Parameter Estimation of Finite Mixtures Using the EM Algorithm and Information Criteria with Application to Medical Image Processing[1)]

Z. Liang, R.J. Jaszczak and R.E. Coleman
Department of Radiology, Duke University Medical Center
Durham, North Carolina 27710

Abstract

A method for parameter estimation in image classification or segmentation is studied within the statistical frame of finite mixture distributions. The method models an image as a finite mixture. Each mixture component corresponds to an image class. Each image class is characterized by parameters, such as the intensity mean, the standard deviation and the number of image pixels in that class. The method uses a maximum likelihood (ML) approach to estimate the parameters of each class, and employs information criteria of Akaike (AIC) and/or Schwarz and Rissanen (MDL) to determine the number of classes in the image. In computing the ML solution of the mixture, the method adopts the expectation maximization (EM) algorithm. Mathematical formula for the method are presented. The initial estimation and convergence of the ML-EM algorithm are studied. The parameters estimated from a simulated phantom are very close to those of the phantom. The determined number of image classes agrees with that of the phantom. The accuracy in determining the number of image classes using AIC and MDL is compared. The MDL criterion performs better than the AIC criterion. A modifed MDL shows further improvement. The results obtained from experimental and real images are very encouraging.

I. INTRODUCTION

In model-based image segmentation or classification, it is usually required to know the model parameters for the distinct image classes and the number of classes in the observed or reconstructed image [1-7]. The model parameters and the number of image classes can be estimated directly from the reconstructed image. The estimation can be performed either prior to segmenting the image or during the segmentation process. The estimation may be beneficial in reconstructing an image during an iterative projection and backprojection process [8-9]. The estimation directly affects the accuracy of segmentation or classification. Therefore, it is necessary to investigate the estimation problem for image processing.

II. THEORY

The problem of model-based parameter estimation has been widely addressed in the statistical frame of finite mixture distributions [10-11].

Let $\mathbf{Y} = \{Y_{ij}\}$ be the intensities of image pixels on an array $\{(i,j), 1 \le i \le I \text{ and } 1 \le j \le J\}$. Assume that an image consists of K classes and each class k is characterized by a parameter vector θ_k. In terms of image intensity levels, image classes are represented by intensity levels which are distinct from one another. Assume further that the pixel intensities are randomly distributed over the image array. Let $p_k(Y_{ij} \mid \theta_k)$ be the probability distribution of those pixel intensities which are associated with class k. The distribution of pixel intensity Y_{ij} is then expressed, over all the K classes, as a mixture [10-11]

$$g(Y_{ij} \mid W, \Theta) = \sum_{k=1}^{K} w_k \, p_k(Y_{ij} \mid \theta_k) \qquad (1)$$

where $W = \{w_k\}$ represents the weights of the components $\{p_k\}$ in the mixture, $\sum_k w_k = 1$, and $\Theta = \{\theta_k\}$. The weight w_k represents the ratio of the number of pixels in class k and the total number of pixels ($N = I \times J$) in the image. The elements of the parameter vector θ_k may be the intensity mean μ_k for class k, the standard deviation ν_k around the mean, and/or the correlation coefficient $\rho_{(ij)k}$ between image pixels within that class [9]. Assume that all pixels are independent, the likelihood for the image distribution is, given all the weights W and the class parameters Θ,

$$G(\mathbf{Y} \mid W, \Theta) = \prod_{(i,j)}^{(I,J)} g(Y_{ij} \mid W, \Theta) \, . \qquad (2)$$

The model parameters including the weights $\{w_k\}$ and the vectors $\{\theta_k\}$ can be estimated using a maximum likelihood (ML) approach, given the image data $\{Y_{ij}\}$, [10-11].

The number K of image classes can be determined using information criteria [12-14]. Let $L(\mathbf{Y} \mid W, \Theta)$ be the maximal value of likelihood $G(.)$ with respect to the model parameters $\{w_k, \theta_k\}$. The information criteria of Akaike (AIC) and Schwarz and Rissanen (MDL) are then expressed as [12]

$$AIC(K) = -2 \ln[\, L(\mathbf{Y} \mid W, \Theta) \,] + 2M \qquad (3)$$

and [13-14]

$$MDL(K) = -\ln[\, L(\mathbf{Y} \mid W, \Theta) \,] + \omega M \ln(N) \qquad (4)$$

where M is the number of freely adjustable parameters in W and Θ, N is the total number of pixels in the image defined above, and ω is a constant. There are $K - 1$ freely adjustable weights in W since the constraint $\sum_k w_k = 1$ eliminates a freedom. If each parameter vector θ_k has τ elements, there are τK freely adjustable elements in Θ. So the total number

1) This work was supported by Grant #HL44194, awarded by the National Heart, Lung, and Blood Institute, and in part by Grant #CA33541, awarded by the National Cancer Institute and by Grant #DEFG05-89ER60894, awarded by the Department of Energy.

0018-9499/92$03.00 © 1992 IEEE

of freely adjustable parameters is $M = \tau K + (K-1)$. The constant $\omega = 1/2$ was derived by specifying a particular prior [13] and $\omega = 1$ by another [15]. During our experimental studies, we found that $\omega = 5/2$ is more appropriate. The optimal number K_s of mixture components or image classes for the given image data $\mathbf{Y}$ is determined by minimizing the information criterion AIC or MDL [12-14].

