Statistical segmentation of multidimensional brain datasets

M Desco[a], JD Gispert[a], S Reig[a], A Santos[b], J Pascau[a], N Malpica[b], P Garcia-Barreno[a]

[a]Hospital General Universitario "Gregorio Marañón", E-28007 Madrid, Spain
[b]Universidad Politécnica de Madrid, E-28040 Madrid, Spain

ABSTRACT

This paper presents an automatic segmentation procedure for MRI neuroimages that overcomes part of the problems involved in multidimensional clustering techniques like partial volume effects (PVE), processing speed and difficulty of incorporating *a priori* knowledge. The method is a three-stage procedure: 1) Exclusion of background and skull voxels using threshold-based region growing techniques with fully automated seed selection. 2) Expectation Maximization algorithms are used to estimate the probability density function (PDF) of the remaining pixels, which are assumed to be mixtures of gaussians. These pixels can then be classified into cerebrospinal fluid (CSF), white matter and grey matter. Using this procedure, our method takes advantage of using the full covariance matrix (instead of the diagonal) for the joint PDF estimation. On the other hand, logistic discrimination techniques are more robust against violation of multi-gaussian assumptions. 3) *A priori* knowledge is added using Markov Random Field techniques. The algorithm has been tested with a dataset of 30 brain MRI studies (co-registered T1 and T2 MRI). Our method was compared with clustering techniques and with template-based statistical segmentation, using manual segmentation as a 'gold-standard'. Our results were more robust and closer to the gold-standard.

Keywords: MRI segmentation, robust segmentation, EM Algorithm, Markov Random Fields, Partial Volume, logistic regression.

1. INTRODUCTION

The study of many brain disorders involves tissue segmentation from magnetic resonance (MR) images. The study of schizophrenia, for instance, pays great attention to the accurate quantification of CSF in the frontal lobes of the brain as an indirect measure of neurodegeneration[1].

Currently, in many clinical studies segmentation is still mainly manual or strongly supervised by a human expert. The level of operator supervision impacts the performance of the segmentation method in terms of time consumption, leading to unfeasible procedures for large datasets, and increases observer variability, deteriorating the precision of the analysis of the segmentation.

Automatic and reliable intensity based segmentation is difficult to achieve due to the overlap of MR intensities of the different tissue types, the acquisition noise, and by the presence of partial volume effects (PVE). This artefact derives from the finite resolution of MR scanners, resulting in voxels that contain more than one tissue type. MR images may present different contrasts –i.e. different intensity levels for the same tissue– depending on the particular sequence selected, thus being able to provide multispectral datasets. MRI data can also be combined with other modalities, such as PET or CT, in order to take full profit of their complementary information. Thus, multispectral or multimodality MR segmentation is becoming more important to maximize tissue differentiation, as well as to map functional metrics.

A review of automatic segmentation methods can be found in [2,3]. Recently, great attention is being paid to multispectral segmentation [4], Expectation-Maximization (EM) algorithm [5], partial volume modelling [6], robust estimators [7], and logistic regression [8] for MRI segmentation.

Correspondence: Dr. M. Desco. Hospital General Universitario "Gregorio Marañón". c/ Dr. Esquerdo, 46. E-28007. Madrid. SPAIN. e-mail: desco@mce.hggm.es; web: http://www.hggm.es/image.

Medical Imaging 2001: Image Processing, Milan Sonka, Kenneth M. Hanson, Editors,
Proceedings of SPIE Vol. 4322 (2001) © 2001 SPIE · 1605-7422/01/$15.00

The segmentation method presented in this article models the intensity distributions of multispectral MRI scans as a mixture of Gaussians. To this end, the Expectation-Maximization (EM) algorithm was modified in order to overcome some of the main drawbacks in automatic intensity based segmentations. The use of robust estimators, logistic regression techniques and full covariance matrices shields to some extent the algorithm against model assumption violations and PVE. Another important feature of this algorithm is the inclusion of *a priori* knowledge by means of a Markov Random Field filtering while avoiding the need of anatomical or statistical atlases. The method has been validated on 30 MRI scans of schizophrenic patients. Results were compared with a manual segmentation used as gold-standard, a clustering technique and the standard algorithm provided with SPM99[9].

2. ALGORITHM

The proposed segmentation method has been applied to a multispectral MRI dataset –T1- and T2-weighted–, and involves the three following steps: 1) automatic background and skull/scalp extraction, 2) tissue segmentation by means of an EM algorithm employing logistic regression techniques, and 3) application of a Markov Random Field filter.

Background and skull/scalp extraction has several purposes. Scalp tissues show intensity levels that overlap with brain tissues and, therefore, a correct classification is impossible to achieve only by means of their grey level. Besides, background and skull present very low signal levels both in T1- and T2-weigthed scans. These regions are dominated by noise, whose intensity distribution is governed by a Rice distribution and, consequently, cannot be accurately modelled by a Gaussian [10]. Previous removal of these non-interesting voxels prevents against violation of multi-gaussian assumptions and reduces the number of voxels to classify resulting in a more accurate and faster result.

