.mdp.edu.ar*

ABSTRACT

Tissue classification in Magnetic Resonance (MR) brain images is an important issue in the analysis of several brain dementias. This paper presents a modification of the classical K-means algorithm taking into account the number of times specific features appear in an image, employing, for that purpose, a weighted mean to calculate the centroid of every cluster.

Pattern Recognition techniques allow grouping pixels based on features similarity. In this paper, multispectral gray-level intensity MR brain images are used. T_1, T_2 and PD-weighted images provide different and complementary information about the tissues. Segmentation is performed in order to classify each pixel of the resulting image according to four possible classes: cerebro-spinal fluid (CSF), white matter (WM), gray matter (GM) and background. T_1, T_2 and PD-weighted images are used as patterns.

The proposed algorithm weighs the number of pixels corresponding to each set of gray levels in the feature vector. As a consequence, an automatic segmentation of the brain tissue is obtained. The algorithm provides faster results if compared with the traditional K-means, thereby retrieving complementary information from the images.

Keywords: Pattern-Recognition, Classification, Images, Brain, Tissue.

1. INTRODUCTION

Visual brain atrophy evaluation is very limited. The Multicentre Consortium to Establish a Registry for Alzheimer's Disease (CERAD) determined that the visual evaluation of the generalized brain atrophy is insufficient and that image interpretation is greatly subjective. The CERAD stated that more sensible techniques were needed when measuring global atrophy. [1]. Therefore, efforts should aim at achieving brain tissue quantification in a more accurate, precise and, especially, non-subjective way. Magnetic resonance images show loss of brain tissue, mainly of gray matter, and consequently an increase in cerebrospinal fluid [2] [3].

Conventional MR images provide morphological information. T_1, T_2 and PD-weighted images supply information about the physical properties of water in tissues. T_1, spin-lattice relaxation time, is inversely proportional to the number of molecular motions in tissues at resonance frequency. T_2, spin-spin relaxation time, is inversely proportional to the number of motions in tissues at frequencies below and equal to resonance frequency. And PD provides proton density images. Understanding the physics of these interactions is not necessary; the key lies in the fact that the three imaging modalities: T_1, T_2 and PD reveal independent information about several biological components (fat, water, blood and bone to name a few).

Segmentation using multispectral MR images has yielded satisfactory results. Although the distinctive features of the different tissues are vast, multispectral analysis is still notoriously successful. In 1990, Raya, S. segmented MR images using proton density and T_2-weighted parameters [4]. In 1991, Chen et al. used T_1 and T_2-weighted images together in order to interpret and improve segmented MR images [5].

The use of pattern recognition methods to segment MR image data sets has been widely described in literature [6] [7] [8] [9]. Nearest neighbor, maximum likelihood and Parzen window classifiers [10] are among the supervised classification algorithms found in literature. In spite of this, references to non-supervised techniques [11] such as K-means algorithm [12], minimum distance, maximum and hierarchical clustering [13] can also be found.

Perhaps the most widely used non-supervised technique is the K-means algorithm. Undoubtedly, each approach has its own pros and cons. Many attempts have been made to improve the performance of the basic K-means algorithm. The C-means, for instance, incorporates a fuzzy criterion function [14]; and Krishna suggested using genetic algorithms [15] [16]. Besides, many authors have improved the basic K-means mapping through a neural network [17]. However, most of these improvements on the K-means algorithm are computationally demanding.

This paper proposes a new variation of the K-means algorithm using complementary information available in T_1, T_2 and PD MR images. Moreover, the suggested algorithm takes into consideration the

times the set of gray levels combined in T_1, T_2 and PD appear in the images. The main goal of this paper is to reduce the calculations carried out during the clustering process, thereby providing a high-speed good-quality segmentation of MR brain images.

2. PATTERN RECOGNITION

Pattern recognition techniques have been widely applied to the identification of regions with different textures in images employing different features. Once a group of features has been obtained for each region, an indicator of the existing texture is assigned to each of them. This process can be carried out with a supervised algorithm. In such process, the patterns associated with each texture are already known. On the other hand, in non-supervised algorithms, regions are grouped according to some criteria of pattern similitude. The algorithm herein proposed is non-supervised.

