

SEGMENTATION AND CLASSIFICATION OF BRAIN SPECT IMAGES USING 3D MARKOV RANDOM FIELD AND DENSITY MIXTURE ESTIMATIONS

M. Mignotte ‡ *J. Meunier* ‡ *J.-P. Soucy* • *C. Janicki* •

‡ DIRO, Département d'Informatique et de Recherche Opérationnelle, C.P. 6128,
Succ. Centre-Ville, Montréal, Canada (Quebec), H3C 3J7.

• Ottawa Hospital - Civic Campus ; CHUM - HND.

E-mail : <mignotte,meunier>@iro.umontreal.ca

ABSTRACT

Thanks to its ability to yield functionally rather than anatomically-based information, the SPECT imagery technique has become a great help in the diagnostic of cerebrovascular diseases. Nevertheless, SPECT images are very noisy and consequently their interpretation is difficult. In order to facilitate this visualization, we propose an unsupervised 3D Markovian model allowing to segment a brain SPECT image into three classes, corresponding to the three existing cerebral tissues, respectively ; “cerebrospinal fluid”, “white matter” and “grey matter”. This unsupervised Markovian model relies on a preliminary distribution mixture parameter estimation step which takes into account the diversity of the laws in the distribution mixture of a SPECT image. Next, in order to help the classification of these images, some features extracted from this segmentation map and the distribution mixture parameters are then exploited to constitute a discriminant feature vector for each image from our database. These feature vectors are then clustered into two distinct classes, namely ; “healthy brains” and “diseased brains” (i.e., brains with possible cerebrovascular diseases) by using once more a distribution mixture-based clustering procedure. The effectiveness of this classification scheme was tested on a database of 46 healthy and diseased brain images. The rate of good classification (74%) indicates that the proposed method may be useful in a first screening for a brain disease prior to a more thorough examination by the nuclear physician.

1. INTRODUCTION

SPECT images (Single Photon Emission Computed Tomography) are obtained by the measure of radiations (gamma rays) coming from radioactive isotopes injected in the human body. Contrary to other clinical medical imaging techniques, such as X-ray, CT (Computer Tomography) scanning, MRI (Magnetic Resonance Imaging), etc., this imagery process is able to give functionally rather than anatomically-based information, such as the 3D metabolic behavior of human brain, by visualizing the level of blood flow of a set of cross-sectional images. This study of Regional Cerebral Blood Flow can aid in the diagnostic of cerebrovascular diseases (e.g., Alzheimer’s disease, Parkinson’s disease, etc.) by indicating lower, or abnormal higher, metabolic activity in some brain regions. Due to the imaging process, SPECT is the cheapest but also the lowest resolution 3D functional visualization technique. Resulting cross-sectional SPECT images are blurred and consequently, their interpretation is often difficult and subjective.

In order to facilitate this interpretation and help the clinical diagnostic, we propose a segmentation model allowing to segment the SPECT image into three kinds of regions, each one associated to a specific brain anatomical tissue, respectively ; “CSF” or cerebrospinal fluid, “white matter” and “grey matter” [1]. In order to take into account the available 3D contextual information of these images and to improve the segmentation map, we propose a 3D Markovian segmentation model. In order to make this segmentation “unsupervised”, the proposed segmentation model rely on a preliminary estimation step which estimates the grey level statistical distribution parameter of each aforementioned class. In our application, this segmentation map result and the distribution mixture estimated parameters of each brain SPECT image from our data set can efficiently be exploited to achieve a rough pre-classification into two classes, namely, “healthy brain” and “diseased brain” (i.e., brains with possible cerebrovascular diseases), by a classical pattern classification approach. This classification step is the second problem we deal with in this paper.

Once the feature extraction step is achieved, different kinds of classification methods can be applied. A widely used approach consists in using supervised classifiers. In these approaches, the “prototypical” representation of each class is determined *a priori* thanks to a training data set. Among the commonly used methods, we can cite for instance neural networks, Bayesian classifiers or the K-nearest neighbor classifier. An inherent drawback of these methods is due to their “supervised” characteristic and the assumptions made during the training step ; these methods require a reliable pre-classification step of the training data set. Besides, they assume that the chosen learning set is large enough and sufficiently informative to represent the classes to be recognized. Another approach consists in using an unsupervised classification method. In this case, pattern classification can then be viewed as a cluster analysis problem within the feature space. To this end, we can use the well known K-means clustering procedure [2]. Nevertheless, one of the major drawback of this algorithm is to assume, often wrongly, the presence of spherical clusters with equal volume. In order to overcome this drawback and take into account the scattering variability of each cluster, an alternative approach, adopted in our application, consists in assuming that each sample of a same class is a realization of an unknown random process with a distribution law whose parameters are unknown. In this way, the clustering problem requires to solve, in a preliminary step, a parameter estimation problem of a multivariate distribution mixture. Then, given the mixture model parameters, we can use any statistical criterion in order to classify the unknown samples.

