

Quantification and Segmentation of Brain Tissues from MR Images: A Probabilistic Neural Network Approach

Yue Wang, *Member, IEEE*, Tülay Adalı, *Member, IEEE*, Sun-Yuan Kung, *Fellow, IEEE*, and Zsolt Szabo

Abstract—This paper presents a probabilistic neural network based technique for unsupervised quantification and segmentation of brain tissues from magnetic resonance images. It is shown that this problem can be solved by distribution learning and relaxation labeling, resulting in an efficient method that may be particularly useful in quantifying and segmenting abnormal brain tissues where the number of tissue types is unknown and the distributions of tissue types heavily overlap. The new technique uses suitable statistical models for both the pixel and context images and formulates the problem in terms of model-histogram fitting and global consistency labeling. The quantification is achieved by probabilistic self-organizing mixtures and the segmentation by a probabilistic constraint relaxation network. The experimental results show the efficient and robust performance of the new algorithm and that it outperforms the conventional classification based approaches.

Index Terms—Finite mixture models, image segmentation, information theoretic criteria, model estimation, probabilistic neural networks, relaxation algorithm.

I. INTRODUCTION

QUANTITATIVE analysis of brain tissues refers to the problem of estimating tissue quantities from a given image, and segmentation of the image into contiguous regions of interest to describe the anatomical structures. The problem has recently received much attention largely due to the improved fidelity and resolution of medical imaging systems. Because of its ability to deliver high resolution and contrast, magnetic resonance (MR) imaging has been the dominant modality for research on this problem [1]–[5]. In clinical practice, MR images are typically analyzed by qualitative, or semi-quantitative visualization and evaluation. The main focus

of most automatic MR image analysis schemes has been on image segmentation [2], [4], [6]–[9]. Tissue quantification, on the other hand, alone or together with tissue segmentation, also provides valuable information for brain tissue analysis [1], [5]. Pathological studies show that many neurological diseases are accompanied by subtle changes in brain tissue quantities and volumes [3]. Because of the practical difficulty for clinicians to identify all pathological changes directly from medical images, development of accurate and efficient image analysis systems is of great importance.

Stochastic model-based methods have been by far the most popular approach for quantification of brain tissues from MR images [1], [3], [5], [10]–[12]. The stochastic model-based approach typically employs a finite mixture model that is shown to be a very suitable model for the task [1], [3], [10]. Neural networks have also been employed for image segmentation [2], [6], [7], [13], and recently, a cross fertilization of the stochastic model based and neural network approaches, probabilistic neural networks have emerged as a powerful tool in MR image analysis [7], [14]. Probabilistic neural networks provide valuable insight for designing and learning in neural networks and offer efficient online computation of the quantities of interest, a feature especially important for evaluation of studies in a clinical setting, such as MR image sequence analysis [12]. Furthermore, probabilistic neural networks are particularly suitable for application to quantitative analysis of MR images. In this paper, we present a probabilistic neural network approach for efficient analysis of brain tissues by using single-valued MR brain scans. The major differences of our work from the previous research described in [1], [3], [5], [9] are as follows.

- 1) We present two theorems to show that the correct use of the standard finite normal mixture (SFNM) model in MR brain tissue quantification does not require the pixel images to be statistically independent.
- 2) We introduce and briefly describe a new information theoretic criterion formulation following Jaynes' principle: the minimum conditional bias and variance (MCBV) criterion. We use three information theoretic criteria to determine the appropriate number of tissue types in a particular MR brain scan.
- 3) We introduce an on-line algorithm for parameter estimation associated with tissue quantification: the probabilistic self-organizing mixtures (PSOM) algorithm.

Manuscript received December 13, 1996; revised November 28, 1997. This work was supported in part by the Office of Naval Research under Grant N00014-94-1-0743 and by the U.S. Army Medical Research and Materiel Command under DAMD 18-97-1-2078 DAR.

Y. Wang is with the Department of Electrical Engineering and Computer Science, The Catholic University of America, Washington, DC 20064 USA, and is affiliated with the Department of Radiology, Georgetown University School of Medicine, Washington, DC 20007 USA (e-mail: wang@pluto.ee.cua.edu).

T. Adalı is with the Department of Computer Science and Electrical Engineering, University of Maryland-Baltimore County, Baltimore, MD 21250 USA (e-mail: adali@engr.umbc.edu).

S.-Y. Kung is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: kung@ee.princeton.edu).

Z. Szabo is with the Department of Radiology, Johns Hopkins Medical Institutions, Baltimore, MD 21205 USA (e-mail: zszabo@welchlink.welch.jhu.edu).

Publisher Item Identifier S 1057-7149(98)05319-6.

We present comparative results to show its superior performance in terms of faster rate of convergence and lower floor of estimation error and introduce global relative entropy (GRE) as the objective function for error measure.

- 4) We introduce an efficient procedure for pixel classification associated with tissue segmentation that is realized by a probabilistic constraint relaxation network (PCRN). PCRN might be considered as an extension of the inhomogeneous Markov random field based approaches [15].

Experimental results demonstrate the efficient and reliable performance of the proposed scheme, in terms of the quantification achieved by PSOM, consistent order determination using three information criteria including MCBV, and the satisfactory segmentation results by PCRN.

The paper is organized as follows. In Section II, we present the stochastic modeling formulation, for both the tissue quantification and the segmentation stages. We present the algorithms to solve these problems in Section III along with results using simulated data. Section IV presents examples, with real MR data. These results demonstrate the accuracy and reproducibility of the method and show the performance of efficient automatic quantification and segmentation. We present discussion of the results in Section V.

II. PROBLEM STATEMENT

In this section, we present the problem formulation and the probabilistic network models used for tissue quantification and segmentation.

A. Stochastic Modeling

In order to validate the use of a suitable stochastic model for MR image analysis with a specified objective, we have studied MR imaging statistics and observed several useful statistical properties of MR images [16], [19]. These results are strongly supported by the analysis of actual MR image data [20]. In particular, based on the statistical properties of MR pixel images, where pixel image is defined as the observed gray level associated with the pixel, use of an SFNM distribution is justified to model the image histogram, and it is shown that the SFNM model converges to the true distribution when the pixel images are asymptotically independent [21]. Furthermore, by incorporating statistical properties of context images, where context image is defined as the membership of the pixel associated with different tissue types, a localized SFNM formulation is proposed to impose local consistency constraints on context images in terms of a stochastic regularization scheme [16].

Assume that each pixel in the MR image can be decomposed into pixel image x and context image l . By ignoring information regarding the spatial ordering of pixels, we can treat context images (i.e., pixel labels) as random variables and describe them using a multinomial distribution with unknown parameters $\pi_k, k = 1, \dots, K$. Since it reflects the distribution of the total number of pixels in each tissue type (or component), π_k can be interpreted as a prior probability of

pixel labels determined by the global context information. Thus, the relevant (sufficient) statistics are the pixel image statistics for each component and the number of pixels of each component. The marginal probability measure for any pixel image, i.e., the SFNM distribution, can be obtained by writing the joint probability density of x and l and then summing the joint density over all possible outcomes of l , i.e., $f_{\mathbf{r}}(x) = \sum_l f(x, l)$, resulting in a sum of the following general form according to the Bayes law:

$$f_{\mathbf{r}}(x) = \sum_{k=1}^K \pi_k g(x | \mu_k, \sigma_k^2) \quad (1)$$

with $\sum_{k=1}^K \pi_k = 1, \pi_k \geq 0$, and

$$g(x | \mu_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x - \mu_k)^2}{2\sigma_k^2}\right)$$

where μ_k and σ_k^2 are the mean and variance of the k th Gaussian kernel. We use K to denote the number of Gaussian components and $\mathbf{r} \in \mathcal{R}^{3K-1}$ to denote the total parameter vector that includes μ_k, σ_k^2 , and π_k for all K components. Several observations are worth reiterating.

- 1) All pixel images are identically distributed from a maximum-entropy mixture distribution and treated as *unclassified* data [22]–[24].
- 2) The SFNM model uses the prior probabilities of pixel label in the formulation instead of realizing its true value for each pixel image.
- 3) Since the calculation of the histogram of pixel images relies on the same mechanism as SFNM modeling, it can be considered to be a sampled version of the true pixel distribution [1].

Since the structure of the likelihood function in SFNM model follows an identical distribution [25], the corresponding ML estimation will be unbiased [26]. However, the price to be paid for the stationary structure is that we cannot represent local context explicitly, i.e., the pixel labels are hidden. Because context information is of particular importance in tissue segmentation, by assuming that the context images are random variables with Markovian property [15], a localized SFNM model is formulated. It explicitly incorporates local context regularities into a consistent network structure. For each pixel i , we define the spatial constraint as a local set of all pairs (l_i, l_j) such that the consistency between l_i and l_j can be represented by the indicator function $I(l_i, l_j)$ [2], [13], [27]. Under this configuration, all pairs of labels are either compatible (produce an output “1”) or incompatible (produce an output “0”) [28]. We define the neighborhood of pixel i , denoted by ∂i , by opening a $b \times b$ window with pixel i being the central pixel, where b is assumed to be an odd integer. Similar to the approach taken in [2], [4], and [13], we compute the frequency of neighbors of pixel i with labels compatible to a given label k , conditioning the labels of its neighbors $l_{\partial i} \in \mathcal{R}^{b^2-1}$ by

$$\pi_{ik} = p(l_i = k | l_{\partial i}) = \frac{1}{b^2 - 1} \sum_{j \in \partial i} I(k, l_j | \partial i) \quad (2)$$

and the localized SFNM distribution for x_i directly follows by

$$q(x_i|I_{\partial_i}) = \sum_{k=1}^K \pi_{ik} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right). \quad (3)$$

The calculation of π_{ik} is same with that of π_k , however, its scale is local and thus can be interpreted as the *conditional* prior of the pixel label determined by the uncertainty contained in I_{∂_i} . The localized SFNM model hence provides a more evident meaning than the SFNM model for tissue segmentation [15], while the SFNM model has a better structure for tissue quantification [23].