First of all, a set of n features is measured in the set of the different tissues to be classified. These measurements make up a feature vector in the n-dimensional space. The total number of tissues to be classified constitutes a set Ω. Let $W = \{ W_1 , W_2 , ... , W_K \}$ be a partition of Ω. Each subset $W_1, W_2, ... , W_K$ of Ω is known as a class. In order to assign each tissue to a class, the partition $H = \{ H_1 , H_2 , ... , H_K \}$ is defined in R^n. The aim is to find the partition H that reflects the best possible partition W. That is to say, if the component of the feature vector is part of a different tissue region H_i of R^n, it could be practically ascertained that this pixel belongs to the class W_i [8] [18].

In this work, the objects to be classified are the pixels of brain multispectral MR images. The three images used are the calculated T_1, T_2 and PD images, which share the same size. Each pixel has to be assigned to four possible classes, CSF, gray matter, white matter and background [19] [20].

The feature vector X = (x_1, x_2, x_3) , where x_1, x_2 and x_3 are the gray levels of the images T_1, T_2 and PD respectively, represents a pattern in the R^3 space. The objective is to obtain four clusters or groups of patterns: H_1, H_2, H_3 and H_4 by using feature vectors X. Thus, all pixels whose vectors belong to H_i partition are assigned to a W_i class. The result of the performed algorithm is a segmented brain image.

2.1. Minimum Distance Classifier

This algorithm assigns a pattern X prototype or a representative of the same, P_i, to each K class, where i = 1,2,...K. Then, each pattern is assigned to the class whose prototype is nearest (minimum Euclidean distance). Then, if

$$D_i^2 = \|x - p_i\|^2 \qquad (1)$$

is the squared distance of the pattern to prototype P_i, the minimum is chosen. In this case, a lineal decision function is obtained, thus regions H_i in the plane are separated by straight lines. The selection of K pixels patterns is carried out manually from the image itself. The proper selection of prototypes is essential if a good performance of the algorithm is sought.

2.2. Maximin Algorithm

For a maximin algorithm, a randomized cluster center is initially chosen, in this case, a pixel corresponding to the image background. Afterwards, the pixel whose feature vector has the greatest Euclidean distance in R^3 space is selected, electing it as the second center. The distance to both centers of all remaining samples are calculated, choosing the minimum of the two (the sample is assigned to the nearest cluster). Then, the maximum of these minimum distances is selected and the corresponding pattern is chosen as the third cluster center. The procedure is repeated until the four cluster centers are obtained, recalculating the distances of the remaining samples to the same, assigning them to the class whose center is nearest. By applying this method, the centers are not necessarily the most representative of the class.

2.3. K-Means Algorithm

The algorithms introduced in sections 2.1 and 2.2 are basically intuitive procedures. This section deals with an algorithm based on the minimization of a criterion function (or performance index) consisting, in this case, in the sum of the squares of the distances of all cluster points to the center of the same. Three initial centers are randomly selected for the three clusters; and each sample is assigned to the cluster whose center is nearest (minimum Euclidean distance). The center of each cluster is recalculated minimizing the criterion function, which turns out to be the mean X of the samples corresponding to the cluster. The algorithm ends when, in the next iteration, the cluster centers do not get modified. K-means behavior is influenced by the selection of the initial centers. Prototypes are not required, as they result from the data itself.

3. MATERIALS AND METHODS

3.2. Weighted K-Means Algorithm

The weighted K-means algorithm is a variation of the classic K-means algorithm.

As mentioned above, the patterns are the components of the X feature vector whose elements (x_1, x_2, x_3) are the pixels gray levels with equal spatial location in T_1, T_2 and PD images. Although there are many pixels with the same set of values (x_1, x_2, x_3), such as background pixels; there are some patterns appearing only in one pixel location. To calculate the centroids, the weighted K-means algorithm takes into account the times a specific pattern X appears in the images, using a weighted mean to compute the centroids.