Both the AIC and MDL criteria have been widely used in model selection [16]. They measure the Kullback-Leibler mean distance [17] between the estimated mixture and the mixture which generates the observed image, and select that model which minimizes the criterion of (3) or (4). If the ML is identical for two models, they select that one with fewer freely adjustable parameters $\{\theta_k\}$.

III. METHODS

In this section, the mathematical formula for computing the ML solution $\{w_k, \theta_k\}$ are presented, and the procedure for determining the image class number K is described. That particular number K_s which gives the minimum of AIC or MDL is chosen together with the corresponding ML estimate $\{w_k, \theta_k\}$, $k = 1, 2, ..., K_s$, as the model parameters for the observed image $\mathbf{Y}$.

A. Iterative Approach Algorithm

Although there are many numerical techniques can be used to compute the ML solution, the expectation-maximization (EM) algorithm [18] is employed.

In terms of EM terminology, the likelihood $G(.)$ of (2) represents the family of sampling densities for the incomplete data or observed data $\mathbf{Y}$. In image processing $\mathbf{Y} = \{Y_{ij}\}$ is the reconstructed image. Each datum Y_{ij} is associated with an unobserved image class θ_k which is to be estimated. In order to completely specify the association between datum Y_{ij} and class θ_k, an unobserved indicator vector $\mathbf{Z}_{ij}$ of length K is introduced for Y_{ij} [18]. The K components $\{Z_{k(ij)}\}$, $1 \le k \le K$, of the vector $\mathbf{Z}_{ij}$ are, in ideal case, all zero except for one equal to unity indicating that datum Y_{ij} is associated with the unobserved class θ_k. In practice, $Z_{k(ij)}$ may be defined as the conditional probability that datum Y_{ij} belongs to class k, given the data $\mathbf{Y}$ and the parameters $\{W, \Theta\}$. Thus, the complete data are $\{Y_{ij}, Z_{k(ij)}\}$ which completely classify all image pixels into the K classes. The mixture for the complete datum $\{X_{ij} = (Y_{ij}, Z_{k(ij)})\}$ is then [11]

$$f(X_{ij} \mid W, \Theta) = \prod_{k=1}^{K} w_k^{Z_{k(ij)}} \, p_k^{Z_{k(ij)}}(Y_{ij} \mid \theta_k) \tag{5}$$

and the family of sampling densities for the complete data $\mathbf{X} = \{X_{ij}\}$ is

$$F(\mathbf{X} \mid W, \Theta) = \prod_{(i,j)}^{(I,J)} f(X_{ij} \mid W, \Theta) \,. \tag{6}$$

Once the family of sampling densities for the complete data is formulated, an iterative approach algorithm can be derived easily following the E- and M-steps described by Dempster et al [18]. The E-step is expressed, at the n-th iteration, as

$$\begin{aligned} Q(W, \Theta \mid W^{(n)}, \Theta^{(n)}) &= E[\, \ln F(\mathbf{X} \mid W, \Theta) \mid \mathbf{Y}, W^{(n)}, \Theta^{(n)} \,] \\ &= \sum_{ij} \{ \sum_k [\, Z_{k(ij)}^{(n)} \ln(w_k) + Z_{k(ij)}^{(n)} \ln p_k(Y_{ij} \mid \theta_k) \,] \} \end{aligned} \tag{7}$$

where [11,19]

$$Z_{k(ij)}^{(n)} = E[\, Z_{k(ij)} \mid Y_{ij}, W^{(n)}, \Theta^{(n)} \,] = \frac{w_k^{(n)} \, p_k(Y_{ij} \mid \theta_k^{(n)})}{g(Y_{ij} \mid W^{(n)}, \Theta^{(n)})} \tag{8}$$

is understood as the conditional probability of datum Y_{ij} belonging to class k at the n-th iteration and $\sum_k Z_{k(ij)}^{(n)} = 1$ means that the conditional probability distribution over the K classes is normalized. It is clear that the summation over all pixels for each class k, $\sum_{ij} Z_{k(ij)}^{(n)}$, is the estimated number of pixels in that class k at the n-th iteration.

The M-step generates the $(n+1)$-th iterated result which maximizes $Q(W, \Theta \mid W^{(n)}, \Theta^{(n)})$ with respect to W and Θ respectively. For parameter W, function $Q(.)$ is maximized, subject to the constraint $\sum_k w_k = 1$. By introducing a Lagrange multiplier λ, the solution $w_k^{(n+1)}$ is given by

$$\left. \frac{\partial \{ Q(.) - \lambda (\sum_k w_k - 1) \}}{\partial w_k} \right|_{w_k^{(n+1)}} = 0$$

or

$$\lambda = N \quad \text{and} \quad w_k^{(n+1)} = \frac{1}{N} \sum_{ij} Z_{k(ij)}^{(n)} \tag{9}$$

is the estimated ratio of the pixel number in class k and the total pixel number N at the $(n+1)$-th iteration.