B. Tissue Quantification

Tissue quantification addresses the combined estimation of tissue parameters (π_k, μ_k, σ_k^2) and the detection of the tissue structural parameter K in (1) given the pixel images \mathbf{x} . The two main approaches used to determine these parameters are classification-based estimation and distance minimization approaches [25], [29]. In the classification-based approach, all pixels are first classified into different components according to a specified distance measure, and then, the model parameters are estimated using sample averages by the ergodic theorems [6], [9], [29]. In the distance minimization approach, the mixture density is fitted to the histogram of pixel images by finding the optimal parameters with respect to a distance measure [1], [3], [21]. We use relative entropy (the Kullback–Leibler distance) [26] for tissue quantification in MR images such that it measures the information theoretic distance between the histogram of the pixel images, denoted by $f_{\mathbf{x}}$, and the estimated SFNM distribution $f_{\mathbf{r}}(x)$, and is given by [26]

$$D(f_{\mathbf{x}}||f_{\mathbf{r}}) = \sum_{x \in \mathcal{X}} f_{\mathbf{x}}(x) \log \frac{f_{\mathbf{x}}(x)}{f_{\mathbf{r}}(x)}. \quad (4)$$

Note that the use of the relative entropy cost also overcomes problems such as convergence at the wrong extreme faced by the squared error cost function as it weighs errors more heavily when probabilities are near zero and one, and diverges in the case of convergence at the wrong extreme [17], [30]. We have shown that, when relative entropy is used as the distance measure, distance minimization is equivalent to maximum likelihood (ML) estimation of SFNM parameters. The conclusion is summarized by the following theorem [38].

Theorem 1: Consider a sequence of random variables x_1, \dots, x_N in \mathcal{R}^N . Assume that the sequence $\{x_i\}$ is independent and identically distributed (i.i.d.) by the distribution $f_{\mathbf{r}}$. Then, the joint likelihood function $\mathcal{L}(\mathbf{r})$ is determined only by the histogram of data and is given by

$$\mathcal{L}(\mathbf{r}) = \exp(-N[H(f_{\mathbf{x}}) + D(f_{\mathbf{x}}||f_{\mathbf{r}})]) \quad (5)$$

where H denotes the entropy with base e [26]. Hence, maximization of joint likelihood function $\mathcal{L}(\mathbf{r})$ is equivalent to the minimization of relative entropy $D(f_{\mathbf{x}}||f_{\mathbf{r}})$.

Thus, tissue quantification is formulated as a distribution learning problem and quantification is achieved when the relative entropy (4) is minimized, or by Theorem 1, when the joint likelihood function $\mathcal{L}(\mathbf{r})$ is maximized. However,

spatial statistical dependence among pixel images is one of the fundamental issues in problem formulation since the calculation of the image histogram treats all pixel images as independent random variables [1], [5]. In order to validate the correct use of (4) in tissue quantification, we prove the following theorem in [38] to show that the image histogram $f_{\mathbf{x}}$ converges to the true distribution f^* with probability one as $N \rightarrow \infty$.

Theorem 2: Consider a sequence of random variables x_1, \dots, x_N in \mathcal{R}^N . Assume that the sequence $\{x_i\}$ is asymptotically independent [26] and identically distributed by the SFNM distribution f^* . For a closed convex set $E \subset \mathcal{F}$ and distribution $f_{\mathbf{x}} \notin E$, let $f_{\mathbf{r}} \in E$ be the distribution that achieves the minimum distance to $f_{\mathbf{x}}$, i.e.,

$$f_{\mathbf{r}} = \arg \min_{f_{\mathbf{r}} \in E} D(f_{\mathbf{x}}||f_{\mathbf{r}}). \quad (6)$$

Then, when N approaches infinity, we have

$$\lim_{N \rightarrow \infty} D(f_{\mathbf{r}}||f^*) = 0 \quad (7)$$

with probability one, i.e., the estimated distribution of \mathbf{x} , $f_{\mathbf{r}}$, given that it achieves the minimum of $D(f_{\mathbf{x}}||f_{\mathbf{r}})$, is close to f^* for large N .

Thus, when N is sufficiently large, minimization of the relative entropy between $f_{\mathbf{r}}$ and f^* can be well approximated by the minimization of the relative entropy between $f_{\mathbf{r}}$ and $f_{\mathbf{x}}$. This fitting procedure can be practically implemented by maximizing the joint likelihood function under the independence approximation of pixel images [18].

C. Tissue Segmentation

Anatomical structure, in addition to the results of tissue quantification that reveals different tissue properties, provides very valuable information in medical applications. Tissue segmentation is a technique for partitioning the image into meaningful regions corresponding to different objects. It may be considered as a clustering process where the pixels are classified into attributed tissue types according to their gray-level values and spatial correlation [6]. A reasonable assumption that can be made is that spatially close pixels are likely to belong to the same tissue type [22]. Accordingly, tissue segmentation addresses the realization of context images $l_i, i = 1, \dots, N$, given the observed pixel images \mathbf{x} . Based on the localized SFNM model (3), a deterministic relaxation labeling can be used to update the context images after global tissue quantification. With a motivation similar to the one in [2] and [6], the technique seeks for a consistent labeling solution where the criterion is to maximize global consistency measure by using a system of inequalities. The structure of relaxation labeling is motivated by two basic considerations: i) decomposition of a global computation scheme into a network performing simple local computations, and ii) use of suitable local context regularities in resolving ambiguities.

We can define the consistency of discrete relaxation labeling and formalize its relationship to global optimization as follows: We first define the component in the localized SFNM distribution (3) as a support function consisting of the compatibility

function $\Lambda(l_i, l_{\partial i})$ and local likelihood $p(x_i|l_i)$:

$$\begin{aligned} S_i(k) &= \Lambda(l_i, l_{\partial i})p(x_i|l_i) \\ &= \pi_{ik} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right). \end{aligned} \quad (8)$$

Note that the support function $S_i(k)$ is a function of the component (tissue type) k . Then, tissue segmentation is interpreted as the satisfaction of a system of inequalities as follows:

$$S_i(l_i) \geq S_i(k) \quad (9)$$

for all k and for $i = 1, \dots, N$, where a consistent labeling is defined as the one having maximum support at each pixel *simultaneously*. We further define the average local consistency measure

$$A(\mathbf{l}) = \sum_{i=1}^N \sum_k I(l_i, k) S_i(k) \quad (10)$$

to link consistent labeling to global optimization [28]. It is shown that when the spatial compatibility measure is symmetric and $A(\mathbf{l})$ attains a local maximum at \mathbf{l} , then \mathbf{l} is a consistent labeling [2], [8], [13], [28]. Hence, a consistent labeling can be accomplished by locally maximizing $A(\mathbf{l})$.

We can view consistency as a "locking-in" property, i.e., since the support function defined for a given pixel depends on the current labels of neighboring pixels, this neighborhood influences the update of the given pixel through probabilistic compatibility constraints. With constraint propagation, the relaxation process iteratively updates the label assignments to increase the consistency, and ideally finds a more consistent labeling with the neighboring labels, such that each pixel is designated a unique label [2], [16].

III. THEORY AND ALGORITHMS

Over the years, several unsupervised approaches have been reported in the literature exploring quantitative analysis of MR brain images [1], [5]. Currently, there are two main approaches to the problem. In the first one, the maximum likelihood quantification scheme, tissue types are first quantified using maximum likelihood principle, where only *soft* classification of the pixel images is required [1]. Further classification of a sample is then performed by placing it into the class for which the posterior probability or the support function is maximum, i.e., by Bayesian consistent labeling [45], [46]. The quantities obtained by sample averages after imperfect pixel classification may not be consistent with the previous quantification result [23]. In the second approach, tissue quantification and segmentation are performed simultaneously with back and forth iterations between the two. In this case, the prior and post quantification results will be consistent, however the quantification and classification errors do interfere with each other during the iterations. In this research, we deal with tissue quantification and segmentation as two separate objectives and use different optimality criteria. However, it is worth reiterating the fact that the proposed method achieves an unbiased ML tissue quantification, a step to be considered independent from the following tissue segmentation step. In

what follows, we present the theory and algorithms for the two stages: i) quantification that involves network order selection and adaptive computation of the parameters to achieve soft classification, and ii) segmentation that uses the order and the parameters computed in the quantification stage to perform hard classification by incorporating local context constraints.