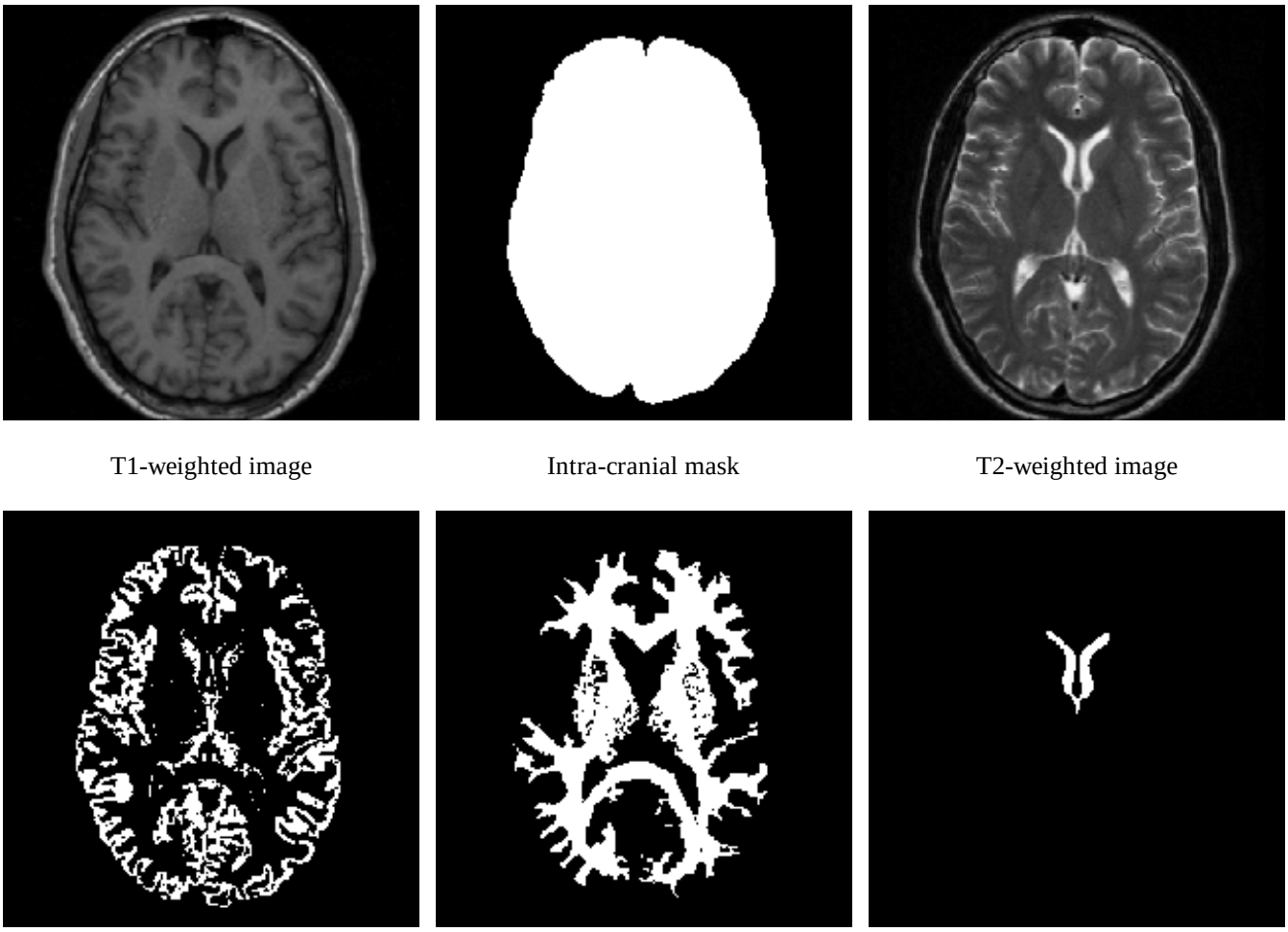
This skull/scalp pre-segmentation is achieved using region growing techniques with fully automatic seed and threshold selection. Voxels are classified as intracranial tissue if a) the sum of intensities in images is above a certain threshold and b) if they are connected to the central voxels of the image. Similarly, white and grey matter are segmented from the T1-weighted image and the ventricles are segmented from the T2-weighted image (Fig. 1).

At this point, the mean, variance and the prior probability of every tissue type is computed, using the median as a suitable robust estimator –trimmed or Winsorized mean could have also been used–.

The second step in the method performs a voxel classification based on the EM algorithm [11] in which a logistic regression technique has been incorporated. Voxel intensities are assumed to be generated by a mixture of known distributions –one for each tissue type– whose parameters are not known. The EM algorithm is an iterative method for finding the maximum-likelihood estimate of an underlying distribution from a given dataset.