Another fundamental problem in pattern classification, discussed in this paper, is the feature extraction step or the selection of features which are most effective in producing an optimum cluster separability. We investigate this problem by combining efficiently an unsupervised distribution mixture-based clustering procedure with an algebraic method of data processing called the Fisher's method.

This paper is organized as follows. In Section 2, we detail the distribution mixture parameter estimation procedure. Section 3 presents the 3D segmentation step and some segmentation results. In Section 4, we describe the feature vector extraction step. Sections 5 and 6 present the feature classification step and the proposed dimensionality reduction strategy.

2. MIXTURE PARAMETER ESTIMATION

Consider a couple of random fields $Z = (X, Y)$, where $Y = \{Y_s, s \in S\}$ represents the field of observations located on q lattices S of N sites s (associated to the N pixels of the q transversal slices of the 3D SPECT image), and $X = \{X_s, s \in S\}$ the label field (related to the qN class labels of the segmented image). Each Y_s takes its value in $\{0, \dots, 255\}$ (256 grey levels), and each X_s in $\{e_1 = \text{"CSF"}, e_2 = \text{"white matter"}, e_3 = \text{"grey matter"}\}$. The distribution of (X, Y) is defined, firstly, by prior distribution $P_X(x)$, supposed to be stationary and Markovian and secondly, by site-wise likelihoods $P_{Y_s/X_s}(y_s/x_s)$ whose parameter Φ depends on class label x_s . Note that, in this application, the shape of these distributions vary with the class (it will be made explicit in the following). Finally, we assume independence between each random variable Y_s given X_s . The observable Y is called the "incomplete data", and Z the "complete data".

Assuming the segmentation result x is known or observable (i.e., we know the "complete data"), the parameters of the grey level statistical distribution associated to each class, also called the distribution mixture parameters, can then be easily computed with the ML estimator of the "complete data".

- First, and in order to take into account the Poisson noise phenomenon inherent to the SPECT imaging process in the "CSF" area, we model $P_{Y_s/X_s}(y_s/e_1)$ by an exponential law [3],

$$\mathcal{E}_Y(y; \alpha) = \frac{1}{\alpha} \exp\left(-\frac{y}{\alpha}\right),$$

with $y > 0$. Let now $Y = (Y_1, \dots, Y_M)$ be M random variables, independent and identically distributed according to a "single" exponential law $\mathcal{E}_Y(y; \alpha)$, and $y = (y_1, \dots, y_M)$ a realization of Y . The ML estimator of α_{ML} for the "complete data" is simply the mean of the sample y [2].

- In order to describe the luminance within the "white matter" and the "grey matter" regions, we model the conditional density function for these regions by a Gaussian law. We have found that this distribution model was particularly successful in our application. Theoretically, this assumption of normality is a reasonable approximation due to the reconstruction physical process used in SPECT imagery in which the grey level of a given pixel (herein considered as a random variable) are sums of many variables and the *central limit theorem* can be applied [2]. The corresponding ML estimator $\hat{\Phi}_{ML} = (\hat{\mu}_{ML}, \hat{\sigma}_{ML}^2)$ of the "complete data", for a sample y distributed according to a normal law, is defined simply by the empirical mean and the empirical variance.

When the segmentation result is unknown (i.e, the class label of each pixel is not supposed to be known), the considered problem is more complex. By assuming that the X_i are independent, this problem can also be viewed as the estimation of the parameters of K-component mixture. In this case, the observed image, or the sample $y = \{y_1, \dots, y_N\}$, is a realization of Y which distribution has for density,