A. Adaptive Model Selection

Since the prior knowledge of the true structure of a real image is generally not available, it is most often desirable to have a neural network structure that is adaptive, in the sense that the number of local components (i.e., hidden nodes) is not fixed beforehand. Both for PSOM and PCRN, using a smaller or larger number of mixture components (local components in the network) than the number of tissue types actually represented on a particular slice will result in incorrect identification and quantification of the tissues in a particular slice. This situation is particularly critical in a real clinical application where the structure of the individual slice for a particular patient may be arbitrarily complex. The objective of adaptive model selection is to propose a systematic strategy for the determination of the structure of the network, i.e., the number of hidden nodes (or mixture components) K in the two probabilistic neural networks: the PSOM and the PCRN. One approach to determine the optimal number K_0 is to use information theoretic criteria, such as the Akaike information criterion (AIC) [31], [32], and the minimum description length (MDL) [5], [33]. The major thrust of this approach has been the formulation of a structural learning in which a model fitting procedure is utilized to select a model from several competing candidates such that the selected model *best fits the observed data*.

For example, AIC will select the model that gives the minimum of

$$\text{AIC}(K_a) = -2 \log(\mathcal{L}(\hat{\mathbf{r}}_{\text{ML}})) + 2K_a \quad (11)$$

where $\mathcal{L}(\hat{\mathbf{r}}_{\text{ML}})$ is the likelihood of $\hat{\mathbf{r}}_{\text{ML}}$, the ML parameter estimates, and K_a is the number of free adjustable parameters in the model. The AIC tries to formulate the problem explicitly as a problem of *approximation* of the true structure by the model. It implies that the correct number of distinctive image regions K_0 can be obtained by minimizing AIC. From a quite different point of view, MDL reformulates the problem explicitly as an information coding problem in which the best model fit is measured such that high probabilities are assigned to the observed data while at the same time the model itself is not too complex to describe [33]. The model is selected by minimizing the total description length defined by

$$\text{MDL}(K_a) = -\log(\mathcal{L}(\hat{\mathbf{r}}_{\text{ML}})) + 0.5K_a \log N. \quad (12)$$

Note that, different from AIC, the second term in MDL takes into account the number of observations. However, the justifications for the optimality of these two criteria with respect to tissue quantification or classification are somewhat indirect and remain unresolved [5], [21], [25], [31].

In this section, we present a new formulation of information theoretic criterion, the *minimum conditional bias and variance criterion*, to address the model selection problem. Akaike and

Rissanen's work on information theoretic criteria have certainly been the inspirational source to this work, however, in our work, we present a new interpretation and justification by information theoretic means [18]. Our approach has a simple optimal appeal in that it selects a minimum conditional bias and variance model, i.e., if two models are about equally likely, MCBV selects the one whose parameters can be estimated with the smallest variance.

New formulation is based on the fundamental argument that the value of the structural parameter can not be arbitrary or infinite, because such an estimate might be said to have low "bias" but the price to be paid is high "variance" [34]. We can obtain a formulation by using Jaynes' principle, which states that "the parameters in a model which determine the value of the maximum entropy should be assigned values which minimize the maximum entropy" [35]. Let joint entropy of \mathbf{x} and $\hat{\mathbf{r}}$ be $H(\mathbf{x}, \hat{\mathbf{r}}) = H(\mathbf{x}|\hat{\mathbf{r}}) + H(\hat{\mathbf{r}})$. It is shown that the maximum of conditional entropy $H(\mathbf{x}|\hat{\mathbf{r}})$ is precisely the negative of the logarithm of the likelihood function $\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}})$ corresponding to the entropy-maximizing distribution for \mathbf{x} [33]. We have

$$\max_{P_{\mathbf{x}}} H(\mathbf{x}|\hat{\mathbf{r}}) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}))|_{P_{\mathbf{x}}=\Pi_{i=1}^N f_{\hat{\mathbf{r}}}(x_i)} \quad (13)$$

where uniform randomization in the SFNM modeling corresponds to the maximum uncertainty [22], [23]. Furthermore, maximizing the entropy of the parameter estimates $H(\hat{\mathbf{r}})$ results in

$$\max_{P_{\hat{\mathbf{r}}}} H(\hat{\mathbf{r}}) = \sum_{k=1}^{K_a} H(\hat{r}_k) \quad (14)$$

where we have used the result that, given the variance of parameter estimate determined by the corresponding sample average, the normal and independent distribution $P_{\hat{\mathbf{r}}}$ gives the maximum entropy [24], [26], [36].

Since the joint maximum entropy is a function of K_a and $\hat{\mathbf{r}}$, by taking the advantage of the fact that model estimation is separable in components and structure, we define the MCBV criterion as

$$\text{MCBV}(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{\text{ML}})) + \sum_{k=1}^{K_a} H(\hat{r}_{k\text{ML}}) \quad (15)$$

where $-\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{\text{ML}}))$ is the conditional bias (a form of information theoretic distance) [24], [26], and $\sum_{k=1}^{K_a} H(\hat{r}_{k\text{ML}})$ is the conditional variance (a measure of model uncertainty) [24], [36], of the model. As both of these two terms represent natural estimation errors about the true models, we treat them on an equal basis. A minimization of the expression in (15) leads to the following characterization of the optimum estimation:

$$K_0 = \arg \left\{ \min_{1 \leq K \leq K_{\text{MAX}}} \text{MCBV}(K) \right\}. \quad (16)$$

That is, if the cost of model variance is defined as the entropy of parameter estimates, the cost of adding new parameters to the model must be balanced by the reduction they permit in the ideal code length for the reconstruction error. A practical

MCBV formulation with code-length expression is further given by [18], [26]

$$\text{MCBV}(K) = -\log(\mathcal{L}(\mathbf{x}|\hat{\mathbf{r}}_{\text{ML}})) + \sum_{k=1}^{K_a} \frac{1}{2} \log 2\pi \cdot e^{-\text{Var}(\hat{r}_{k\text{ML}})} \quad (17)$$

where the calculation of $H(\hat{\mathbf{r}}_{k\text{ML}})$ requires the estimation of the true ML model parameter values. It is shown that, for sufficiently large number of observations, the accuracy of the ML estimation tends quickly to the best possible accuracy determined by the Cramér–Rao lower bounds (CRLB's) [36]. Thus, the CRLB's of the parameter estimates are used in the actual calculation to represent the "conditional" bias and variance [37]. We have found that, experimentally, the MCBV formulation for determining the value of K_0 exhibits very good performance consistent with both the AIC and the MDL criteria. It should be noted, however, that it is not the only plausible one; other criteria such as cross validation techniques may also be useful in this case [46], [48], [50], [52], [53].

We present a simulation study to test the performance of model selection with the proposed criterion (MCBV) and the two frequently used methods, AIC and MDL. We generate a test data with four overlapping normal components. Each component represents one local cluster. The value for each component is set to a constant value and normally distributed noise is then added to this simulation phantom with a signal-to-noise ratio (SNR) of 10 dB [38]. The phantom is shown in Fig. 1(a). The AIC, MDL, and MCBV curves, as functions of the number of local clusters K , are plotted in the same figure, Fig. 1(b). According to the information theoretic criteria, the minima of these curves indicate the correct number of the image components. From this experimental figure, it is clear that the number of local clusters suggested by these criteria are all correct. More application of the MCBV to the identification of real data structures will be presented in Section IV.

B. Probabilistic Self-Organizing Mixtures

There are many numerical techniques to perform the ML estimation of finite mixture distributions [25]. The most popular method is the expectation-maximization (EM) algorithm [44]. EM algorithm first calculates the posterior Bayesian probabilities of the data through the observations and the current parameter estimates (E -step) and then updates parameter estimates using generalized mean ergodic theorems (M -step). The procedure cycles back and forth between these two steps. The successive iterations increase the likelihood of the model parameters. A neural network interpretation of this procedure is given in [39]. However, EM algorithm has the reputation of being slow, since it has a first order convergence in which new information acquired in the expectation step is not used immediately [40]. Recently, on-line versions of the EM algorithm are proposed for large scale sequential learning. Such a procedure obviates the need to store all the incoming observations, changes the parameters immediately after each data point allowing for high data rates. Titterton [25] has developed a stochastic approximation procedure which is closely related to our approach, and shows that the solution