In order to compute parameters θ_k, the probability distribution $p_k(Y_{ij} \mid \theta_k)$ of pixel intensities for class k should be specified first. A few such probabilities were mentioned in our previous work [9]. For example, a Gaussian probability $p_k(Y_{ij} \mid \theta_k)$ is given by

$$p_k(Y_{ij} \mid \theta_k) = (2\pi \nu_k)^{-1/2} \, exp[-(Y_{ij} - \mu_k)^2 / 2\nu_k] \tag{10}$$

where $\theta_k = (\mu_k, \nu_k)$, μ_k is the intensity mean of image class k and ν_k is the variance around the mean. The estimate for the parameters at $(n+1)$-th iteration, which maximizes function $Q(.)$, is then given by [11,19-20]

$$\mu_k^{(n+1)} = \frac{\sum_{ij} Z_{k(ij)}^{(n)} \, Y_{ij}}{\sum_{ij} Z_{k(ij)}^{(n)}} \tag{11}$$

i.e., the conditional expectation of the pixel intensities belonging to class k and

$$\nu_k^{(n+1)} = \frac{\sum_{ij} Z_{k(ij)}^{(n)} \, [\, Y_{ij} - \mu_k^{(n+1)} \,]^2}{\sum_{ij} Z_{k(ij)}^{(n)}} \tag{12}$$

i.e., the conditional variance of the pixel intensities belonging to class k at the $(n+1)$-th iteration.

The convergence of the iterative approach or ML-EM algorithm (8)-(12) can be studied by the difference of log

likelihood

$$\psi_1(n) = \ln G(\mathbf{Y} \mid W^{(n+1)}, \Theta^{(n+1)}) - \ln G(\mathbf{Y} \mid W^{(n)}, \Theta^{(n)}) \quad (13)$$

and the difference of successive estimates

$$\psi_2(n) = \sum_k [\, w_k^{(n+1)} - w_k^{(n)} \,]^2 / \sqrt{K-1} \quad \text{for} \quad K > 1 \,. \quad (14)$$

The ML-EM approach for image reconstruction problem has been detailed in [21-22], and the related stopping criteria were studied in [23].

B. Determination of the Number K of Image Classes

The information criteria AIC of (3) and/or MDL of (4) are employed to select that K_s from a number of candidates ($1 \le k \le K_{\max}$) for the assumed mixture $g(.)$ of (1) and the given image $\mathbf{Y}$, where $K_{\max}$ is a prespecified maximum number of classes in the image. The numerical value of K_s can be calculated as follows:

(a) determine the relation between K and M for the assumed probability p_k of (10);

(b) assume an initial K value, which is in a prespecified range $[K_{\min}, K_{\max}]$, where $K_{\min} \le K_s \le K_{\max}$;

(c) maximize the likelihood $G(\mathbf{Y} \mid W, \Theta)$ of (2) with respect to W and Θ for the assumed K via the iterative ML-EM algorithm of (8)-(12);

(d) compute the criterion value of AIC(K) [see (3)], MDL(K) with $\omega = 1/2$, or modified MDL(K) with $\omega = 5/2$ [see (4)];

(e) repeat steps (b)-(d) for $K \pm 1$ values;

(f) if the calculated criterion value for K using AIC or MDL's is less than those for $K \pm 1$, then $K_s = K$ is selected; otherwise, $K \pm 2$ are assumed, and the steps (c)-(f) above are repeated.

In step (a), $M = 3K - 1$ for the Gaussian probability of (10). In step (b), the starting value K is chosen arbitrarily, dependending on one's prior knowledge about the observed image $\{Y_{ij}\}$. If the starting value K is chosen closer to K_s, then less computation time is needed. The starting value K is usually chosen as 3 or 4.

IV. RESULTS

Since the information criteria are defined by the ML value $L(.)$ and the ML solution $\{w_k, \theta_k\}$ is computed iteratively, the convergence properties of the iterative ML-EM algorithm (8)-(12) should be studied first. Figs.1 and 2 show the convergence performances of the algorithm for the images on the top right of Figs.3 and 4, respectively, using functions ψ_1 of (13) and ψ_2 of (14). For both sets of the image data, the algorithm converged after 50 iterations. From 50 to 100 iterations, the curves of ψ_1 and ψ_2 are almost flat. In order to show clearly the rapid convergence during the first 10 iterations, the curves from 50 to 100 iterations are not shown in the figures. The computation time was approximately 10 seconds per iteration for an image size of 256×256.

Theoretically the model parameters $\{w_k, \theta_k\}$ would be uniquely fitted iteratively via the ML-EM algorithm if there are infinite numbers of data available, independent from the initial estimate. For finite number of data, as in practical situations, the initial estimates $\{w_k^{(0)}, \theta_k^{(0)}\}$ are quite important. We studied three choices: (a) $w_k^{(0)}$ and $\theta_k^{(0)}$ were uniform; (b) $w_k^{(0)}$ or $\theta_k^{(0)}$ was uniform; and (c) $w_k^{(0)}$ and $\theta_k^{(0)}$ were non-uniform (as described later). When (a) was chosen, all iterated $\{w_k^{(n)}, \theta_k^{(n)}\}$ were the same as $\{w_k^{(0)}, \theta_k^{(0)}\}$. In other words, the ML fitting was trapped into a local maximum for the finite number of data. When (b) was used, we have found a case in which the fitting was trapped into a local maximum, although in most cases studied in this paper the parameters were correctly fitted. The choice (c) fitted the model parameters correctly in all the cases being studied.