$$P_Y(y) = \sum_{k=1}^K \pi_k P_{Y/X_i}(y/e_k, \Phi_k),$$

where the π_k are the mixing proportion ($0 \leq \pi_k \leq 1$ for all $k = 1, \dots, K$ and $\sum_k \pi_k = 1$). In this case, the mixture component $P_{Y/X_i}(y/e_k, \Phi_k)$ associated to classes e_1, e_2 and e_3 are an exponential law and two different Gaussian laws respectively. In order to obtain a reliable estimation of the parameter, we resort to the ICE algorithm. This procedure, described in detail in [4, 5], is briefly recalled here. This method relies on an estimator $\hat{\Phi}(X, Y)$ with good asymptotic properties (like the ML estimator) for completely observed data case. When X is unobservable, this procedure starts from an initial parameter vector $\Phi^{[0]}$ (not too far from the optimal one) and generates a sequence of parameter vectors leading to the optimal parameters. To this end, $\Phi^{[p+1]}$ at step $(p+1)$ is chosen as the conditional expectation of $\hat{\Phi}$ given $Y = y$, computed according to the current value $\Phi^{[p]}$. It is the best approximation of Φ in terms of the mean squares error [4, 5]. By denoting E_p , the expectation relative to parameter vector $\Phi^{[p]}$, $\Phi^{[p+1]}$ is computed from $\Phi^{[p]}$ and $Y = y$ by $\Phi^{[p+1]} = E_p[\hat{\Phi}(X, Y)|Y = y]$. The computation of this expectation is impossible in practice, but we can approach it thanks to the law of large numbers by,

$$\Phi^{[p+1]} = \frac{1}{n} [\hat{\Phi}(x_{(1)}, y) + \dots + \hat{\Phi}(x_{(n)}, y)],$$

where $x_{(i)}$, $i = 1, \dots, n$ are realizations of X drawn according to the posterior distribution $P_{X/Y}(x/y, \Phi^{[p]})$. In order to decrease the computational load, we can take $n=1$ without altering the quality of the estimation [6]. Finally, we can use the Gibbs sampler algorithm [7] to simulate realizations of X according to the posterior distribution. For the local *a priori* model of the Gibbs sampler, we adopt a three-dimensional isotropic Potts model with a first order neighborhood [8]. In this model, there are three parameters, called "the clique parameters" denoted $\beta_1, \beta_2, \beta_3$ and associated to the horizontal, vertical, and transverse binary cliques respectively. Given this *a priori* model, the prior distribution $P_X(x)$ can be written as,

$$P_X(x) = \exp\left\{-\sum_{\langle s,t \rangle} \beta_{st}(1 - \delta(x_s, x_t))\right\},$$

where summation is taken over all pairs of neighboring sites and $\delta(\cdot)$ is the Kronecker delta function. In order to favor homogeneous regions with no privileged orientation in the Gibbs sampler simulation process, we choose $\beta_{st} = \beta_1 = \beta_2 = \beta_3 = 1$. Finally, the distribution mixture parameter estimation procedure for the "incomplete data" using the ICE procedure is outlined below :

Parameter initialization : we can use the initialization method described in [9] or an initial guess for $\Phi^{[0]}$ (not "too far" from the optimal one). Then $\Phi^{[p+1]}$ is computed from $\Phi^{[p]}$ in the following way :

ECI Procedure			
$\Phi_{(e_1)}^{\text{final}}$	0.52(π)	11(α)	
$\Phi_{(e_2)}^{\text{final}}$	0.26(π)	100(μ)	648(σ^2)
$\Phi_{(e_3)}^{\text{final}}$	0.22(π)	172(μ)	383(σ^2)

TAB. 1: Estimated parameters for the picture reported in Fig. 2. π stands for the proportion of the three classes within the SPECT image. α are the exponential law parameter. μ and σ^2 are the Gaussian law parameter.

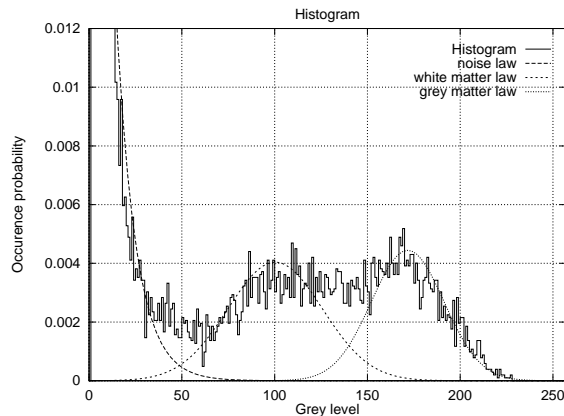


FIG. 1: Image histogram of the picture reported in Fig. 2 (solid curve) and estimated probability density mixture obtained with the ICE procedure (dotted and dashed curves).

1. Stochastic step : using the Gibbs sampler, one realization x is simulated according to the posterior distribution $P_{X/Y}(x/y)$, with parameter vector $\Phi^{[p]}$.
2. Estimation step : the parameter vector $\Phi^{[p+1]}$ is estimated with the ML estimator of the “complete data” corresponding to each class.
3. Repeat until convergence is achieved ; i.e., if $\hat{\Phi}^{[p+1]} \not\approx \hat{\Phi}^{[p]}$, for $k = 1, \dots, K$, we return to step 1.