The algorithm steps are as follows:

1-Centroids of K classes are chosen: Z_1 (1), Z_2 (1) ... Z_K (1)

2- At the pth iterative step, X samples are distributed into K clusters, assigning each sample to the cluster whose centroid is nearest, using the following relation:

$$X \in S_j(p) \quad if \quad \left\| X - Z_j(p) \right\| < \left\| X - Z_i(p) \right\| \quad (2)$$

$$\forall \quad i = 1, 2, ..., K \quad ; i \neq j$$

where $S_j(p)$ denotes the set of patterns whose cluster centroid is $Z_j(p)$

3-From the results obtained in the second step, the new cluster centroids are recalculated Z_1 $(p+1)$, Z_2 $(p+1)$ ... Z_K $(p+1)$ so

$$J_j = \sum_{i,j} \left\| n_{i,j} X_{i,j} - Z_j(p+1) \right\|^2 \quad (3)$$

being $i=1,2,...,Q$, where Q_j is the number of patterns of the jth class; $j = 1, 2, ..K$ and $n_{i,j}$ is the number of pixels of the images with the same pattern X.

The value of Z_j $(p+1)$ that minimizes this index is the new centroid of the cluster given by:

$$Z_j(p+1) = \frac{1}{N_j} \sum_{i,j} n_{i,j} X_{i,j} \quad (4)$$

where N_j is the number of pixels whose X belongs to the class of centroid Zj.

4-If Z_j $(p+1) = Z_j$ (p) when $j = 1, 2,...K$; the algorithm ends due to convergence. Otherwise, it returns to the second step.

In order to reduce processing time, the algorithm ceases when $\left\| Zj(p+1) - Zj(p) \right\| < \varepsilon$, with an ε small enough not to affect the result (considering that X elements are the whole numbers corresponding to the gray levels appearing in the images).

4. RESULTS

All MR data was acquired with a 1.5 Tesla system. The image protocol included coronal 3D T1-weighted gradient echos orthogonal to the AC-PC line (TR/TE= 24/5 ms, slice thickness= 1.5 mm); and coronal proton density (PD) and T2-weighted fast spin echos oriented aparallely (TR/TE!/TE= 3,500/32/96 ms, echo train length = 8, slice thickness = 3 mm)

Tests with different standard cluster algorithms, such as maximum and minimum distances and K-means were carried out to compare the results of the weighted K-means algorithm in a group of more than 200 sample images. Figure 1 shows the resulting images. The same characteristic vector was used to independently evaluate the performance of different algorithms.

Confusion matrices were calculated to evaluate the number of pixels correctly as well as misclassified. Along these lines, algorithms were compared with more detail than if only the recognition rate had been computed (percentage of pixels well classified).

Table 1 shows the confusion matrix resulting from the maximin algorithm. This table clearly depicts that although CSF correctly classified 92.13% of the pixels, grey matter was notoriously misclassified in about 27%. Grey matter was over classified in almost 8% of the pixels corresponding to CSF and in 18.82% of the pixels corresponding to WM. 89.10% of the pixels were underclassified as white matter. In addition, those patterns with a maximum Euclidean distance were set as initial points. As seen in Figure 1, the lack of flexibility led to unfavorable results.

Table 2 shows the confusion matrix resulting from the minimum distance algorithm. This algorithm is not stable and depends extremely on the selection of the initial points. Figure 1 shows that sometimes classification is quite acceptable,

Maximin Algorithm	Nº of pixels classified as CSF class And...	Nº of pixels classified as GM class and...	Nº of pixels classified as WM class And...
... that should have been classified as CSF	92.13 %	1.01 %	7.96 %
... that should have been classified as GM	0.00 %	89.10 %	18.82 %
... that should have been classified as WM	0.00 %	1.01 %	73.22 %

Table 1. Confusion Matrix resulting from maximin algorithm: distribution of classified pixels percentages compared with experts' classification.