In all the following studies, the iterative algorithm was run for 100 iterations in order to make sure that the ML value of $L(.)$ was closely reached. Non-uniform initial estimates were chosen for the weights and the model parameters of the Gaussian probability (10): $\{\mu_k^{(0)}\}$, $k = 1, 2, ..., K$, were set in the range from minimal to maximal values of Y_{ij} in a constant increment; $\{v_k^{(0)}\}$ were set from 1 to the maximum of Y_{ij} in a constant increment (note that $v_k^{(0)} \neq 0$); and $\{w_k^{(0)}\}$ were set from maximal value of Y_{ij} to 1 in a constant decrement and then were normalized, $\sum_k w_k^{(0)} = 1$. Those image pixels within the circle of radius of 63 units were input into the computer program of the ML-EM fitting algorithm for images with rectangular size of 128×128. So some pixels outside image support region were included, and the total number of pixels was $N = 12492$. For images of 256×256, the chosen circle had radius of 127 units, and the total number of pixels within the circle was 50696.

A. Computer Simulations

Fig.3 shows the computer simulated brain images. The top left is a bit map (256×256) of a slice from its Hoffman 3D brain phantom [24]. It contained three intensity levels or image classes: the means $\{\mu_k\}$ were scaled to 0.0, 1.25, and 5.0, respectively; the variances $\{v_k\}$ were zero for the 3 classes (noise-free); the ratios $\{w_k\}$ of pixel numbers in the classes to the total pixel number (i.e., $N = 50696$) were 0.6461, 0.1523, and 0.2016, respectively. The image on the top right contained Gaussian noise with variances equal to 1.0, 1.5, and 5.0 for the three classes, respectively. The middle curves are the one-pixel-width horizontal profiles through the centers of the images respectively. On the bottom are the histograms of the top images respectively. In computing the histograms, the image intensities were normalized from 0 to 100. The y-axis represents the log frequencies.

Table 1 below lists the values of AIC, MDL($\omega = 1/2$), and modified MDL($\omega = 5/2$) (M-MDL) for the noise-free image in Fig.3. All the three criteria selected the correct class number 3. The iterative ML-EM algorithm fitted the model parameters $\{w_k, \theta_k\}$ accurately for the three classes. It is noted that all the criteria increased rapidly as K went down below the correct class number, and the modified MDL

increased faster than the other two criteria as K went up above the correct number. It is clear that the class number is much less likely to be underdetermined.

Table 1

K	AIC	MDL	M-MDL
1	212184	106101	106144
2	108330	54187	54296
3	90010	45040	45214
4	90016	45057	45295
5	90022	45073	45376
6	90028	45089	45457
7	90035	45106	45539
8	90041	45122	45621
9	90045	45137	45701

The values of AIC, MDL, and M-MDL for the noisy image of Fig.3 are listed in table 2.

Table 2

K	AIC	MDL	M-MDL
1	228272	114145	114188
2	173314	86679	86787
3	172111	86091	86266
4	172059	86078	86317
5	172072	86098	86401
6	172082	86116	86485
7	172078	86127	86561
8	172076	86139	86638
9	172097	86163	86727

The modified MDL selected the correct class number 3 and the other two criteria overestimated the number of classes (i.e., 4).

The fitted model parameters (i.e., the means $\{\mu_k\}$, the variances $\{v_k\}$, and the pixel ratios $\{w_k\}$) using the ML-EM algorithm are given by table 3 when 3 and 4 classes were assumed in the noisy image.

Table 3

	μ_k	v_k	w_k
$K=3$			
	0.01	0.27	0.676
	1.69	1.56	0.161
	5.66	3.78	0.163
$K=4$			
	0.00	0.26	0.667
	1.44	1.22	0.145
	5.33	4.75	0.159
	4.96	2.82	0.029

The modified MDL criterion selected 3 classes for the noisy image. The classes had means of 0.01, 1.69, and 5.66, variances of 0.27, 1.56, 3.78, and pixel numbers of 34270, 8162, and 8263, respectively. The criteria AIC and MDL selected 4 classes for the noisy image in which the class with mean 5.0 was splited into two classes. It is noted that the class number is more important in guiding image segmentation or classification as compared to the model parameters.

B. Experimental Phantom Studies

Fig.4 shows the transaxial images reconstructed from the Hoffman 3D brain phantom using filtered backprojection methods. The top left is a positron emission tomography (PET) image. It was acquired using a CTI Model 911/2 Scanner with radiopharmaceutical of [F-18]-FDG. The top right is a single photon emission computed tomography (SPECT) image. It was acquired using a Triad Three-Headed Scanner with radiopharmaceutical of Tc-99m. The experimental phantom was added small support structures that link the "white" matter to the "grey" matter (Data Spectrum Corp). These small structures were not included in the bit map of Fig.3. These support structures are visible only in the reconstructed images of Fig.4. The middle curves are the one-pixel-width horizontal profiles through the image centers respectively. The bottom curves are the histograms of the top images respectively. The image size was 128×128. The number of pixels used was $N = 12492$. The negative pixel intensities of the SPECT image were truncated to zero, while the PET images' were not. Since the images were usually in the form of 2-byte integers, the pixel intensities were rescaled to the range [0.0, 20.0] before running the computer program of the ML-EM algorithm. It is noted that the Gaussian probability p_k of (10) is an approximation for the rescaled images. If the accurate probability is known for the reconstructed images. that probability should be used.