Fig. 1 represents the estimated distribution mixture of the brain SPECT image shown in Fig. 2. The three site-wise likelihoods $P_{Y_s/X_s}(y_s/e_k)$, $k = 1, 2, 3$, (weighted by the estimated proportion π_k of each class e_k) are superimposed to the image histogram. Corresponding estimates obtained by the ICE procedure, requiring about fifteen iterations, are given in Table 1. We can appreciate the quality of estimation by comparing the image histogram (solid curve) with the probability density mixture corresponding parameters (dotted and dashed curves).

3. 3D MARKOVIAN SEGMENTATION

Based on the estimates given by the ICE procedure, we can compute an unsupervised 3D Markovian segmentation of brain SPECT images. In this framework, the Markovian segmentation can be viewed as a statistical labeling problem according to a global Bayesian formulation in which the posterior distribution $P_{X/Y}(x/y) \propto \exp -U(x, y)$ has to be maximized [8]. The cor-

responding posterior energy is,

$$U(x, y) = \underbrace{\sum_{s \in S} -\ln P_{Y_s/X_s}(y_s/x_s)}_{U_1(x, y)} + \underbrace{\sum_{\langle s, t \rangle} \beta_{st} (1 - \delta(x_s, x_t))}_{U_2(x)}$$

where U_1 expresses the adequacy between observations and labels, and U_2 represents the energy of the *a priori* model. We use the deterministic algorithm ICM [8] to minimize this global energy function. For the initialization of this algorithm, we exploit the segmentation map obtained by a ML segmentation. Fig. 2 displays examples of unsupervised three-class segmentation, exploiting parameters estimated with the ICE procedure. In this segmentation, the “CSF”, the “white matter” and the “grey matter” are represented by a dark, a grey, and a white region respectively, in order to visually express the activity level of the blood flow. Now we are able to identify the different functional regions of a given brain SPECT image, we can turn our attention to the actual classification problem.

4. FEATURE EXTRACTION

The aim of this feature extraction process is to simplify the representation of the brain SPECT image with a relevant parameter vector giving maximal information about the observed brain functionality. The efficiency of the classification technique, used in the next step, depends a lot on the aptness and the discriminating power of these extracted features. In our application, the distribution mixture parameter (i.e., the parameters associated to the statistical distribution law of the white and grey matter regions) turned out to be discriminating for the classification of a given brain SPECT image and to distinguish between healthy and diseased brains (see Section 7 and Fig. 3 in which a selection of feature vector scattering are shown for some parameters of the estimated distribution mixture). For instance, the statistical mean ratio between the white and grey matter classes and the proportion of the white matter seems to be lower in the case of a diseased brain. We will use these interesting characteristics in the following.

We use the three features extracted from the mixture distribution parameters and/or the segmentation map ; namely, 1) the statistical mean ratio between the white and grey matter classes (μ_1/μ_2), the entropy of the grey matter class (E_2) (close to the variance of this class) and finally the proportion of the white matter class (π_1). The proportion of the grey matter class and the entropy or variance of the grey matter class turned out not to be discriminant in our application. These three aforementioned parameters have to be computed for each brain SPECT volume and they make up the feature vector that will be used for the classification step.

5. CLASSIFICATION STEP

The objective of unsupervised pattern classification is to separate the feature vectors (or samples) into groups or classes so that the set of samples of a same class differ from each other as little as possible, according to a chosen criterion. To this end, we can use a statistical approach and suppose that each sample of a given class is a realization of an unknown random process. In that prospect, we consider once again a couple of random fields (\mathbf{X}, W) where $\mathbf{X} = \{\mathbf{X}_i, 1 \leq i \leq M\}$ is the data field associated to the M