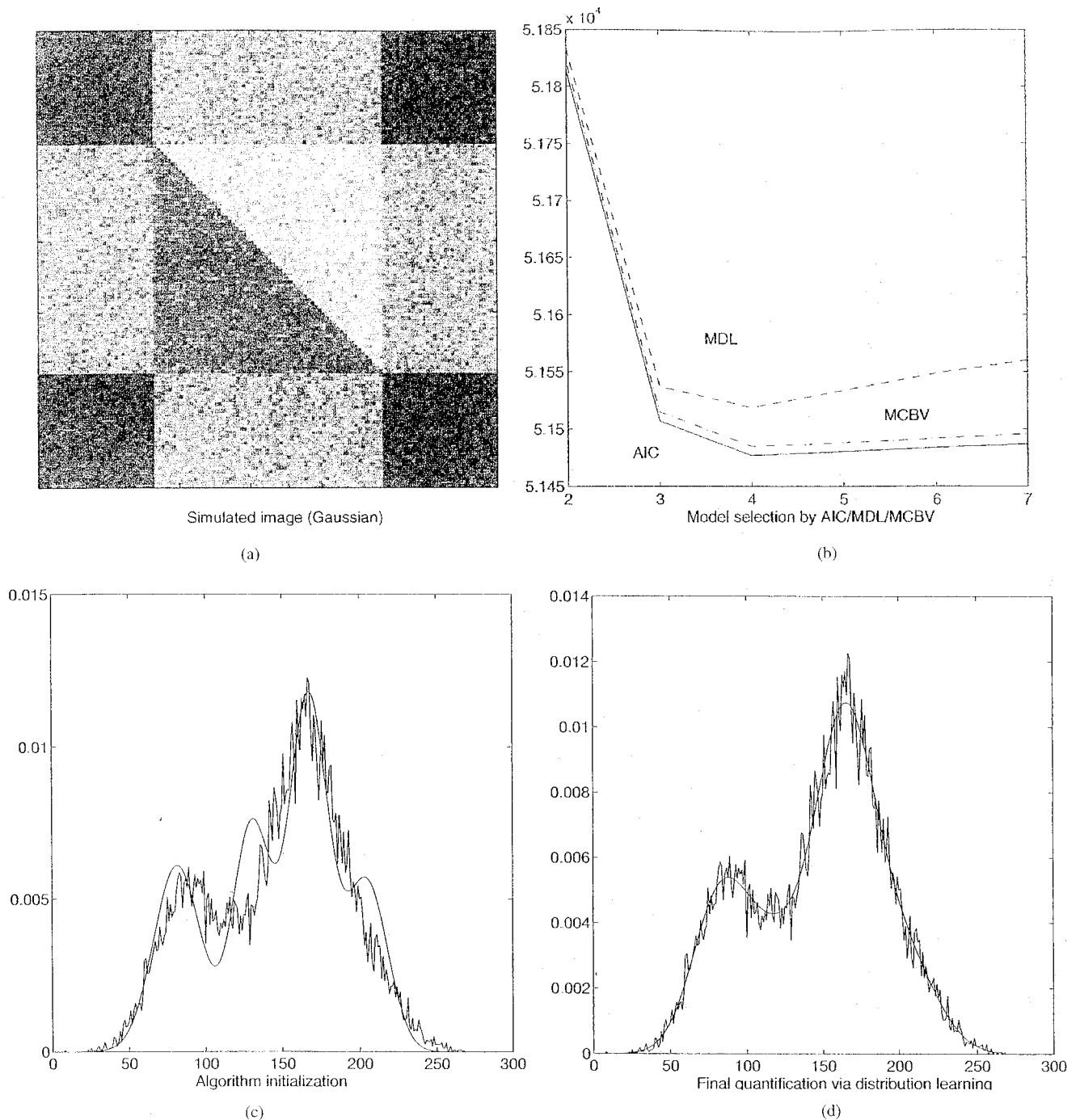


Fig. 1. Experimental results of model selection, algorithm initialization, and final quantification on the simulated image. (a) Original image with four components. (b) Curves of the AIC/MDL/MBV criteria where the minimum corresponds to $K_0 = 4$. (c) Initial histogram learning by the ALMHQ algorithm. (d) Final histogram learning by the PSOM algorithm.

can be made consistent. Other similar formulations are due to Marroquin *et al.* [29] and Weinstein *et al.* [41].

The PSOM we present here is a fully unsupervised and incremental stochastic learning algorithm, and is a generalized adaptive structure version of the SOFM algorithm we presented in [21]. The scheme provides winner-takes-in probability (Bayesian "soft") splits of the data, hence allowing the data to contribute simultaneously to multiple tissues. By differentiating $D(f_{\mathbf{x}}||f_{\mathbf{r}})$ given in (4) with respect to the unconstrained parameters, μ_k and σ_k^2 , we obtain the following

standard gradient descent learning rule for the mean and variance parameter vectors:

$$\begin{aligned} \mu_k^{(t+1)} &= \mu_k^{(t)} + \frac{\lambda}{N} \sum_{i=1}^N (x_i - \mu_k^{(t)}) \frac{z_{ik}^{(t)}}{\sigma_k^{2(t)}} \\ \sigma_k^{2(t+1)} &= \sigma_k^{2(t)} + \frac{\lambda}{N} \sum_{i=1}^N [(x_i - \mu_k^{(t)})^2 - \sigma_k^{2(t)}] \frac{z_{ik}^{(t)}}{2\sigma_k^{4(t)}} \end{aligned} \quad (18)$$

$$k = 1, \dots, K \quad (19)$$

TABLE I
TRUE PARAMETER VALUES AND THE ESTIMATES FOR THE SIMULATED IMAGE OF FIG. 1

k	True				Initial				Final			
	1	2	3	4	1	2	3	4	1	2	3	4
π	0.25	0.125	0.5	0.125	0.234	0.234	0.364	0.185	0.23	0.135	0.48	0.157
μ	86	126	166	206	81	131	167	205	84	121	164	201
σ^2	400	400	400	400	235	158	157	177	354	365	373	463

where λ is the learning rate and $z_{ik}^{(t)}$ is the posterior Bayesian probability, defined by

$$z_{ik}^{(t)} = \frac{\pi_k^{(t)} g(x_i | \mu_k^{(t)}, \sigma_k^{2(t)})}{f(x_i | \mathbf{r}^{(t)})} \quad (20)$$

By adopting a stochastic gradient descent scheme for minimizing $D(f_{\mathbf{x}} || f_{\mathbf{r}})$ [29], the corresponding on-line formulation is obtained by simply dropping the summation in (18) and (19) which results in

$$\mu_k^{(t+1)} = \mu_k^{(t)} + a(t)(x_{t+1} - \mu_k^{(t)})z_{(t+1)k}^{(t)} \quad (21)$$

$$\sigma_k^{2(t+1)} = \sigma_k^{2(t)} + b(t)[(x_{t+1} - \mu_k^{(t)})^2 - \sigma_k^{2(t)}]z_{(t+1)k}^{(t)} \quad (22)$$

$k = 1, \dots, K$

where the variance factors are incorporated into the learning rates while the posterior Bayesian probabilities are kept, and $a(t)$ and $b(t)$ are introduced as the learning rates, two sequences converging to zero, ensuring unbiased estimates after convergence. This modified version of the parameter updates is motivated by the principle that assigning different learning rates to different parameters of a network and allowing those to vary over time increases the rate of convergence [42]. Based on generalized mean ergodic theorem [26], updates can also be obtained for the constrained regularization parameters, π_k , in the SFNM model. For simplicity, given an asymptotically convergent sequence, the corresponding mean ergodic theorem, i.e., the recursive version of the sample mean calculation, should hold asymptotically. Thus, we define the interim estimate of π_k [43] by

$$\pi_k^{(t+1)} = \frac{t}{t+1}\pi_k^{(t)} + \frac{1}{t+1}z_{(t+1)k}^{(t)}. \quad (23)$$

Hence, the updates given by (21)–(23) provide the incremental procedure for computing the SFNM component parameters. Their practical use however requires strongly mixing condition and a decaying annealing procedure (learning rate decay) [26], [27], [36]. These two steps are currently controlled by user-defined parameters which may not be optimized for a specific problem. In addition, algorithm initialization must be chosen carefully and appropriately. In [43], we introduce an adaptive Lloyd–Max histogram quantization (ALMHQ) algorithm for threshold selection which is also well suited to initialization in ML estimation. In this work, we employ ALMHQ for initializing the network parameters μ_k , σ_k^2 , and π_k .

We tested the proposed technique using the same simulated image shown in Fig. 1(a). After the algorithm initialization by ALMHQ [43], network parameters are finalized by the PSOM algorithm. The GRE value is used as an objective measure to evaluate the accuracy of quantification. The results of the distribution learning are shown in Fig. 1(c) and (d). The GRE in the initial stage achieves a value of 0.0399 nats, and after

the final quantification by PSOM, is down to 0.008 nats. The numerical results are given in Table I where the unit of μ and σ^2 simply represents the observed gray levels of the pixel images while π is the probability measure. To simplify the representation, we omit their units as in [1], [5].

We also present a comparison of the performance of PSOM with that of the EM [23], [40], [44] and the competitive learning (CL) [29] algorithms in MR brain tissue quantification (see Section IV). We evaluate the computational accuracy and efficiency of the algorithm in standard finite normal mixture (SFNM) distribution learning, based on an objective criterion and its learning curve characteristics. For comparison, we applied all methods to the same example and used the GRE value between the image histogram and the estimated SFNM distribution as the goodness criterion to evaluate the quantification error. Fig. 2(a) shows learning curves of the PSOM and competitive learning (CL) algorithms, averaged over five independent runs. As observed in the figure, PSOM outperforms CL learning by faster convergence and lower quantification error, and reaches a final GRE value of about 0.04 nats. Fig. 2(b) presents the comparison of the performance of the PSOM algorithm with that of the EM algorithm for 25 epochs. As seen in the learning curves, PSOM algorithm again shows superior estimation performance. Note that since the EM algorithm uses intrinsically a batch learning mode, the learning curve appears very smooth when each point on the curve corresponds to a completed learning cycle in this case. The final quantification error is about 0.02 nats for PSOM with a faster convergence rate.