For the PET image, M-MDL selected 3 classes, MDL chose 6 classes, and AIC selected 7 classes in the image. The ML fitted model parameters for 3 and 4 classes are listed below.

Table 4

	μ_k	v_k	w_k
$K=3$			
	0.51	0.30	0.273
	7.07	11.53	0.395
	14.63	4.76	0.332
$K=4$			
	0.51	0.30	0.273
	5.67	7.33	0.282
	12.35	6.36	0.291
	15.93	2.45	0.154

For the SPECT image, M-MDL again selected 3 classes, MDL chose 5 classes, and AIC did not reach its minimum up to 9 classes (i.e., AIC assumed more than 9 classes in the image). The ML fitted model parameters for 3 and 4 classes are shown in table 5.

For both the PET and SPECT images, AIC and MDL overestimated the number of classes in the images. The two classes on the bottom of tables 4 and 5 should be one class. The modified MDL selected the correct class number 3. It is noted that since the negative pixel intensities of SPECT image were truncated to zero, the mean and variance of its least level are expected to be lower than that of the PET

image (see the tables). Since the brain in the SPECT image occupied a smaller area on the 128×128 array, so the pixels in its least class should be less than that of the PET image (see the tables). The middle level (or class) of the PET image had larger variance v_k as compared to that of the SPECT image. The larger variance may indicate a slower varying of pixel intensities between the least and highest levels. Despite the truncation and rescale of pixel intensities, the modified MDL selected the same class number.

Table 5

	μ_k	v_k	w_k
$K=3$			
	0.03	0.16	0.501
	5.80	7.42	0.245
	12.78	6.32	0.254
$K=4$			
	0.03	0.16	0.502
	4.61	4.29	0.167
	10.94	7.45	0.264
	14.95	2.77	0.068

C. Results from Real Images

Fig.5 shows a magnetic resonance (MR) image (on the top left) and a PET image (on the top right) of a patient brain. The MR image was a slice from a 3D gradient echo acquisition using a 1.5 Tesla GE Signa Scanner. The PET image was acquired using the same scanner mentioned above. Both images had the same size of 256×256. The curves in the middle are the horizontal profiles of one-pixel-width through the image centers respectively. The curves on the bottom are the histograms of the top images respectively. The pixel intensities were again rescaled to [0.0, 20.0] before input to the computer program of the iterative algorithm without truncation. The Gaussian probability of (10) is again assumed for the images.

Table 6

	μ_k	v_k	w_k
$K=5$			
	0.46	0.16	0.460
	4.40	5.01	0.121
	9.64	1.17	0.280
	10.29	0.45	0.110
	13.98	3.94	0.029
$K=6$			
	0.45	0.16	0.455
	2.28	1.16	0.053
	7.47	6.10	0.141
	9.99	0.76	0.299
	9.69	0.84	0.028
	14.40	3.34	0.024

For the MR image, M-MDL selected 6 classes. Both AIC and MDL assumed more than 9 classes (i.e., they did not reach their minimums up to 9 classes). The fitted model parameters for 5 and 6 classes are given in table 6 above. These two sets of model parameters are most likely to be selected. The selected 6 classes correspond to those classes from least to highest: (a) outside region and the skull (bone); (b) the region of cerebral spinal flow (CSF); (c) the marrow of skull; (d) the white mater; (e) the grey matter; and (f) the skin. Some artifacts due to the motion of CSF during scanning appeared near the image center. They were fitted into the skin class. If 5 classes were selected from the image, the three classes of (c)-(e) should be combined into two classes, so one of the three classes should be mis-segmented.

For the PET image, M-MDL selected 4 classes, both MDL and AIC chose 7 classes. The fitted model parameters for 4 and 5 classes, which are most likely to be selected, are listed in table 7. From the table, it is seen that the two classes on the bottom should be one class and so the PET image had 4 classes as M-MDL criterion predicted. The selected 4 classes, from least to highest, are: (a) outside dark region and the white and grey matter (dark region inside the image support area); (b) the thalami; (c) the substantia nigra; and (d) the cortexes. The visible regions outside the image support area are fitted into the class of thalami.

Table 7

	μ_k	v_k	w_k
$K=4$			
	1.98	0.26	0.723
	3.72	1.58	0.139
	7.55	6.19	0.101
	13.53	5.47	0.037
$K=5$			
	1.98	0.26	0.717
	3.53	1.33	0.132
	6.60	3.96	0.089
	11.93	7.32	0.046
	12.96	8.31	0.016

It has been seen that AIC and MDL again overestimated the numbers of classes in the MR and PET images. The numbers chosen by M-MDL were reasonable. Since the outside of the brain in the MR image is less noisy, the pixel intensities there are lower and there are less pixels there, the fitted mean, variance and pixel ratio of the least class in the image should be smaller than that of the PET image. These expectations are shown in the tables.