Mindist Algorithm	Nº of pixels classified as CSF class and...	Nº of pixels classified as GM class and...	Nº of pixels classified as WM class and...
... that should have been classified as CSF	60.55 %	5.00 %	37.06 %
... that should have been classified as GM	0.00 %	83.24 %	4.73 %
... that should have been classified as WM	0.00 %	7.47 %	46.3 %

Table 2. Confusion Matrix resulting from minimum distance algorithm: distribution of classified pixels percentages compared with experts' classification.

Weighted K-means Algorithm	Nº of pixels classified as CSF class and...	Nº of pixels classified as GM class and...	Nº of pixels classified as WM class and...
... that should have been classified as CSF	98.77 %	0.00 %	0.00 %
... that should have been classified as GM	0.00 %	97.52 %	1.91 %
... that should have been classified as WM	0.00 %	2.48 %	98.09 %

Table 3. Confusion Matrix resulting from weighted K-means algorithm: distribution of classified pixels percentages compared with experts' classification.

whereas others it is not. In addition, the recognition rates evaluated for 10 images were really poor as seen in the confusion matrix. CSF is correctly classified only in a 60.55% in defect. White matter classification is quite poor (only 83%); and grey matter is in general over classified. Only 46.3% of the pixels were correctly classified as grey matter; a 37.06% should have been classified as CSF and a 5% as white matter.

Finally, Table 3 presents the confusion matrix resulting from the weighted K-means algorithm. In this case, a large percentage of pixels are correctly classified (above 97%). A 2.48% of grey matter pixels were misclassified as white matter; while only 2% of white matter was classified as grey matter.

The previous tables do not include the classical K-means algorithm as it yields results highly similar to those of the weighted K-means algorithm with respect to the percentages of properly classified pixels.

In spite of this, the weighted K-means algorithm is, indeed, more stable with regard to the selection of initial points than the classical version. Several tests were performed with the same image, though the selection of initial points was changed. This led to variations in the confusion matrix of less than 0.02%. Figure 2 shows the stability of the proposed algorithm in relation to the variation of its initial
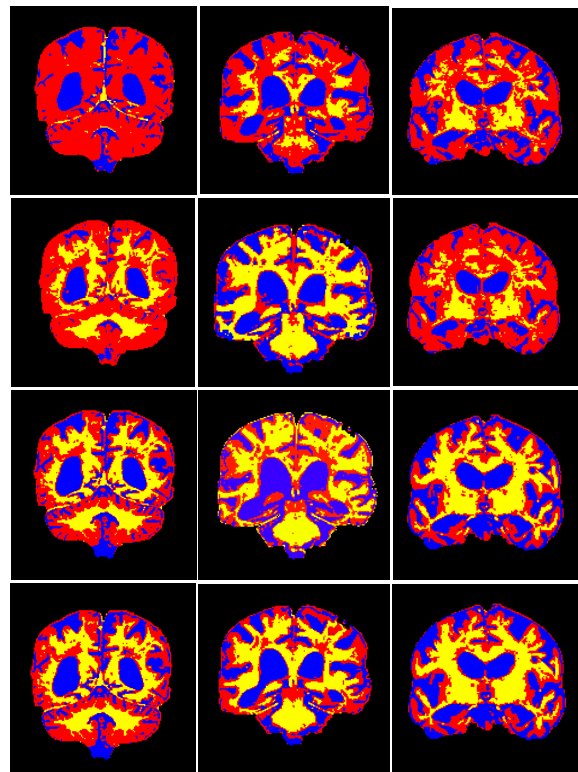

Fig.1 Coronal MR images resulting from the three different algorithms. First row: maximin algorithm results; second row: minimun distance results; third row: weighted K-means algorithm results; and forth row: experts classified images.

points. In some tests, it could be observed that the standard version inverted the centroids of the image background and gray matter. This fault is corrected in the weighted version as the great number of times the background pattern appears in the image is taken into account, notoriously modifying the calculus of the corresponding centroid.