The algorithm operates as follows. Let the intensities in the MR images be indicated by an array $x = \{x_{11}, \dots, x_{n1}; \dots; x_{1d}, \dots, x_{nd}\}$ of size N x D where N is the total number of voxels in one image and D is the number of images used –two in this work–. Let the underlying segmentation be represented by $y = \{y_{11}, \dots, y_{n1}; \dots; y_{1d}, \dots, y_{nd}\}$ where y_i denotes the tissue type that voxel *i* belongs to. The number of distributions is assumed to be K and $y_i = e_k$ for a given *k*, $1 \leq k \leq K$ where e_k is a unit vector whose *k*th component is equal to one, being all the other components zero. The underlying distribution is described by a probability density function $f(y|\Phi_y)$ –in this case a Gaussian distribution– that is parameterised by the parameter set Φ_y. Then, the Bayes' rule can be applied to the initial estimation of the mean, variance and prior probability of each tissue type that were obtained in the first step, allowing to calculate the posterior conditional probability $p(y_i|x_i,\Phi_y)$, of the voxel *i* with intensity level x_i belonging to the tissue type y_i, given the parameter set Φ_y,:

$$p(y_i = e_k \mid x_i, \Phi_y) = \frac{p(x_i \mid y_i = e_k, \Phi_y) \cdot p(y_i = e_k)}{\sum_{j=1}^{K} p(x_i \mid y_i = e_j, \Phi_y) \cdot p(y_i = e_j)} \tag{1}$$



T1-weighted image | Intra-cranial mask | T2-weighted image

Gray Matter mask | White Matter mask | Ventricle mask

Figure 1. Rough tissue masks are used to initially estimate the mean and variance of each tissue

In the particular case of a mixture of Gaussians:

$$p(x_i \mid y_i = e_k, \Phi_y) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2\sigma_k^2} (x_i - \mu_k)^2 \right) \tag{2}$$

Calculation of $p(y_i|x_i,\Phi_y)$ is called the Expectation step. These conditional probabilities are then used to re-calculate the new set of estimates Φ_y that maximizes the likelihood –the Maximization step–. These new estimates are given by:

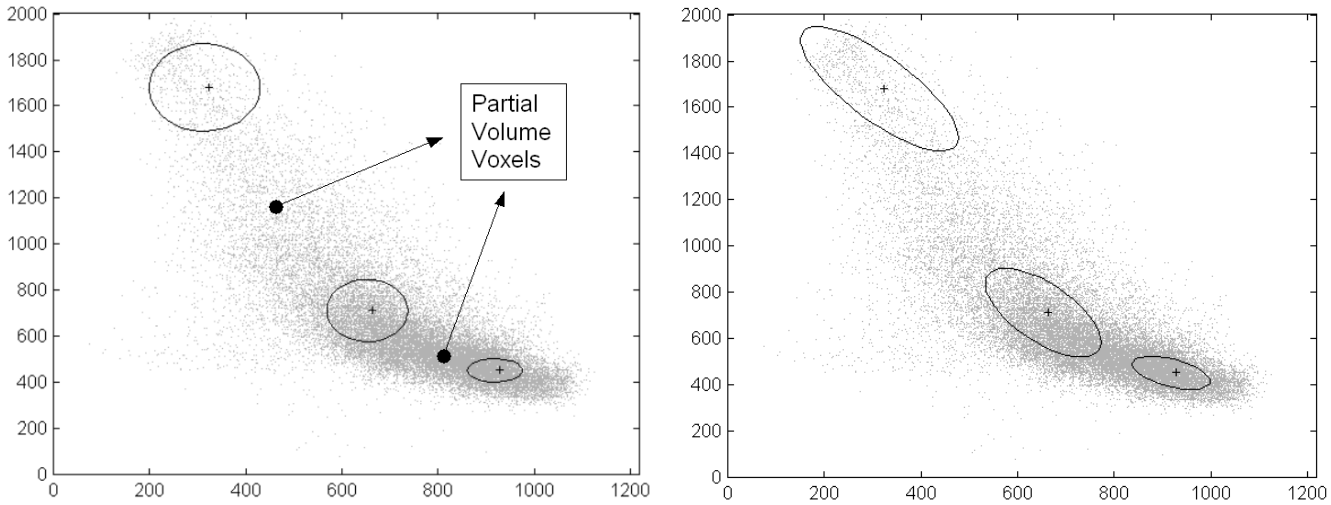
$$p(y_k) = \frac{1}{N} \sum_{i=1}^{N} p(y_i = e_k \mid x_i, \Phi_y) \tag{3}$$

$$\mu_k = \frac{\sum_{i=1}^{N} p(y_i = e_k \mid x_i, \Phi_y) \cdot x_i}{\sum_{i=1}^{N} p(y_i = e_k \mid x_i, \Phi_y)} \tag{4}$$

$$\Sigma_k = \frac{\sum_{i=1}^{N} p(y_i = e_k \mid x_i, \Phi_y) \cdot (x_i - \mu_k) \cdot (x_i - \mu_k)^T}{\sum_{i=1}^{N} p(y_i = e_k \mid x_i, \Phi_y)} \tag{5}$$

In which $p(y_k)$ is the prior probability of the tissue type k, μ_k is the mean and Σ_k is the full covariance matrix.

Noise in high signal regions of MR images can be modelled as a random process governed by a Gaussian distribution. Consequently, noise correlation between two different MR acquisitions is not expected and statistical independence of the voxels is assumed. Nevertheless, since the signal intensity in partial volume voxels is the weighted mean of the intensities of the tissue types contained, a correlation between classes is clearly introduced, as can be seen in Fig. 2. Modelling of this correlation leads to the use of full covariance matrices instead of diagonal ones that cannot take this effect into consideration. Another consequence of the bias introduced by PVE is the violation of the multi-Gaussian assumption. Due to this bias, the estimation of the parameters of the distribution mixture may suffer from lack of accuracy if the estimators are not robust enough. For this purpose, the initial robust estimation of the mean is kept and the EM algorithm is used to only estimate the prior class probabilities and the covariance matrices. In spite of this, PVE still biases severely the covariance matrix estimation, and a further correction is needed. To deal with biased estimations an additional step is introduced into the EM algorithm: the calculation of the posterior probabilities $p(y_i|x_i,\Phi_y)$ by means of the logistic function[12].