In summary, the information criterion MDL performed better than AIC. The modified MDL($\omega = 5/2$) showed improvement as compared to MDL. The MDL($\omega = 1$) of Akaike [15] was studied. It performed between MDL($\omega = 5/2$) and MDL($\omega = 1/2$). MDL($\omega = 1$) overestimated the class numbers for the experimental and real images. The iterative ML-EM approach fitted the model parameters effectively.

V. DISCUSSIONS

A method for model-based parameter estimation has been studied using computer simulations, experimental phantoms, and real patient images. The method used the ML

approach to estimate the model parameters and the information criteria to select the optimal number of classes in the images. The initial estimation and convergence of the iterative ML-EM approach (8)-(12) was investigated. The model parameters were effectively fitted by the algorithm. The MDL criterion performed better than AIC for those image data, while the modified MDL demonstrated improvement in selecting the number of image classes. The overestimation with AIC has also been noted in the literature [25-26].

Note that the information criteria and the ML approach provide a possible mean to theoretically determine the number of image classes and the model parameters. The determined results may be used either to directely segment the image (see Appendix) or to simultaneously reconstruct and segment the image via an empirical Bayesian method in which the determined class number and/or fitted model parameters are treated as the *a priori* image information [8-9].

Although the weights in the mixture were assumed independent from nearby pixels, a Markov property can be included into the weights [27-28]. Although only the mixture distributions were studied, other forms of the likelihood of (2) can be investigated in a similar manner [4,7].

The ML approach with EM algorithm discussed in this paper can be extended to incorporate additional constraints into the approach so that a penalized ML and/or Bayesian frame work can be established.

VI. ACKNOWLEDGMENT

The authors would appreciate Drs. T. Turkington, J. MacFall and J. Bowsher for their assistance in acquiring the images and in preparing the manuscript.

VII. APPENDIX

In this section, the image segmentation given the estimated class parameters is outlined. The details are described in [29-30].

As defined before, the probability distribution of the intensities of those pixels associated with class k is $p_k(Y_{ij} \mid \theta_k)$. If those pixels can be assumed as uncorrelated (as in high spatial resolution imaging modalities, where the resolution is the same as the pixel size), then the simplest segmentation is to assign the label of class k to the pixel (i,j) such that p_k is maximal among all the K_s classes [29]. If the correlation among those pixels is not negligible, and it is known and can be formulated approximately into kernels, then a restoration should be performed on the image before estimating the class parameters and segmenting the image. If the correlation is unknown, but it may be modeled by some *a priori* knowledge about the image, then a prior $p_k(\theta_k)$ can be imposed upon the likelihood $p_k(Y_{ij} \mid \theta_k)$, and the segmentation assigns the label of class k to pixel (i,j) such that $p_k(Y_{ij} \mid \theta_k)\, p_k(\theta_k)$ is maximal among the K_s classes [30].

Although the image intensities within class k can vary significantly, the pixel labels should be the same, i.e., the label of class k. The label distribution or the segment of the image is expected to be piecewise constant within different regions of an image, and so the Markov random field model is applicable

$$p_k(\theta_k) = 1/C_k \; exp\,[-U(\theta_k)/\beta] \qquad (15)$$

where C_k is the normalization constant and β the free parameter for the "potential function" $U(\theta_k)$ [6]. Among the neighborhood of pixel (i,j), $U(\theta_k)$ is chosen to be proportional to the number of the neighboring pixels belonging to class k [30].

VIII. REFERENCES

[1]: H. Derin and H. Elliott, "Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields." IEEE Trans. PAMI, vol.9, pp.39-55, 1987

[2] T. Lei and W. Sewchand, "A New Stochastic Model-Based Image Segmentation Technique for X-Ray CT Image." SPIE, VCIP III, Boston, pp.252-260, 1988

[3] S. Lakshmanan and H. Derin, "Simultaneous Parameter Estimation and Segmentation of Gibbs Random Field Using Simulated Annealing." IEEE Trans. PAMI, vol.11, pp.799-813, 1989

[4] H. Choi, D. Haynor and Y. Kim, "Multivariate Tissue Classification of MRI Images for 3D Volume Reconstruction - A statistical approach." SPIE, Med. Imaging III, vol.1092, pp.183-193, 1989

[5] J. Modestino and J. Zhang, "A Model-Fitting Approach to Cluster Validation with Application to Stochastic Model-Based Image Segmentation." SPIE, VCIP IV, vol.1199, pp.1381-1392, 1989

[6] D. Geman, S. Geman, C. Graffigne and P. Dong, "Boundary Detection by Constraned Optimization." IEEE Trans. PAMI, vol.12, pp.609-628, 1990

[7] Z. Wu and R. Leahy, "Tissue Classification in MR Images Using Hierarchical Segmentation." Proc. IEEE Med. Imaging, vol.2, pp.1410-1414, 1990