To conclude the discussion on PSOM, we address two issues regarding the nature of PSOM as it relates to neural computation. These are, the adjustment of structures in the feature space by the algorithm and the temporal dynamics of the learning process at the single neuron and the modular levels. Mapping the self-organizing operation to the PSOM, we design a network where both the structure and weights are updated according to an unsupervised learning procedure. Information theoretic criteria are shown to provide a reasonable approach for the solution of the problem. Another issue relating to the neural computational aspect of the PSOM procedure is the temporal dynamics of the learning process. As given by (21)–(23), learning in PSOM is a dynamic feedback competitive learning procedure in a self-organizing map (SOM) [27]. In particular, both the structure and the weights of the PSOM “compete” for the assignment order of each model and assignment probability of each observation. Overall convergence dynamics of the PSOM are similar to SOM in that a solution is obtained by “resonating” between input data and an internal representation. Such a mechanism

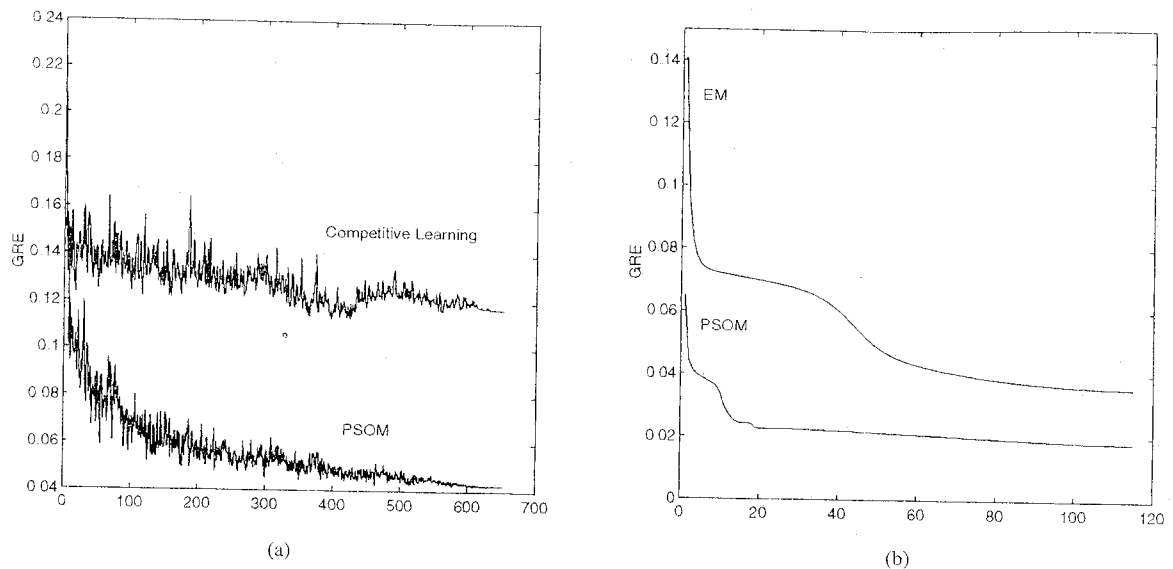


Fig. 2. Comparison of the learning curves of PSOM and CL (left) and EM (right).

can be considered as a more realistic learning than the batch EM procedure. In addition, temporal dynamics of the learning process for PSOM on the structure level, suggest the adjustment of the internal structure of a neural network as more information is acquired, i.e., the addition of new clusters.

C. Probabilistic Constraint Relaxation Networks

Given the SFNM parameters, i.e., the image components computed by the ML principle, there are several approaches to perform pixel classification. When the true pixel labels l_i^* are considered to be functionally independent and nonrandom constants, competitive learning approaches can be used for the segmentation of different tissue types [6], [8]. ML classification directly maximizes the individual likelihood function of pixel images by placing pixel i into the k th region, if

$$l_i^* = \arg \left\{ \min_k (\log(\sigma_k^2) + (x_i - \mu_k)^2 \sigma_k^{-2}) \right\} \quad (24)$$

where the term in parentheses is the modified Mahalanobis distance. On the other hand, when pixel labels are considered to be random variables, and the global context is taken as the prior information, probabilistic neural networks are most commonly used for tissue segmentation [4], [7]. By minimizing the expected value of the total Bayes classification error, pixel i will be classified into the k th region if

$$l_i = \arg \left\{ \min_k (\log(\sigma_k^2) - 2 \log(\pi_k) + (x_i - \mu_k)^2 \sigma_k^{-2}) \right\} \quad (25)$$

where the term in parentheses, since it incorporates the global prior information π_k , is called the Bayesian distance.

The major problem with these approaches is that the classification error will be high when the observed images are noisy, and possibly, there will be a high bias in the model parameters computed with sample averages after classification. We propose a probabilistic constraint relaxation network (PCRN) to perform tissue segmentation by imposing neighborhood context regularities to alleviate the two problems mentioned above. It operates on an initial segmented image, preferably

one with uniformly distributed classification errors, such as the one segmented by the classification-maximization (CM) algorithm [29]. PCRN uses stochastic discrete gradient descent procedure where each pixel is randomly visited and its label is updated [16], [45], i.e., pixel i is classified into the k th region if

$$l_i = \arg \left\{ \min_k (\log(\sigma_k^2) - 2 \log(\pi_{ik}) + (x_i - \mu_k)^2 \sigma_k^{-2}) \right\} \quad (26)$$

where π_{ik} is defined in (2) and the decision follows a probabilistic compatibility constraint given by

$$p(l_i | x_i, l_{\partial i}) = \sum_k I(l_i, k) \frac{p(l_i | l_{\partial i}) p(x_i | l_i)}{p(x_i | l_{\partial i})}$$

As discussed in Section II-C, by employing local maximization, relaxation labeling searches for a consistent labeling such that the average total consistency measure given by (10) is maximized for the given support function (8) [2]. It has been shown that relaxation labeling based on the stochastic discrete gradient descent principle converges to a steady point such that no label needs to be updated and the solution corresponds to at least one local maximum of $A(l)$ (10), [2], [16], [28], [29]. Iterations are needed to search for a consistent labeling, i.e., to maximize (10) for the given support function (8). During this relaxation process, our numerical experiments show that classification error decreases at every iteration and converges to a local maximum. Although a complete consistent labeling may not be reached in a practical implementation, the relaxation labeling algorithm, can provide a quite reasonable and accurate segmentation usually within few iterations [28]. The procedure can be summarized as follows.

PCRN Algorithm: Given $l^{(0)}$, $m = 0$

- 1) randomly visit each pixel for $i = 1, \dots, N$ (by random permutation of pixel ordering), and update its label l_i according to (26);
- 2) when the percentage of label changing less than $\epsilon\%$, stop. Otherwise, $m = m + 1$ and repeat step 2.

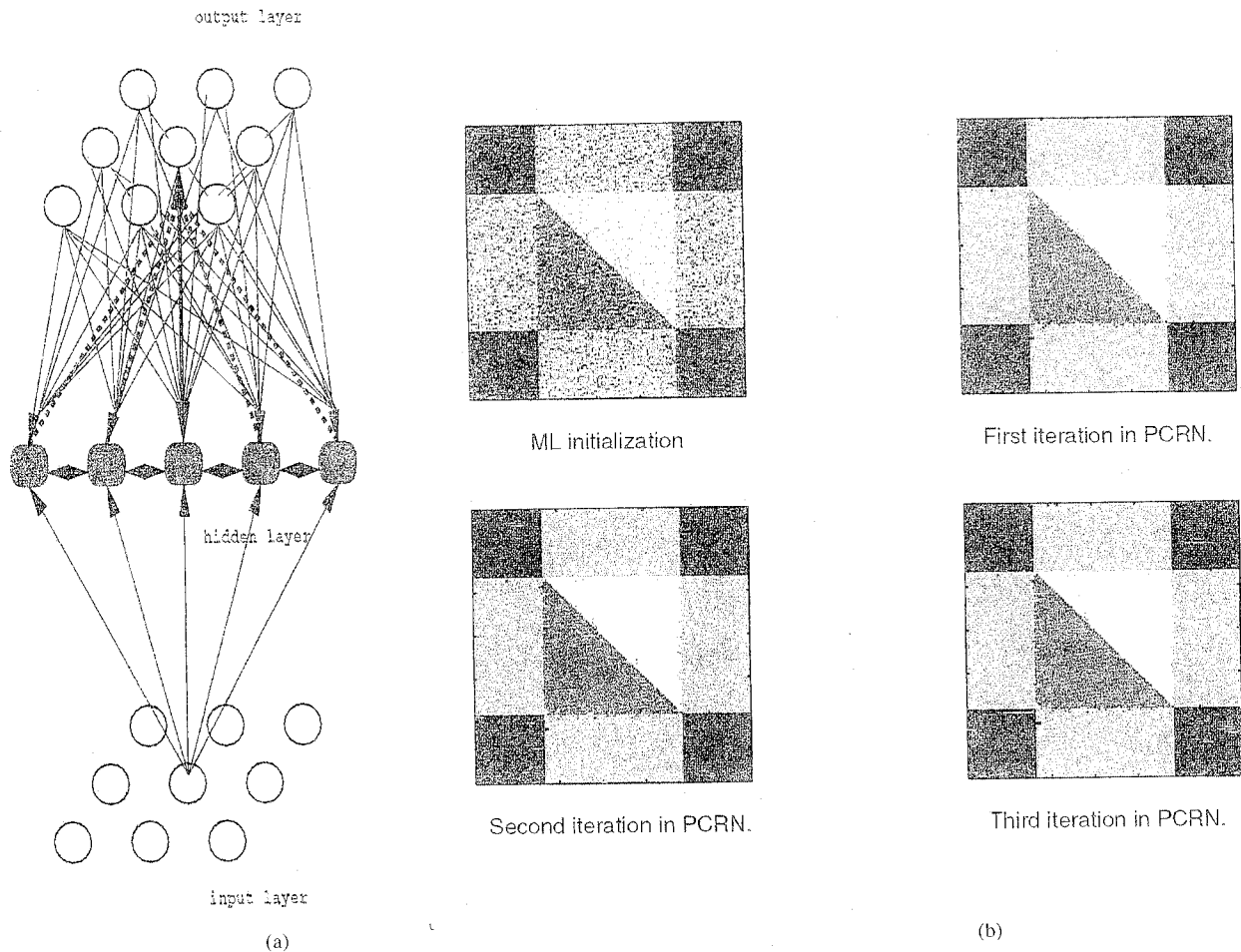


Fig. 3. (a) PCRN structure. (b) Image segmentation by PCRN on simulated image (with initialization by ML classification).