(a): Diagonal covariance matrices cannot accurately model partial volume effect.

(b): Full covariance matrices allow better modelling of data containing partial volume effect.

Figure 2. Joint histogram of the T1-weighted (x-axis) and T2-weighted (y-axis). Crosses indicate Gaussian means and ellipses the standard deviation of the tissue classes, given the covariance matrices.

In the case that covariance matrices were equal for all the classes, equation (1) could be re-written as follows: [13]

$$p(y_i = e_k \mid x_i, \Phi_y) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \sum_{k \neq i} \beta_k \cdot x_i\right)\right)} \tag{6}$$

Under this assumption, all β_k can be analytically calculated. There are strong arguments in favour of logistic discrimination over standard estimators under non-normal conditions [13]. When the assumption of equal covariance matrices does not hold, these parameters must be estimated by maximizing a likelihood function, as in the case of a mixture of Gaussians. The estimation of the posterior probabilities using (6) instead of (1) introduces a risk function that modulates the probability density functions according to the actual data.

Finally, the third step of the segmentation scheme involves the filtering of the posterior probabilities obtained in the previous step with a Markov Random Field (MRF)-based technique, in order to include both spatial and *a priori* information. In a MRF the probability of a voxel *i* to belong to a certain tissue y_i only depends on the tissue-probabilities of its neighbours $y' \in N_i$. For instance, a voxel surrounded of white matter must have a high probability to belong to the same class, while it is very unlikely that it belongs to a non-brain tissue. This probability follows a Gibbs distribution [14]:

$$f(y = e_k \mid \Phi_{mrf}) = \frac{\exp\left[-U_{mrf}(e_k \mid p_{Ni}, \Phi_{mrf})\right]}{\sum_{j=1}^{K} \exp\left[-U_{mrf}(e_j \mid p_{Ni}, \Phi_{mrf})\right]} \tag{7}$$

$$U_{mrf}(y \mid \Phi_{mrf}) = \sum_{i=1}^{N} U_{mrf}(y_i \mid y' \in N_i, \Phi_{mrf}) \tag{8}$$

$$U_{mrf}(y_i \mid y' \in N_i, \Phi_{mrf}) = e_{ik}^T G g_i \tag{9}$$

The expression $U_{mrf}(e_k|p_{Ni}, \Phi_{mrf})$ denotes a potential that increases with disparity between a voxel and its neighbourhood. This is represented by the vector g_i which counts the number of neighbours of *i* that belong to each class and by the matrix *G* which sets the potential between tissue types. The elements of *G* are the parameters to be estimated at this step. MRF provides a methodology to smooth homogeneous regions while preserving edges and allowing for introducing *a-priori* knowledge by penalizing unlikely tissue transitions.

In its present implementation, our algorithm performs an optimisation of the parameter set of the EM algorithm, keeping fixed the logistic regression and MRF parameters. The resulting segmentations are presented in the Fig. 3.