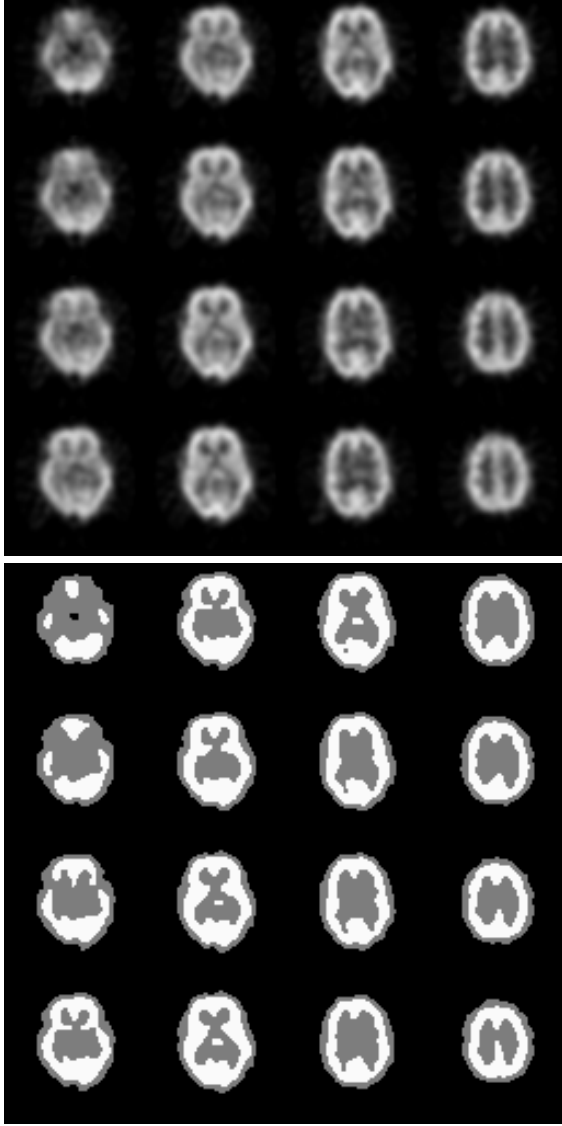


FIG. 2: Example of an unsupervised three-dimensional Markovian segmentation of a brain SPECT volume using the ICM deterministic relaxation technique and based on the parameters estimated by the ICE procedure. Top : real brain SPECT volume (16 central transversal slices). Bottom : three-class Markovian segmentation.

extracted d -dimensional feature vectors. Each of the \mathbf{X}_i takes its value in \mathbb{R}^d and each $W = \{W_i, 1 \leq i \leq M\}$ in the finite set $\{\epsilon_1 = \text{“healthy brain”}, \epsilon_2 = \text{“diseased brain”}\}$. In this classification step, we assume that the distribution of (\mathbf{X}, W) is defined by, firstly, $P_{W_i}(\epsilon_k) = p_k$, with $k = 1, \dots, K$, the proportion of the class ϵ_k and, secondly, by the distribution family $P_{\mathbf{X}/W_i}(\mathbf{x}/\epsilon_k)$. We suppose that the \mathbf{X}_i 's are independent given W and the W_i 's are also independent for $i = 1, \dots, M$. Finally, we assume that $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_M$ is a realization in \mathbb{R}^d , of \mathbf{X} whose density takes the form of the following K -component mixture,

$$P_{\mathbf{X}}(\mathbf{x}) = \sum_{k=1}^K p_k P_{\mathbf{X}/W_i}(\mathbf{x}/\epsilon_k, \Psi_k). \quad (1)$$

In this study, each mixture component $P_{\mathbf{X}/W_i}(\mathbf{x}/\epsilon_k, \Psi_k)$ is assumed to be a multivariate Gaussian distribution with mean vector μ_k and covariance matrix Σ_k ($\Psi_k = (\mu_k, \Sigma_k)$),

$$P_{\mathbf{X}/W_i}(\cdot) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu_k)^t \Sigma_k^{-1}(\mathbf{x} - \mu_k)\right\}.$$

The mixture model is determined by the mixing coefficients p_k , the mean vector μ_k , the covariance matrix Σ_k and the number of terms in the mixture K ($= 2$ in our application). Given the mixture model parameters, we can use the Bayes criteria, also called the MAP criteria, in order to classify the extracted feature vectors [2].

In supervised classification, patterns known to belong to a class are used to train a classifier or to compute, with the ML estimators of the “complete data”, an estimation of the distribution mixture parameters; \hat{p}_k and $\hat{\Psi}_k$ (for all $k = 1, \dots, K$). In unsupervised classification approach, the estimation problem is more complex because the class of each pattern are not supposed to be known. Given initial values for the mixture parameters, Expectation Maximization (EM) algorithm [10] or the Stochastic EM algorithm (SEM) [11] provide an effective and iterative Maximum Likelihood estimates of these parameters. Nevertheless, the initial parameter values have a significant impact on the convergence of these iterative procedures and on the quality of the final estimates. In our application these initial parameters can be given by a K -means clustering procedure. In doing so, we assume, as first approximation, that the considered clusters are spherical with equal volumes. The obtained partitions allow to obtain a rough estimation of the mixture parameters which are then used to initialize the EM or the SEM clustering and estimation procedure.