As mentioned before, it is desirable to start with an initial labeling $l^{(0)}$ which has classification errors that have spatial uniform distribution on the initial segmented image. Our experience has shown ML classification described by (24) to be a very good candidate to perform the initialization, i.e., to compute $l^{(0)}$ since it results in uniformly distributed classification errors. Also, a reasonable stopping criterion, suggested by our experimental results is 1%, i.e., choosing $\epsilon = 1$ in step 2.

As shown in Fig. 3(a), PCRN is composed of an N dimensional input array (the pixel images), a K dimensional hidden layer, and an N dimensional output array of pixel labels, such that each takes a value $l_i = k$, where $k = 1, \dots, K$. The number of the hidden units K , corresponding to the number of tissue types, is determined by information theoretic criteria as explained in Section II-A during tissue quantification. The estimates of the model parameters also determine the parameters μ_k and σ_k^2 for each of the K units $g(u|\mu_k, \sigma_k^2)$. Each of these K hidden units combines the local probabilistic constraint with the global intensity distribution information to produce an output which competes with the outputs of other hidden units to produce the labeling for the i th pixel, i.e., to determine the output l_i . The incorporation of the local context information is achieved by a gating function between the hidden units and the output, realizing π_{ik} given in (2), providing feedback from the output units to determine

the activation of the hidden unit. Hence, the network, rather than minimizing an energy function as in [6], [8], [29], looks for a possible local maximum of a global consistency measure by operating on local probabilistic constraints. It is derived directly from probabilistic constraints and can be classified as a recurrent noncausal competitive network with gating functions that incorporate context constraints. This approach demonstrates how a network of discrete units can be used to search an optimal solution to a problem that benefits the incorporation of context constraints.

Given the configuration of PCRN that is partially determined in model selection and estimation, the input layer of the PCRN has neurons corresponding to each pixel image and the output layer has neurons each corresponding to the labels of the original image. Competition within hidden layer ensures that only one neuron becomes active at any pixel location. This is accomplished by a winner-takes-all scheme among neurons, i.e., by a competitive learning procedure [29]. Gating between output and the hidden layer incorporates the local labeling information to provide locally consistent labeling and hence to remove the ambiguities. This is performed by the use of consistent measures between neighborhood neurons. Reciprocal feedback from output to gating unit allows each hidden neuron to control its activation. Another important difference between the PCRN and the conventional competitive learning network is that the recurrent gating provides a mechanism to

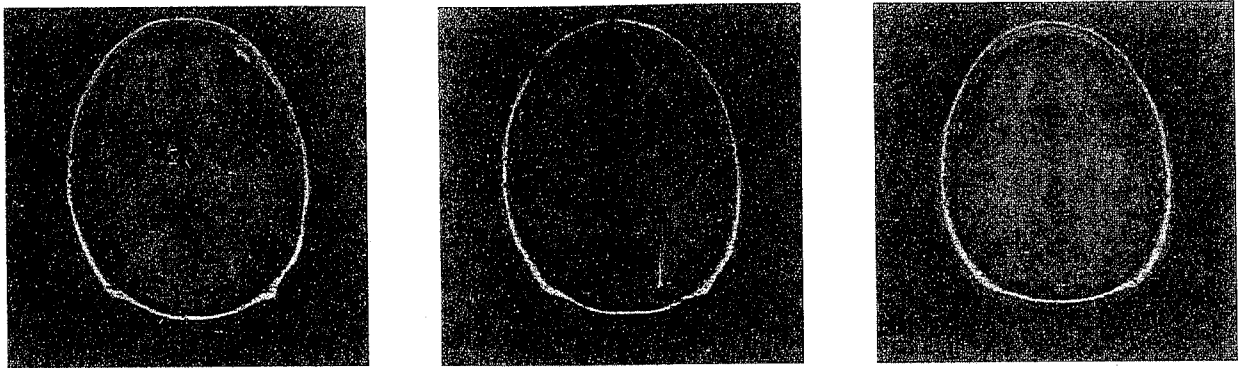


Fig. 4. Test sequence of MR image brain scans (original images).

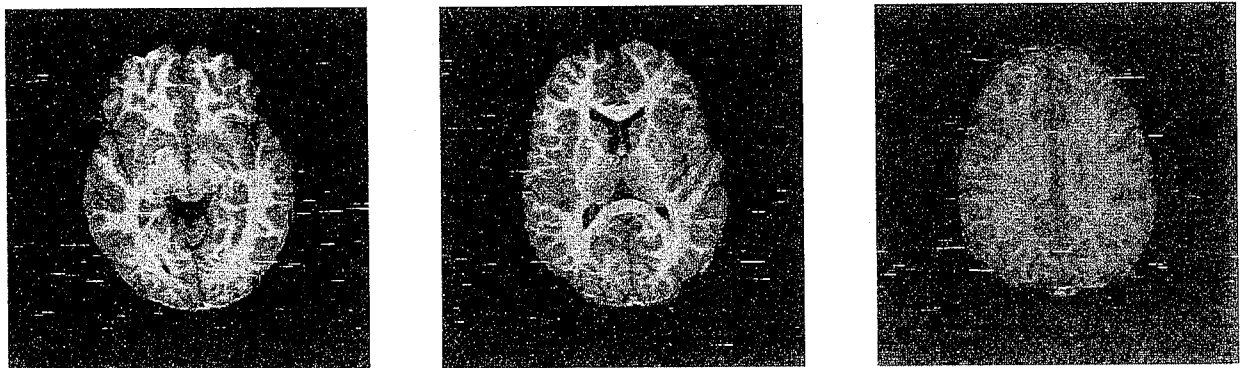


Fig. 5. Pure brain tissues extracted from original images.

incorporate the local Bayesian prior in the decision-making process through a consistency constraint. Without a similar mechanism, the conventional methods can only achieve, at best, a ML or a global Bayesian classification.

For validation of image segmentation using PCRN, we apply the algorithm first to the simulated images shown in Fig. 1(a). We use ML classifier to initialize the image segmentation, i.e., to initialize the quantified image by selecting the pixel label with the largest likelihood at each node by (24). Our experience suggested this to be a very suitable starting point for relaxation labeling [16]. PCRN is then used to fine tune the image segmentation. Since the true scene is known in this experiment, the percentage of total classification error is used as the criterion for evaluating the performance of the segmentation technique. In Fig. 3(b), the initial segmentation by the ML classification and the step-wise results of three iterations in PCRN are presented. In this experiment, algorithm initialization results in an average misclassification of 30%. It can be clearly seen that a dramatic improvement is obtained after several iterations of the PCRN by using local constraints determined by the context information. Also, note that the convergence is fast, as after the first iteration, most misclassifications are removed. The final percentage of classification errors for Fig. 3 is about 0.7935%.

IV. EXPERIMENTS AND RESULTS

In this section, we present results from real MR brain images using the probabilistic neural network based approach we

introduced to quantify and segment tissue types. In Section III, after introducing the algorithms, we presented results using a simulated tone image for which the number and structure of regions were known beforehand. The results presented showed the success of the scheme in determining the correct number of regions and the reliable definition of the boundaries of regions. In this section, we concentrate on application of the method to real MR images, which presents a great challenge to any computerized unsupervised analysis technique because of its complex structure. Furthermore, in addition to the assessment of radiologists, we also introduce application of an objective measure, GRE, to assess the performance of the scheme after quantification and segmentation, i.e., the soft and hard classification stages.