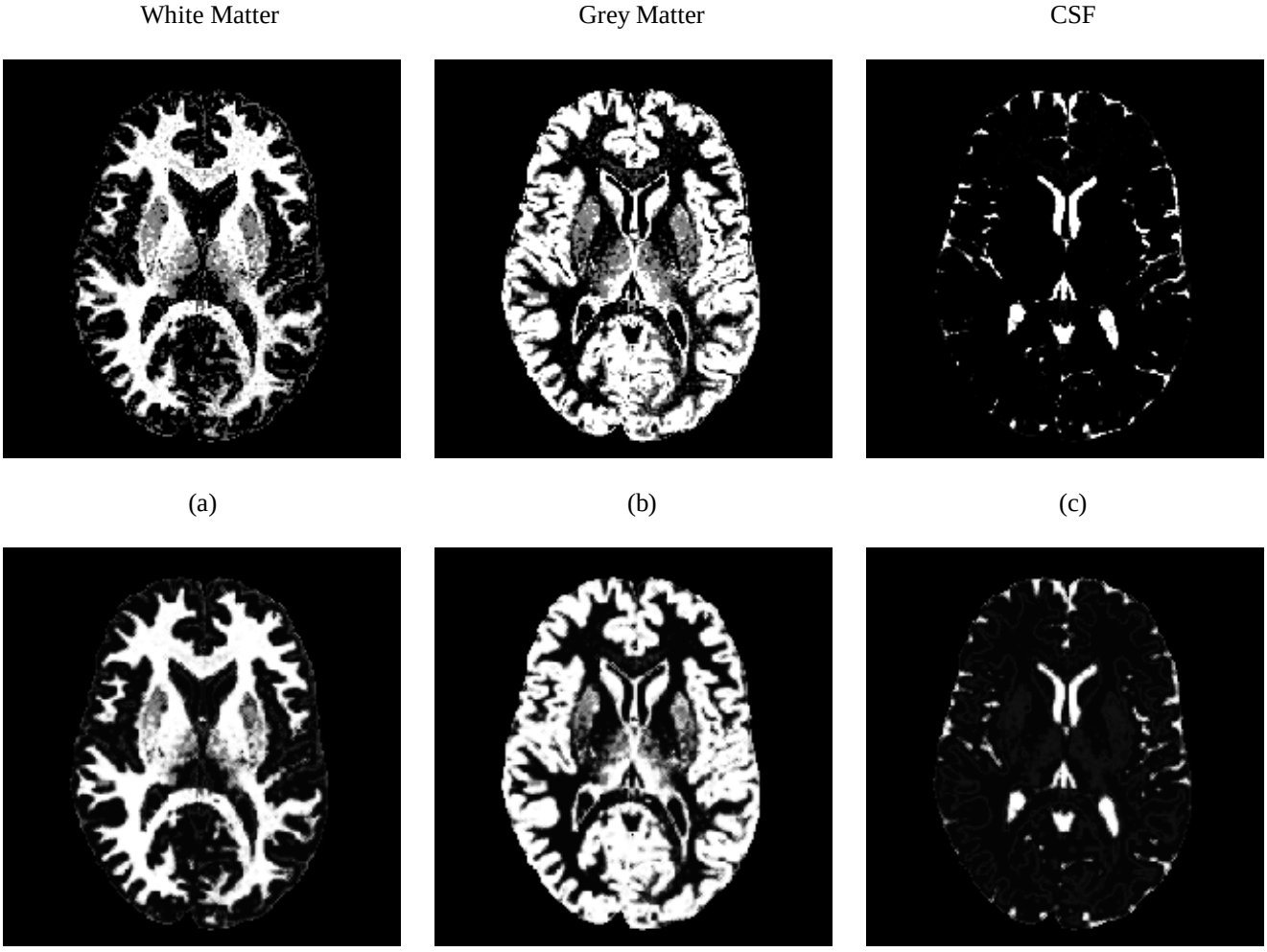


Figure 3. (a),(b),(c): Posterior probability images for white matter, grey matter and CSF
(d), (e), (f): Same images after EM algorithm with posterior MRF-based filtering

3. VALIDATION

The algorithm has been tested on a dataset of 30 MRI scans of 15 schizophrenic patients, using one slice per case. MRI studies were obtained on a Philips ACS Gyroscan 1.5 T scanner. Patients underwent two MRI procedures: a gradient echo T1-weighted 3D sequence and a turbo spin echo T2-weighted 2D sequence. Matrix size ranged from 256 x 256 to 512 x 512 and slice thickness ranged from 1.1 mm to 1.5 mm in T1 studies and from 3.3 mm to 5.5 mm in T2 studies. T1 images were registered and re-sliced using trilinear interpolation to T2 images by means of a mutual information registration algorithm [15]. Segmentation of multispectral or multimodality datasets is particularly sensitive to registration errors. The mutual information algorithm used in this work has been developed and previously tested in our group [16] and has proven to yield more accurate and stable results than other algorithms.[9, 17]

In order to validate the accuracy and reliability of the logistic EM segmentation, multispectral slices were classified into white matter, grey matter, CSF and non-brain tissue by four segmentation methods: 1) manual segmentation by an expert, which was used as a gold standard to establish a ground truth for all comparisons, 2) a supervised clustering technique based on *K-Nearest Neighbours* [18] algorithm with K=4 that had been previously tested in our group[19], 3) an automatic segmentation obtained through Bayes' rule from the prior probabilities calculated with SPM99 [20], and 4) the logistic EM

algorithm. Afterwards, the tissue masks obtained were compared in terms of: a) Overlap, b) False Negatives –type I error–, and c) False Positives –type II error–. The first two segmentation methods are rigid and unable to model PVE. In order to obtain comparable results, voxels in the SPM99 and the logistic EM segmentation were uniquely assigned to the most probable class. Finally, the spatial distribution of the misclassified voxels for a single patient is presented.

The overlap rate is defined as [21] $2 \cdot V^k_{ov}/(V^k_g + V^k_s)$, where V^k_{ov} represents the number of voxels that overlap for class *k*, and V^k_g and V^k_s denotes the total number of voxels in each segmentation. This metric is able to quantify the goodness of the segmentation but does not provide any information about the error type of the misclassified voxels. For this reason, the metric was complemented to detect the direction of the error –type I or type II–. Voxels in a given segmentation not present in the gold standard are called false positives and increase the type I error. Alternatively, voxels in the gold standard that do not appear in the given segmentation are called false negatives and increase type II error. Consequently, false positive and false negatives rates are defined as $V^k_{fp}/(V^k_g + V^k_s)$ and $V^k_{fn}/(V^k_g + V^k_s)$. It should be stressed that in medical applications a type of error is usually more important than the other. Overlap and error rates are presented in Fig. 4.