K-means algorithm

This clustering method [2] assumes the presence of spherical clusters with equal volumes or Gaussian distributions with identical covariance matrix; i.e., $\Sigma_1 = \Sigma_2 = \sigma^2 I$ with σ^2 unknown. In this iterative algorithm, we reassign, at each iteration, every sample or feature vector to the class of the cluster with the nearest mean. Once the $K = 2$ partitions are obtained, we can compute a rough estimates of the proportion \hat{p}_k and the Gaussian mixture parameters $\hat{\Psi}_k$, for $k = 1, 2$, with the empirical ratio, mean and the empirical covariance matrix,

$$\hat{p}_k = \frac{N_k}{M}, \quad (2)$$

$$\hat{\mu}_k = \bar{\mathbf{x}}_k = \frac{1}{N_k} \sum_{\mathbf{x}_i \in C_k} \mathbf{x}_i, \quad (3)$$

$$\hat{\Sigma}_k = E\left[(\mathbf{x}_i - \hat{\mu}_k)(\mathbf{x}_i - \hat{\mu}_k)^t \mid W_i = \epsilon_k\right]. \quad (4)$$

SEM algorithm

The SEM [11] is a recent density mixture estimator which is an improvement of the EM method obtained by the addition of a stochastic component. Using initial value $(\hat{p}_k^{[0]}, \hat{\Psi}_k^{[0]})$, the SEM algorithm can be outlined as follows :

Initialization step : we compute for every sample \mathbf{x}_i , the conditional posterior probability $P_{W_i/\mathbf{X}_i}(\epsilon_k/\mathbf{x}_i)$ (based on parameter vector $(\hat{p}_k^{[0]}, \hat{\Psi}_k^{[0]})$) of its belonging to the class ϵ_k , for $k = 1, \dots, K$ and for $i = 1, \dots, M$.

1. Stochastic step : for each \mathbf{x}_i , we select from the set of classes $\{\epsilon_1, \dots, \epsilon_K\}$ an element according to the distribution

$P_{W_i/\mathbf{X}_i}^{[p]}(\epsilon_1/\mathbf{x}_i), \dots, P_{W_i/\mathbf{X}_i}^{[p]}(\epsilon_K/\mathbf{x}_i)$. This selection defines a partition $C_1^{[p]}, \dots, C_K^{[p]}$ of the sample $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_M$.

2. Maximization step : the SEM algorithm supposes that every \mathbf{x}_i belonging to $C_k^{[p]}$ for $k = 1, \dots, K$ is realized according to the distribution defined by $P_{\mathbf{X}/W_i}(\mathbf{x}/\epsilon_k)$, the density corresponding to the class ϵ_k . By denoting $N_k^{[p]} = \text{card}(C_k^{[p]})$, we can estimate $(\hat{p}_k^{[p]}, \hat{\Psi}_k^{[p]})$, the parameters vector of the distribution mixture with the Maximum Likelihood estimator of the ‘‘complete data’’ (see Equations (2)-(4)).

Estimation step : For each \mathbf{x}_i , we define the next distribution $P_{W_i/\mathbf{X}_i}^{[p+1]}(\epsilon_1/\mathbf{x}_i), \dots, P_{W_i/\mathbf{X}_i}^{[p+1]}(\epsilon_K/\mathbf{x}_i)$ based on the current parameter vector $(\hat{p}_k^{[p]}, \hat{\Psi}_k^{[p]})$.

Repeat until convergence is achieved ; i.e., if $(\hat{p}_k^{[p+1]}, \hat{\Psi}_k^{[p+1]}) \approx (\hat{p}_k^{[p]}, \hat{\Psi}_k^{[p]})$, we return to step 1.

If the convergence is never completely achieved ; for instance, in the case of important overlapping regions between classes, a solution consists in stopping the iterative procedure after P iterations and in choosing the most frequently drawn partition in the $(P - k)$ last iterations. Note also that this algorithm can be viewed as a classical ECI procedure with a multivariate Gaussian mixture and without a *a priori* Markovian assumption on the data.

6. DIMENSIONALITY REDUCTION

Another fundamental problem in clustering procedures is the extraction and the selection of features which are most effective in producing an optimum cluster separability and then an accurate classification result. This also reduces the computation time while improving the quality of the classification by discarding noisy, redundant and less useful features. The dimensionality reduction method that we have investigated in this application is the Fisher’s method [2] and consists in finding and doing optimal projection of the feature space to achieve maximum separability between classes and minimum scattering between samples of a same class. In doing so, we facilitate the classification of patterns into classes. In order to obtain the optimum class separability, the Fisher’s method seeks the linear combination vector \mathbf{u} which minimize the within-class inertia and maximize the between-class inertia. This criterion leads to maximize the following separability criterion,