Fig. 4 shows the original data consisting of three adjacent, T1-weighted images parallel to the AC-PC line. The data are acquired with a GE Sigma 1.5 Tesla system. The imaging parameters are TR 35, TE 5, flip angle 45° , 1.5 mm effective slice thickness, 0 gap, 124 slices with in-plane 192×256 matrix, and 24 cm field of view. Since the skull, scalp, and fat in the original brain images do not contribute to the brain tissue, we edit the MR images to exclude nonbrain structures prior to tissue quantification and segmentation as explained in [16]. This also helps us to achieve better quantification and segmentation of brain tissues by delineation of other tissue types that are not clinically significant [1], [2], [5]. The extracted brain tissues are shown in Fig. 5. For each slice in the test sequence, the corresponding histograms are given in Fig. 6. As seen in the figure, the histogram has a

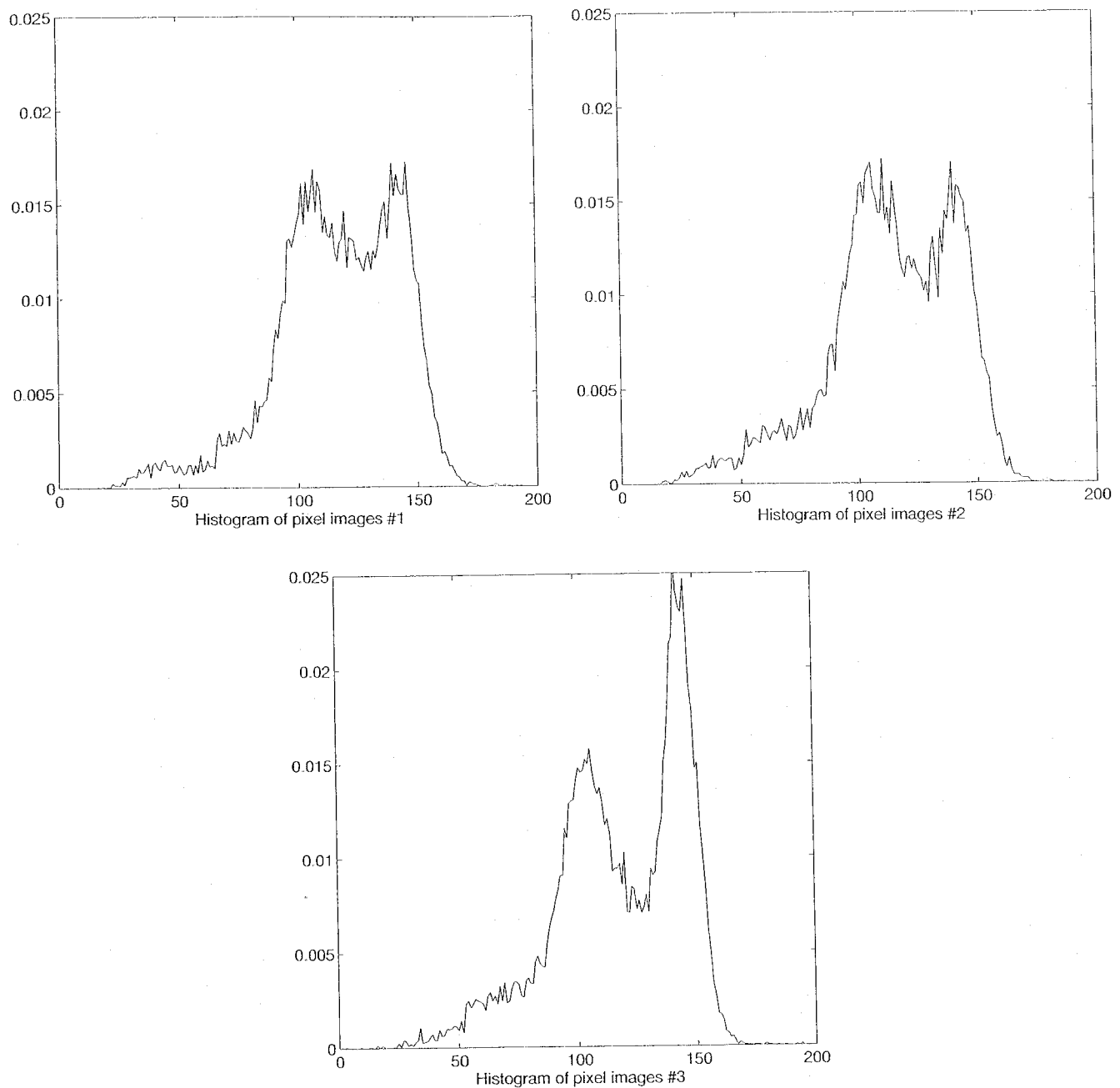


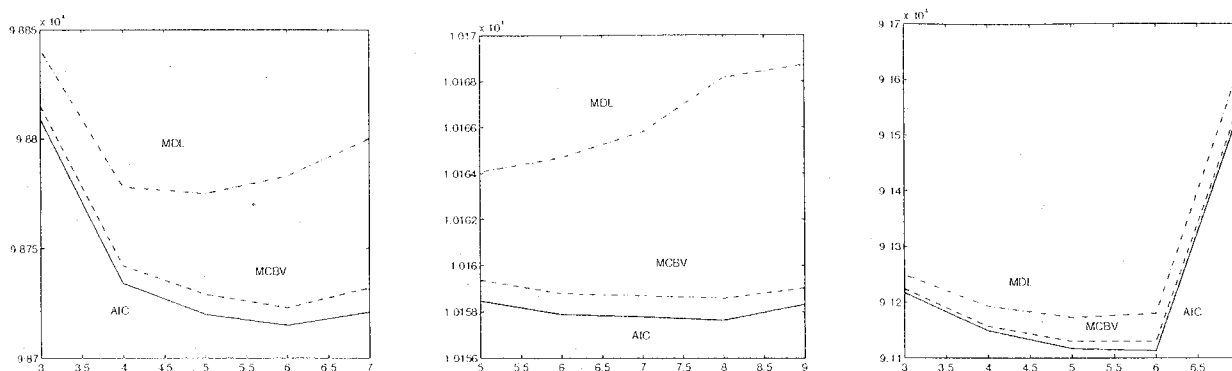
Fig. 6. Histograms of the brain tissue images.

considerably different characteristics from slice to slice and the tissue types are all highly overlapping making the problem quite complex. Our main objective is to assess the accuracy and repeatability of the results obtained with the method on real MR images. Evaluation of different image analysis techniques is a particularly difficult task, and dependability of evaluations by simple mathematical measures such as squared error performance is questionable. Therefore, most of the time, the quality assessment of the quantified and segmented image usually depends heavily on the subjective and qualitative judgements. As mentioned before, in this work, besides the evaluation performed by radiologists, we use the GRE value to reflect the quality of tissue quantification and also present results using EM and CL for image quantification to compare

the results of our scheme in terms of both the accuracy and the efficiency of the procedure. For assessment of tissue segmentation, we use post-segmentation sample averages as an indirect but objective criterion, and again use GRE values and visual inspection.

Based on the pre-edited MR brain image, the procedure for analysis of tissue types in a slice is summarized as follows.

- 1) For each value of K (number of tissue types), $K = K_{\min}, \dots, K_{\max}$, ML tissue quantification is performed by the PSOM algorithm [(20)–(23)].
- 2) Scan the values of $K = K_{\min}, \dots, K_{\max}$, use MCBV (16) to determine the suitable number of tissue types.
- 3) Select the result of tissue quantification corresponding to the value of K_0 determined in step 2.

Fig. 7. Results of model selection for slice 1-3 ($K_0 = 6, 8, 6$, left to right).TABLE II
RESULT OF PARAMETER ESTIMATION FOR SLICE 2

tissue type	1	2	3	4	5	6	7	8
π	0.0251	0.0373	0.0512	0.071	0.1046	0.1257	0.2098	0.3752
μ	38.848	58.718	74.400	88.500	97.864	105.706	116.642	140.294
σ^2	78.5747	42.282	56.5608	34.362	24.1167	23.8848	49.7323	96.7227

- 4) Initialize tissue segmentation by ML classification (23).
- 5) Finalize tissue segmentation by PCRN (by implementing (25) as explained in Section III-C).

The performance of tissue quantification and segmentation is then evaluated in terms of the GRE value, convergence rate, computational complexity, and visual judgement.

As discussed in the literature, the brain is generally composed of three principal tissue types, i.e., white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), and their pairwise combinations, called the partial volume effect. Santiago and Gage [1] have proposed a six-tissue model representing the primary tissue types and the mixture tissue types which were defined as CSF-White (CW), CSF-Gray (CG), and Gray-White (GW). In this work, we also consider the triple mixture tissue, defined as CSF-White-Gray (CWG). More importantly, since the MR image scans clearly show the distinctive intensities at local brain areas, the functional areas within a tissue type need to be considered. In particular, the caudate nucleus and putamen are two important local brain functional areas. In our experiment, as we have noted before, we allow the number of tissue types to vary from slice to slice, i.e., consider adaptability to different MR images. We let $K_{\min} = 2$ and $K_{\max} = 9$ and calculate $AIC(K)$ (11), $MDL(K)$ (12), and $MCBV(K)$ (15) for $K = K_{\min}, \dots, K_{\max}$. The results with these three criteria are shown in Fig. 7, which suggest that the brain images contain six, eight, and six tissue types, respectively. According to the model fitting procedure using information theoretic criteria as explained before, the minima of these criteria indicate the most appropriate number of the tissue types, which is also the number of hidden nodes in the corresponding PSOM (mixture components in SFNM). In the calculation of MCBV using (16), we used the CRLB's to represent the conditional variances of the parameter estimates,

given by [37]

$$\text{Var}(\hat{\pi}_{k\text{ML}}) = \frac{\pi_k(1 - \pi_k)}{N} \quad (27)$$

$$\text{Var}(\hat{\mu}_{k\text{ML}}) = \frac{\sigma_k^2}{N\pi_k}, \quad \text{and} \quad (28)$$

$$\text{Var}(\hat{\sigma}_{k\text{ML}}^2) = \frac{2\sigma_k^4(N\pi_k - 1)}{N^2\pi_k^2} \quad (29)$$

Note that since the true parameter values in above equations are not available, their ML estimates are used to obtain the approximate CRLB's. From Fig. 7, it is clear that, with real MR brain images, the overall performance of the three information theoretic criteria is fairly consistent. Our experience suggests that, however, AIC tends to overestimate while MDL tends to underestimate the number of tissue types [38], and MCBV provides a solution between those of AIC and MDL, which we believe to be more reasonable especially in terms of providing a balance between the bias and variance of the parameter estimates.