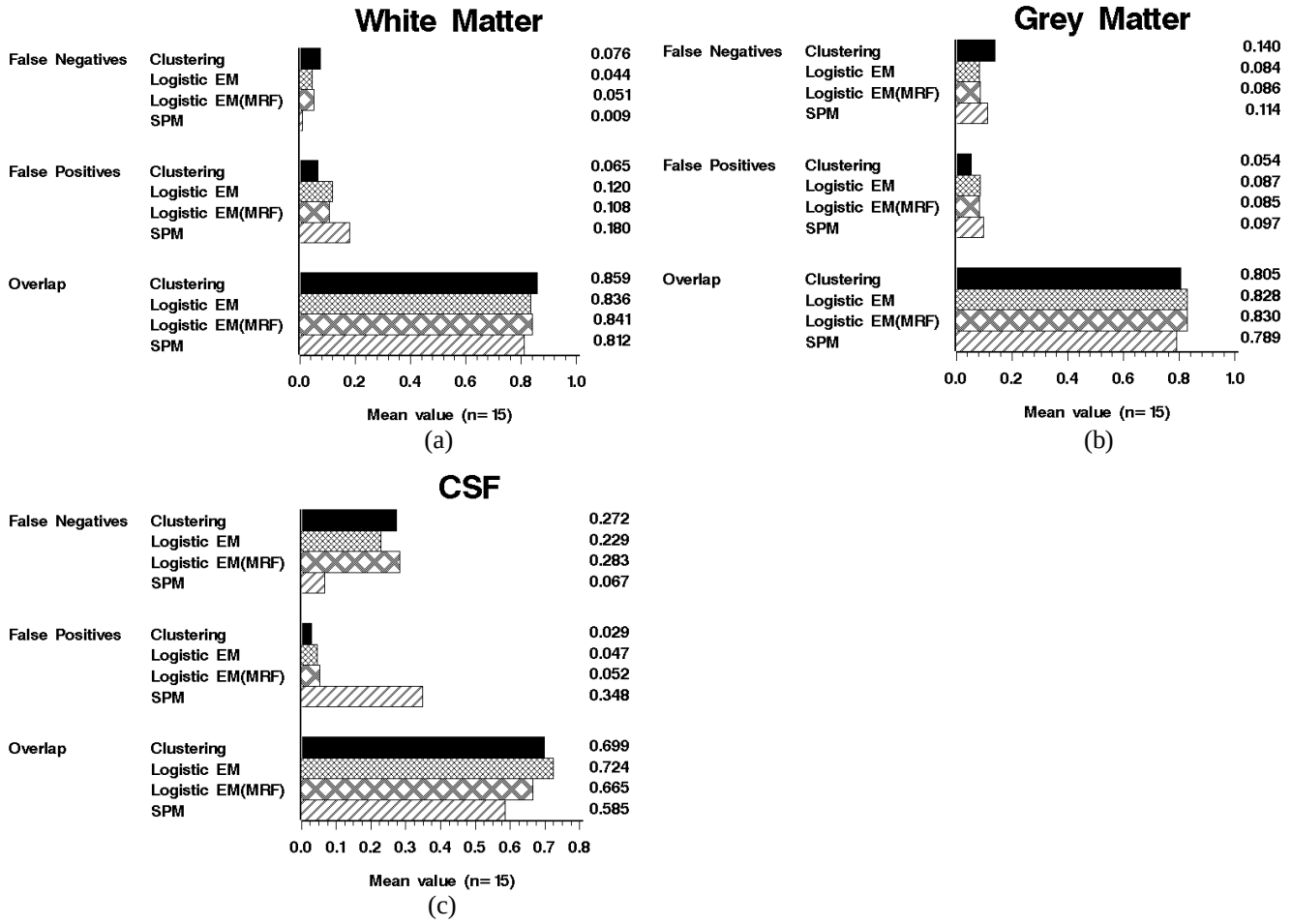


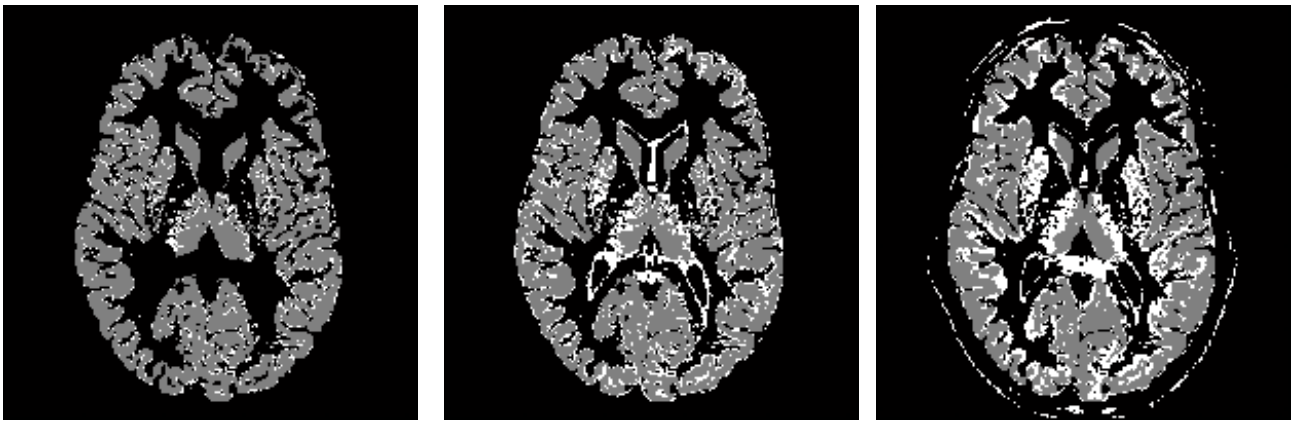
Figure 4. Comparison of the segmentation procedures. Figures represent the average values (N=15) of the False Negative, False Positive and Overlap rates for every tissue.