[8] Z. Liang, R. Jaszczak, E. Coleman and V. Johnson, "Simultaneous Reconstruction, Segmentation, and Edge Enhancement of Relatively Piecewise Continuous Images with Intensity-Level Information." Med. Physics, vol.18, pp.394-401, 1991

[9] Z. Liang, R. Jaszczak and E. Coleman, "On Reconstruction and Segmentation of Piecewise Continuous Images." in *Information Processing in Medical Imaging*, vol.12, pp.94-104, 1991

[10] B. Everitt and D. Hand, *Finite Mixture Distribution.* Chapman and Hall, New York, 1981

[11] D. Titterington, A. Smith and U. Makov, *Statistical Analysis of Finite Mixture Distributions.* John Wiley & Sons, New York, 1985

[12] H. Akaike, "A New Look at the Statistical Model Identification." IEEE Trans. Auto. Control, vol.19, pp.716-723, 1974

[13] G. Schwarz, "Estimating the Dimension of a Model." Ann. Statistics, vol.6, pp.461-464, 1978

[14] J. Rissanen, "Modeling by Shortest Data Description." Automatica, vol.14, pp.465-471, 1978

[15] H. Akaike, "A Bayesian Analysis of the Minimum AIC Procedure." Ann. Inst. Statist. Math., vol.30, pp.9-14, 1978

[16] M. Stone, "Comments on Model Selection Criteria of Akaike and Schwarz." JRSS, vol.41, pp.276-278, 1979.

[17] S. Kullback, *Information Theory and Statistics*. John Wiley & Sons, New York, 1959.

[18] A. Dempster, N. Laird and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm." JRSS, vol.38, 1-38, 1977

[19] R. Redner and H. Walker, "Mixture Densities, Maximum Likelihood and the EM algorithm." SIAM Review, vol.26, pp.195-239, 1984

[20] T. Lei and W. Sewchand, "Image Segmentation by Using Finite Normal Mixture Model and EM algorithm." IEEE Intern. Conf. on Image Processing, Singapore, 1989

[21] L. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," IEEE Trans. Med. Imaging, vol.1, pp.113-122, 1982

[22] K. Lange and R. Carson, "EM Reconstruction Algorithms for Emission and Transmission Tomography," JCAT, vol.8, pp.306-316, 1984

[23] E. Veklerov and J. Llacer, "Stopping Rule for the MLE Algorithm Based on Statistical Hypothesis Testing," IEEE Trans. Med. Imaging, vol.6, pp.313-319, 1987

[24] E. Hoffman, P. Cutler, W. Digby and J. Mazziotta, "3D Phantom to Simulate Cerebral Blood Flow and Metabolic Images for PET." IEEE Trans. Nucl. Sci., vol.37, pp.616-620, 1990

[25] M. Wax and T. Kailath, "Detection of Signals by Information Theoretic Criteria." IEEE Trans. ASSP, vol.33, pp.387-392, 1985

[26] K. Wong, Q. Zhang, J. Reilly and P. Yip, "On Information Theoretic Criteria for Determining the Number of Signals in High Resolution Array Processing." IEEE Trans. ASSP, vol.38, pp.1959-1971, 1990

[27] L. Baum and J. Eagon, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains." Ann. Math. Statist., vol.41, pp.164-171, 1970

[28] G. Lindgren, "Markov Regime Models for Mixted Distributions and Switching Regressions." Scand. J. Statist., vol.5, pp.81-91, 1978

[29] Z. Liang and J. MacFall, "Multivariate Tissue Classification of NMR Images with Maximum Likelihood and Information Criteria," submitted to Proc. RSNA 1992

[30] Z. Liang and J. MacFall, "Parameter Estimation and NMR Image Segmentation," submitted to IEEE Proc. Med. Imaging 1992

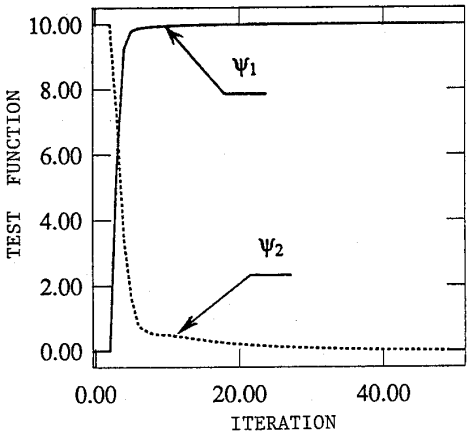

Fig.1: The convergence performance of the ML-EM algorithm for the simulated noisy image in Fig.3.

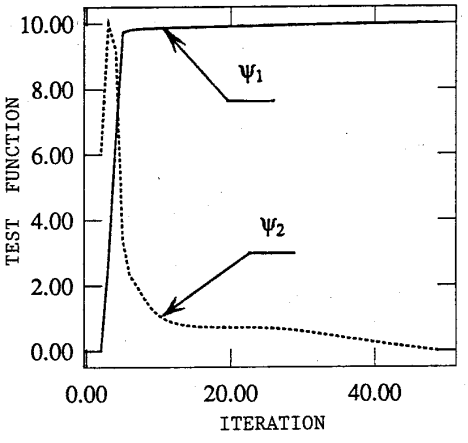

Fig.2: The convergence performance of the ML-EM algorithm for the SPECT phantom image in Fig.4.