$$\mathcal{D} = \frac{\mathbf{u}^t \Sigma_b \mathbf{u}}{\mathbf{u}^t \Sigma_w \mathbf{u}},$$

where Σ_w and Σ_b are the within and between class scatter matrix [2]. We solve this problem by using the Lagrange multiplier and by defining the mixture scatter matrix $\Sigma_m = \Sigma_w + \Sigma_b$ (defined as the covariance of all samples regardless of their class assignments). Fisher have shown that \mathbf{u} is one of the eigenvectors of $\Sigma_m^{-1} \Sigma_b$ related to the (nonzero) largest eigenvalue λ . In order to reduce the d -dimensional problem to a r -dimensional problem (with $r < d$), we have to consider the r eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_r$ related to the r largest eigenvalues of the matrix $\Sigma_m^{-1} \Sigma_b$. By denoting A_u , the matrix whose rows are defined by these r eigenvectors, the new feature vectors, defined in the projected feature space, are simply $A_u \mathbf{x}_i$ ($1 \leq i \leq M$). This new feature space contains most information of the initial input data set (and no correlation between the new features) if the r selected eigenvalues represent the major part of the trace of $\Sigma_m^{-1} \Sigma_b$.

True class	Estimated classes	
	Healthy brain	Diseased brain
Healthy brain	68.2%	20.8%
Diseased brain	31.8%	79.2%

TAB. 2: Classification matrix results using the full feature set and the SEM classification procedure.

7. EXPERIMENTAL RESULTS

The effectiveness of the proposed classification scheme is tested on a database of 46 real brain SPECT volumes (containing 24 healthy brains and 22 diseased brains). Each input 3D SPECT image used in this test is composed of 64 transversal slices of 64×64 pixels with 256 grey levels. We recall that the initial estimates of the mixture parameters, used by the classification iterative procedures, are given by the ML estimator of the ‘‘complete data’’ on the K partitions ($K=2$) obtained by the K -means clustering procedure. The class of each cluster is identified by comparing the ‘‘mean ratio between the white and grey matter classes’’ feature mean of each cluster. The cluster corresponding to the ‘‘healthy brain’’ class is the one associated to the highest mean of this feature (cf. Fig. 3a to see the discriminating power of this feature). Besides the feature reduction step is the Fisher’s method and is based on the clusters obtained by a first unsupervised classification step without dimensionality reduction. In our application, we decide to convert the d -dimensional clustering problem (initially, $d = 3$) into a r -dimensional clustering problem allowing to keep 99.9% of the discriminating power of the initial feature vector.

In addition to the deterministic EM procedure, we have implemented and tested the CEM (Classification Expectation Maximization) algorithm [12]. This procedure can be viewed as the deterministic version of the SEM procedure and consists simply in changing the ‘‘stochastic step’’ by a ‘‘classification step’’ in which we select the class associated to the maximum conditional posterior probability obtained at each iteration of the algorithm. We evaluate now the performance of the different classifiers on their ability to correctly classify the brain images from our database, with and without dimensionality reduction procedure.

- Without dimensionality reduction, we obtain 74% rate of good classification. This iterative procedure require less than 20 iterations in order to converge For our application, classification results are similar for the EM or CEM procedure. Table 2 shows the classification matrix using the full feature set and the SEM classification procedure.

- The dimensionality reduction Fisher’s method allows to reduce the initial three-dimensional clustering problem into a two-dimensional clustering problem ($r = 2$). Besides, this dimensionality reduction procedure allows to decrease the necessary number of iterations of the SEM classification procedure. Nevertheless, even if it allows to efficiently reduce the dimensionality and increase the convergence of the clustering problem, it does not improve the classification accuracy. Fig. 4 shows the scattering of the feature vectors in the reduced feature space. The two clusters can easily be discriminated in this new projected feature space.