The spatial distribution of the misclassifications is presented in Fig. 5 with data from a single patient.

(a): White matter segmented by Clustering.

(b): White matter segmented by the logistic EM algorithm.

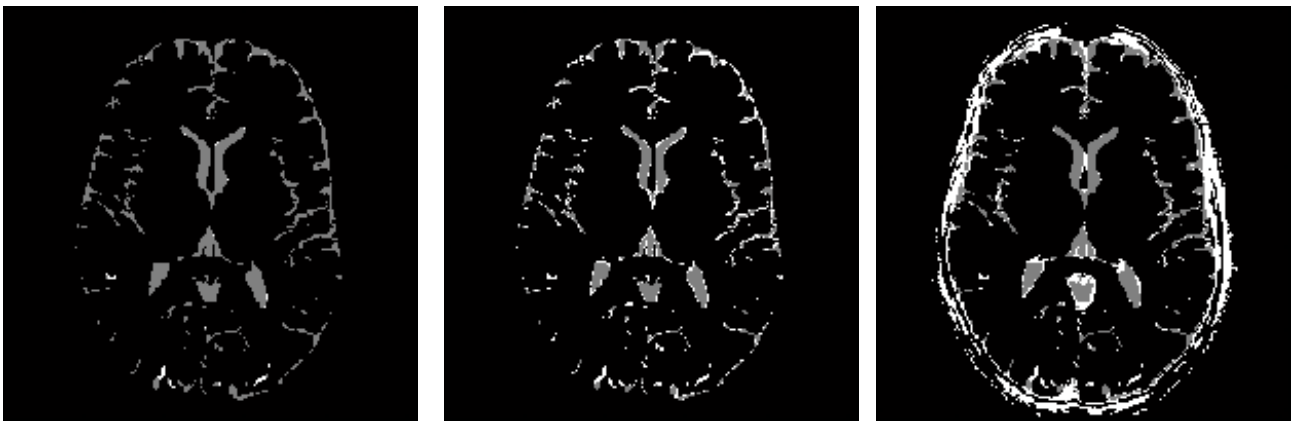
(c): White matter segmented by SPM99.



(d): Grey matter segmented by Clustering.

(e): Grey matter segmented by the logistic EM algorithm.

(f): Grey matter segmented by SPM99.



(g): CSF segmented by Clustering.

(h): CSF segmented by logistic EM.

(i): CSF segmented by SPM99.

Figure 5. Example of the spatial distribution of the error, using manual segmentation used as gold-standard. Overlapping voxels are showed in grey and misclassified voxels in white.

4. DISCUSSION

Logistic EM algorithm showed the best overall performance, except in the case of white matter segmentation. This tissue was better segmented by the clustering method with manual seed selection and editing. Fig. 5b illustrates that the misclassification was mainly localized in the basal ganglia and the occipital lobe. Segmentation of sub-cortical structures is very complicated since their limits are not clearly defined in the image –especially in the thalamus– and their intensity level is not equal to that of cortical grey matter. Consequently, the clustering technique takes advantage of the human expert knowledge of brain morphology. SPM segmentation also failed to detect these regions. Misclassification in occipital lobe probably derives from the presence of an inhomogeneity artefact. This effect, inherent to MRI [22], is not always visible for a human observer but can cause serious misclassifications on intensity-based segmentations [23]. Inclusion in the model of the bias field could theoretically improve our logistic EM segmentation.

In the segmentation of CSF, logistic EM showed its best performance in comparison to the other methods (Fig. 4c). This superior performance can be explained by the use of robust estimators and logistic regression, especially useful when a high percentage of partial volume voxels biases its distribution from a Gaussian.

The SPM99 method showed a poor performance when segmenting CSF, with higher false positive rate (Fig. 4c). This is caused by the misclassification of many scalp voxels as CSF. Intensity based methods cannot accurately segment the scalp, which presents intensities that overlap brain tissue voxels. The probability template used by this algorithm does not exclude scalp voxels. Besides, this method may suffer from a lack of accuracy when registering the MRI to an average template, especially in the case of severely distorted brains. This effect may be accentuated in this work because of the use of MRI scans of schizophrenic patients for the validation.

The MRF-based filter implemented could not significantly improve the logistic EM in the white and grey matter cases, and even decreased the overlap of the CSF segmentation. This behaviour suggests wrong MRF parameters selection, and thus pointing to the need to include their proper estimation in the algorithm.

These considerations suggest that the logistic EM algorithm is particularly indicated for the quantification of CSF. Since the presence of non-Gaussian distributions is not critical, other medical images different from MR could be used. For instance, the joint use of anatomical MRI with some functional imaging method, such as PET, could result in a more accurate characterization of the functional behaviour of the different tissues. Moreover, the use of full covariance matrices makes it applicable to any imaging technique prone to show correlated noise, such as parametric or perfusion images.

The logistic EM algorithm is open to several improvements. The inclusion in the algorithm of the bias field correction and optimisation of the logistic and MRF parameters should result in an increase of its accuracy. The implementation of these refinements together with the excellent results reported in this work make the logistic EM algorithm a very promising technique.