The K-means variation herein suggested, employs a weighted means for the calculation of the class centroids, i.e., it takes into account the times a specific pattern appears in the image. This modification, results in a faster convergence of the algorithm. As a
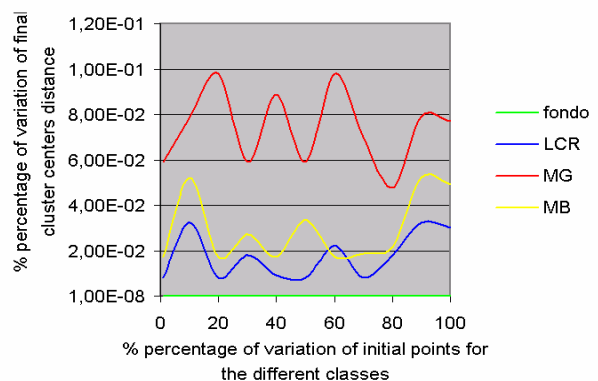

Fig.2 Percentage of variation of final cluster centers distance in relation to the variation of initial points for the different classes.
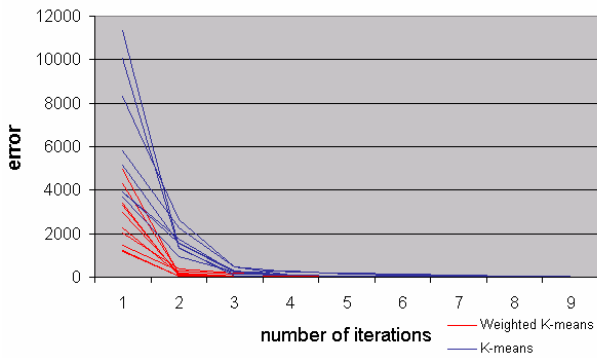
Fig.3 Error variation with respect to number of iterations in traditional K-means and weighted K-means algorithm.

Classes	Different Classifying Algorithms		
	Minimum distance	Maximin	Weighted K-means
CSF	60.55 %	92.13 %	98.77 %
Grey matter	83..24 %	89.10 %	97.52 %
White matter	46.33 %	73..22 %	98.09 %

Table 4. Recognition rate (percentage of pixels correctly classified) of the different algorithms evaluated, contrasted with experts' classification.

consequence, the number of iterations and processing time are lower, thereby compensating for the main disadvantage of the standard algorithm, and keeping stability. Figure 3 shows an error graph with regard to the number of iterations for both K-means versions, which reflects the highest convergence speed of the weighted variant.

Furthermore, the variation of the K-means proposed takes into consideration the times a specific pattern appears in the image to calculate the centroids of the classes. Since the weighted mean is used to compute the centroids, the number of operations involved get reduced. Hence processing times get reduced as well, improving the main disadvantage the standard algorithm has, though maintaining its stability.

Finally Table 4 compares the performance of minimum distance, maximin, standard and weighted K-means algorithms through their recognition rates.

Results show that the weighted K-means algorithm allows to obtain the best brain image segmentation. Feature vector elements with small occurrence in the image play a small role in classes determination. Besides, this method preserves the advantages of the standard K-means if compared with the minimum distance algorithm.

5. DISCUSSION AND CONCLUSIONS

The results previously presented prove that the weighted K-means algorithm offers the best brain image segmentation. To begin with, it prevents characteristics with very little presence from having significant incidence on the determination of classes.

Moreover the method keeps K-means advantages regarding minimum distance, since the clusters obtained depend on the prototypes manually selected from the image.

Pattern recognition techniques proved to be a powerful tool in image segmentation. In this particular application, they allowed the effective use of the complementary information provided by MR T_1, T_2 and PD-weighted images when separating the different brain tissues, to be more precise, in the discrimination of white matter, gray matter and cerebrospinal fluid (CSF). Further improvements in these techniques could lead to an even better brain atrophy quantification allowing the evaluation of the therapeutic effects of the medication given to Alzheimer's patients.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] P. Davis, A. Gray, M. Albert, et al. "The Consortium to Establish a Registry for Alzheimer's Disease, " Neurology, Vol. 42, 1992, pp. 1676-1680.

[2] N. Fox, and P. Freeborough, "Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease," J. Magn. Reson. Imaging, Vol. 7 No.6, 1997, pp. 1069-1075.