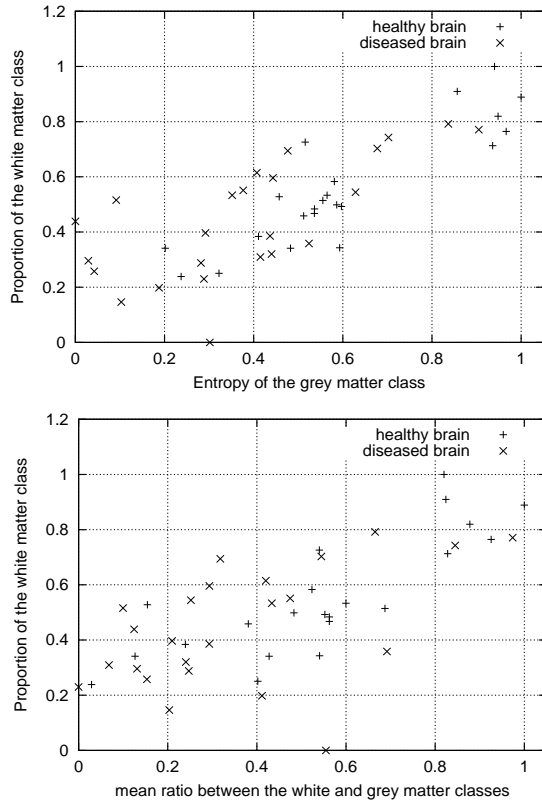


FIG. 3: Scattering of the feature vectors associated to the healthy and diseased brain (ground-truth) for some two-dimensional subsets of the full feature set. Feature subspace : Top : “entropy of the grey matter class” and “mean ratio between the white and grey matter classes”. Bottom : “proportion of the white matter class” and “mean ratio between the white and grey matter classes”.

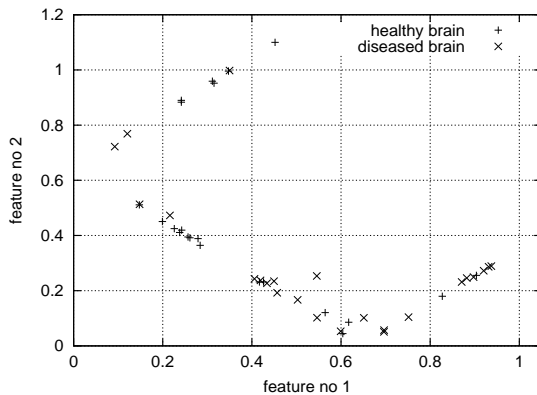


FIG. 4: Scattering of the feature vectors in the new projected feature space (ground-truth). The two clusters can easily be discriminated thanks to the dimensionality reduction procedure.

8. CONCLUSION

In this paper, we have described an unsupervised segmentation method which proved quite adapted for the SPECT image segmentation problem. We have also described an unsupervised classifica-

tion scheme in order to roughly distinguish the healthy brains from the diseased ones. We hope this approach will be useful in a first screening for a disease prior to a more thorough examination by the nuclear physician. We have stated these segmentation and classification tasks within the Bayesian framework. In order to make these methods “unsupervised”, these ones exploit or “cooperate” with a stochastic and Bayesian estimation step which estimates in a Maximum Likelihood sense, the parameters of a Gaussian distribution mixture. The estimation step, involved in the segmentation, takes into account the diversity of the laws in the distribution mixture of a SPECT image and the segmentation Markovian model use a 3D spatial neighborhood whose structure accurately describes the inherent 3D spatial properties of these images. Parameters of the distribution mixture of the SPECT image have proved to be discriminant. Besides, the described method remains sufficiently general to be applied to other medical data classification problems (breast lesions for instance).

9. REFERENCES

- [1] D.C. Costa and P.J. Ell. *Brain Blood Flow in Neurology and Psychiatry*. Series Editor : P.J. Ell, 1991.
- [2] S. Banks. *Signal processing, image processing and pattern recognition*. Prentice Hall, 1990.
- [3] T. S. Curry and J. E. Dowdey and R. C. Murry. *Christensen's Physics of Diagnostic Radiology*. Lea and Febiger, 1990.
- [4] F. Salzenstein and W. Pieczynski. Unsupervised Bayesian segmentation using hidden markovian fields. *In proc. International Conference on Acoustics, Speech, and Signal Processing*, pages 2411–2414, May 1995.
- [5] W. Pieczynski. Champs de Markov cachés et estimation conditionnelle itérative. *Revue Traitement Du Signal*, 11(2) :141–153, 1994.
- [6] B. Braathen, P. Masson, and W. Pieczynski. Global and local methods of unsupervised Bayesian segmentation of images. *GRAPHICS and VISION*, 2(1) :39–52, 1993.
- [7] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, 1984.
- [8] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, B-48 :259–302, 1986.
- [9] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Unsupervised Markovian segmentation of sonar images. *In Proc. International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 2781–2785, Munchen, May 1997.
- [10] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Royal Statistical Society*, pages 1–38, 1976.
- [11] P. Masson and W. Pieczynski. SEM algorithm and unsupervised statistical segmentation of satellite images. *IEEE Trans. on Geoscience and Remote Sensing*, 31(3) :618–633, 1993.
- [12] Gaussian parsimonious clustering models. *Pattern Recognition*, 28(5) :781–793, 1995.