ACKNOWLEDGMENTS

This work was supported in part by the Spanish National research funding programs TIC99 #1085-C02 and FIS #00/0036; Comunidad de Madrid IIIPRICYT and CM#08.1/0049/98; and by 'Fundació La Caixa' #99/042-00. The authors would like to thank Dr. Vicente Molina for contributing the MRI scans.

REFERENCES

[1] R. W. McCarley, C. G. Wible, M. Frumin, Y. Hirayasu, J. J. Levitt, I. A. Fischer, and M. E. Shenton, “MRI anatomy of schizophrenia,” *Biol. Psychiatry*, 45, pp. 1099-119, 1999.

[2] L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger, “MRI Segmentation: Methods and Applications,” *Magnetic Resonance Imaging*, vol. 13, pp. 343-368, 1995.

[3] J. S. Duncan and N. Ayache, “Medical Image Analysis: Progress over Two Decades and the Challenges Ahead,” *IEEE Trans. Patt. Anal. Mach. Intel.*, 22, pp. 85-105, 2000.

[4] A. Lundervold, T. Taxt, L. Ersland, and A. M. Fenstad, "Volume distribution of cerebrospinal fluid using multispectral MR imaging," *Med. Image Anal.*, 4, pp. 123-136, 2000.
[5] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated Model-Based Tissue Classification of MR Images of the Brain," *IEEE Trans. Med. Imag*, 18, pp. 897-908, 1999.
[6] D. H. Laidlaw, K. W. Fleischer, and A. H. Barr, "Partial-Volume Bayesian Classification of Material Mixtures in MR Volume Data Using Voxel Histograms," *IEEE Trans. Med. Imag*, 17, pp. 74-86, 1998.
[7] P. Schroeter, J.-M. Vesin, T. Langenberger, and R. Meuli, "Robust Parameter Estimation of Intensity Distributions for Brain Magnetic Resonance Images," *IEEE Trans. Med. Imag*, 17, pp. 172-186, 1998.
[8] E. T. Bullmore, J. Sucking, S. Overmeyer, S. Rabe-Hesketh, E. Taylor, and M. J. Brammer, "Global, Voxel, and Cluster Tests, by Theory and Permutation, for a Difference Between Two Groups of Structural MR Images of the Brain," *IEEE Trans. Med. Imag*, 18, pp. 32-42, 1999.
[9] K. J. Friston, J. Ashburner, J.-B. Poline, C. D. Frith, J. D. Heather, and R. S. J. Frackowiak, "Spatial Registration and Normalization of Images," *Hum. Brain Mapp.*, 2, pp. 165-189, 1995.
[10] J. Sijbers, A. J. d. Dekker, J. V. Audekerke, M. Verhoye, and D. V. Dyck, "Estimation of the noise in magnitude MR images," *Magn. Reson. Imaging*, vol. 16, pp. 87-90, 1998.
[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc. Ser. B.*, vol. 39, 1977.
[12] P. McCullagh and J. A. Nelder, *Generalized Linear Models*. New York, NY: Chapman & Hall, 1989.
[13] W. R. Dillon and M. Goldstein, *Multivariate analysis*: John Wiley & Sons, 1984.
[14] S. Z. Li, *Markov Random Field Modeling in Computer Vision*: Springer-Verlag, 1995.
[15] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, 16, pp. 187-98, 1997.
[16] J. J. Vaquero, M. Desco, J. Pascau, A. Santos, I. Lee, J. Seidel, and M. Green, "PET and CT Image Registration of the Rat Brain and Skull using the AIR Algorithm," presented at IEEE Nuclear Science Symposium and Medical Imaging Conference (in press), Lyon (France), 2000.
[17] R. Woods, J. Mazziota, and S. Cherry, "MRI-PET registration with automated algorithm," *J. Comput. Assist. Tomogr.*, 17, pp. 536-546, 1993.
[18] J. M. Keller, M. R. Gray, and J. A. Givens, "A Fuzzy K-Nearest Neighbor Algorithm," *IEEE Trans. Sys. Man Cybern.*, SMC-15, 1985.
[19] J. M. Calderon, "Segmentacion de imagenes de resonancia magnetica mediante tecnicas estadisticas," *Univ. Politécnica de Madrid. Escuela Técnica Superior de Ingenieros de Telecomunicación*. Master Thesis, 1998.
[20] J. Ashburner, K. Friston, A. Holmes, and J.-B. Poline, "Statistical Parametric Mapping," *The Wellcome Dept. of Cognitive Neurology. Univ. College London, London, U.K. Available in: http://www.fil.ion.ucl.ac.uk/spm/*.
[21] A. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in MR images: Method and Validation," *IEEE Trans. Med. Imag*, 13, pp. 716-724, 1994.
[22] J. G. Sled and G. B. Pike, "Understanding intensity nonuniformity in MRI" *Proc. Medical Image Computing Computer-Assisted Intervention - MICCAI'98 (Lecture notes in computer science)*, Berlin, Gemany, 1998.
[23] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated Model-Based Bias Field Correction of MR Images of the Brain," *IEEE Trans. Med. Imag*, 18, pp. 885-896, 1999.