[3] J. Tanabe, D. Amend, N. Scuff, et al. "Tissue segmentation of the brain in Alzheimer Disease," Am. J. Neuroradiol., Vol. 18, 1997, pp. 115-123.

[4] S.P. Raya, "Low level segmentation of 3D resonance brain images," IEEE Trans. Med. Imaging, vol. 9 , 1990, pp. 2.

[5] S. Chen, L. Wei-chun, and C. Chen, "Medical image understanding system based on Dempster-Shafer reasoning," SPIE Biomedical Imag. Processing II, 1991, San Jose, California.

[6] Clarke, L.P. et al. (1992): 'Comparison of Supervised Pattern Recognition Techniques and Unsupervised Methods for MRI Segmentation," Proc. SPIE Medical Imaging VI Conference, Vol. 1652 , SPIE Press, Bellingham, WA, 668-677, 1992.

[7] Vannier, M.W., Brunsden, B.S., Hildebolt, C.F., Falk, D., Cheverud, J.M., Figiel, G.S., Perman, W.H., Kohn, A., Robb, R.A. and Yoffie R.L. 'Brain surface cortical sulcal lengths: quantification with three- dimensional MR imaging' Radiology, Vol 180, 479-484 1991.

[8] R. Duda, P. Hart, and D. Stork, "Pattern Classification and Scene Analysis". Ed. Wiley New York, 2000.

 [9] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: A review", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22 No.1, 2000, pp. 4-37.

[10] G. Gerig, J. Martin, and R. Kikinis, "Unsupervised tissue type segmentation of 3D dual-echo MR head data," Image Vision Computer, Vol.10, 1992, pp. 349–360.

[11] B. Alfano, B. Brunetti, et al. "Unsupervised, automated segmentation of the normal brain using multispectral relaxometric magnetic resonance approach," Magn. Reson. Med., Vol. 37, 1997, pp. 84-93.

[12] R. De La Paz, E. Herskovits, V. Gesu, et al. "Cluster analysis of medical magnetic resonance images MRI data: Diagnostic application and evaluation," SPIE Extracting Meaning from Complex Data: Processing, Display, Interaction, Vol. 1259, 1990, pp. 176 –181.

[13] Y. Wei, J. Fritts, and S. Fangting, "A hierarchical image segmentation algorithm,"

ICME '02. Proceedings. IEEE International Conference, Vol. 2, 2002, pp. 221-224.

[14] J.C. Bezdek, "Pattern Recognition with fuzzy objective function algorithms," New York: Plenum Press, 1981.

[15] K. Krishna, K.R. Ramakrishnan, M.A.L. Thathachar, "Vector quantization using genetic K-means algorithm for image compression," Proceedings of the International Conference on Information, Communications and Signal Processing, ,Vol. 3, 1997, pp. 1585-1587.

[16] K. Krishna, and M. Narasimha Murty, "Genetic K-means algorithm," IEEE Trans. on Systems, Man and Cybernetics, Vol. 29,1999, pp. 3.

[17] L.O Hall,, A.M. Bensaid, L.P. Clarke, R.P. Velthuizen, M.S. Silbiger and J.C. Bezdek, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," IEEE Trans. on Neural Networks, Vol. 3, No. 5, 1992, pp. 672- 682.

[18] J. Tou and R. Gonzalez, "Pattern Recognition Principles," Ed. Adisson-Wesley, 1974.

[19] T. Bartlett, M. Vannier, D. Mc Keel, M. Gado, C. Hildebolt and R. Walkup, "Interactive segmentation of cerebral gray matter, white matter, and CSF: photographic and MR images," Comp. Med. Imag. Graphics, Vol. 18 No. 6, 1994, pp. 449-460.

[20] G. Abras, V. Ballarin, y M.González, "Aplicación de Reconocimiento de Patrones en la Clasificación de Tejido Cerebral," Actas del Congreso Argentino de Bioingeniería. Tafí del Valle. Septiembre 2001. (Publicadas en CD).