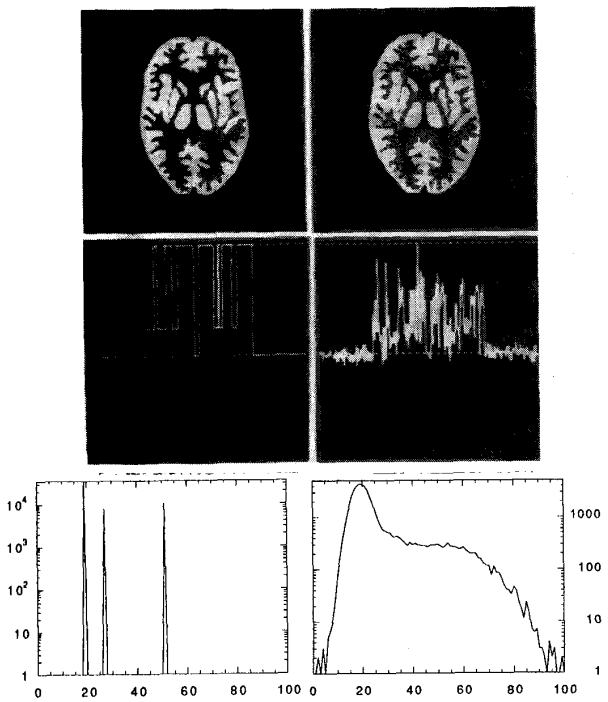

Fig.3: The computer simulated brain phantom images. Top left is the bit map of a slice from Hoffman 3D brain phantom. Top right is the image containing Gaussian noise. The curves in the middle are the horizontal one-pixel-width profiles through the image centers respectively. The curves on the bottom are the histograms of the top images respectively.

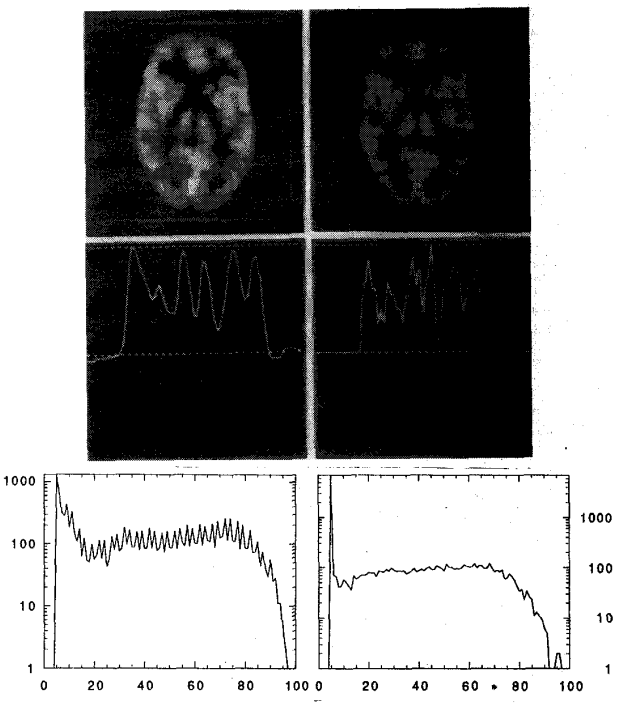

Fig.4: The experimental transaxial images from the Hoffman 3D brain phantom. Top left is a PET image. Top right is a SPECT image. Middle is the horizontal one-pixel-width profiles through the image centers. Bottom is the histograms of the images.

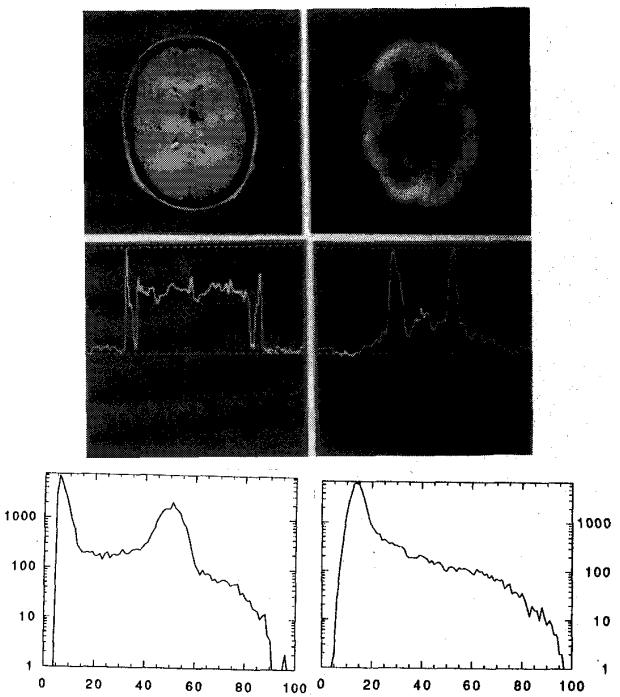

Fig.5: The patient brain images. Top left is a MR image and top right is a PET image. Middle is the one-pixel-width profiles through the image centers horizontally. Bottom is the histograms of the top images.