When performing the computation of the information theoretic criteria, we used PSOM to iteratively quantify different tissue types for each fixed K . The PSOM algorithm is initialized by a fully automatic thresholding technique, i.e., the adaptive ALMHQ procedure that we have introduced in [43]. For slice 2, the results of final tissue quantification with $K_0 = 7, 8, 9$ are shown in Fig. 8. Table II gives the numerical result of final tissue quantification for slice 2 corresponding to $K_0 = 8$, where a GRE value of 0.02-0.04 nats is achieved. It was found that most of the variance parameters are different, which suggests that assuming same variance for each tissue type with distinct image-intensity distribution is not very realistic. These quantified tissue types agree with those of physician's qualitative analysis results [54], [55].

The PCRN tissue segmentation for slice 2 is performed with $K_0 = 7, 8, 9$, and the algorithm is initialized by ML

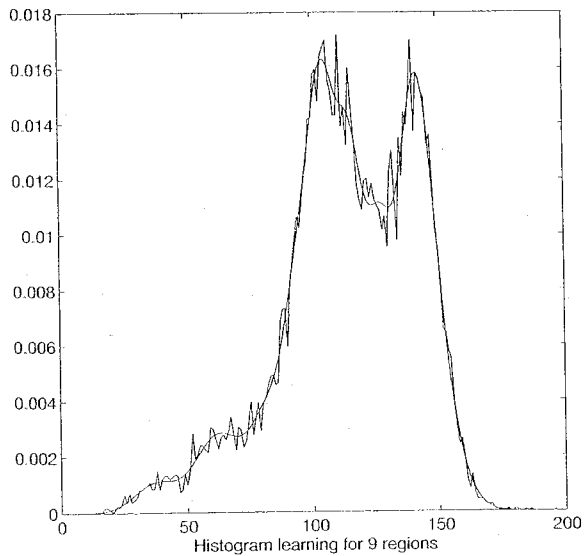
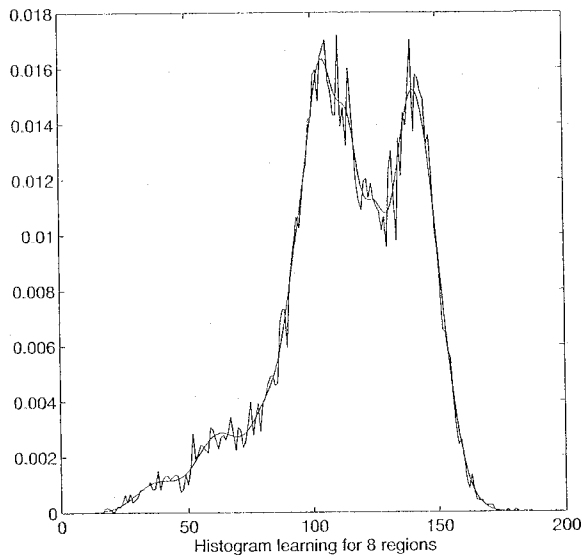
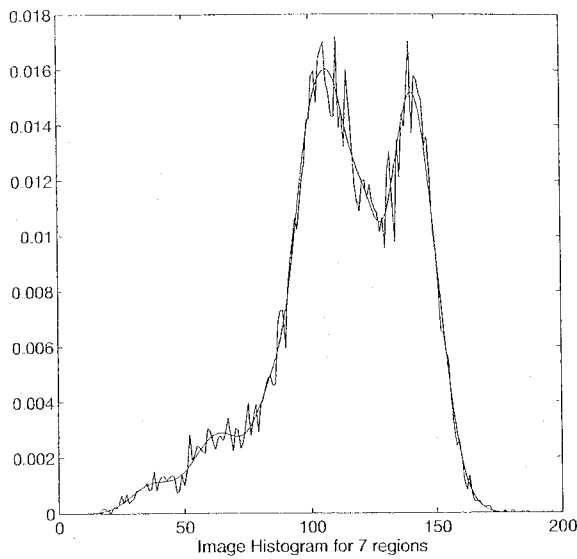


Fig. 8. Histogram learning for slice 2 ($K = 7, 8, 9$ from top to bottom).

TABLE III
COMPARISON OF SEGMENTATION ERROR RESULTING FROM
NONCONTEXTUAL AND CONTEXTUAL METHODS FOR SLICE 2

Method	PSOM (<i>soft</i>)	GBC (<i>hard</i> -GBC)	PCRN (<i>hard</i> -PCRN)
GRE value (nats)	0.0067	0.4406	0.1578

classification [see (24)]. PCRN updates are terminated after five to ten iterations, since further iterations produced almost identical results. The segmentation results are shown in Fig. 9. Although the segmentation result contains some small isolated spots (less than four-pixel size), the PCRN approach is quite encouraging. It is seen that the boundaries of WM, GM, and CSF are delineated very well and successfully. To see the benefit of using information theoretic criteria in determining the number of tissue types, the decomposed tissue type segments are given in Fig. 10 with $K_0 = 8$. As can be observed in Figs. 9 and 10, the segmentation with eight tissue types provides a very meaningful result. The regions with different gray levels are satisfactorily segmented, especially, the major brain tissues are clearly identified. If the number of tissue types were "underestimated" by one, tissue mixtures located within putamen and caudate areas would be lumped into one component, though the results are still meaningful. When the number of tissue type was "overestimated" by one, there is no significant difference in the quantification result, but white matter has been divided into two components. For $K_0 = 8$, the segmented regions represent eight types of brain tissues: (a) CSF, (b) CG, (c) CGW, (d) GW, (e) GM, (f) putamen area, (g) caudate area, and (h) WM as shown in Fig. 10. These segmented tissue types also agree with the results of radiologists' evaluation [54], [55].

We then test the hypotheses that: i) tissue segmentation using the prior constraint that the MR image has a locally piecewise continuous structure provides better results than those of using global regularization together with local intensity values [called global Bayesian classification (GBC)]; and ii) tissue quantification using *soft* classification (i.e., without realizing the value of l_i , by ML quantification) is more accurate than the quantification results obtained by using sample averages computed after *hard* pixel classification, (i.e., by a winner-takes-all scheme), or than those obtained in conjunction with such a scheme. For this task, slice 2 is segmented and postquantified, using the Bayesian approach [i.e., global Bayesian classification based on (25)] and the sample averages. The global Bayesian approach is not iterative and does not require a stopping point. In this work, the performance is evaluated by the post-GRE values for all schemes, which is consistent with model-based ergodic principle and allows for uniform comparison among various techniques. Table III gives the classification errors by these two methods in terms of postquantification errors. It can be seen that quantification by PSOM results in lower error than GBC and PCRN, with PCRN resulting in lower GRE value. This result implies that the intrinsic misclassification in tissue segmentation creates a biased parameter estimate that contributes to the higher quantification error, as also noted in

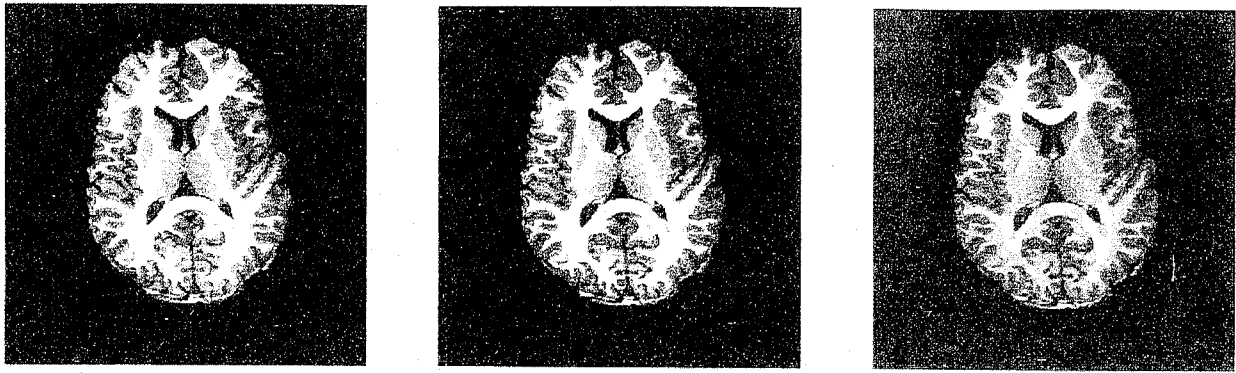


Fig. 9. Results of tissue segmentation for slice 2 with $K_0 = 7, 8, 9$ (from left to right).

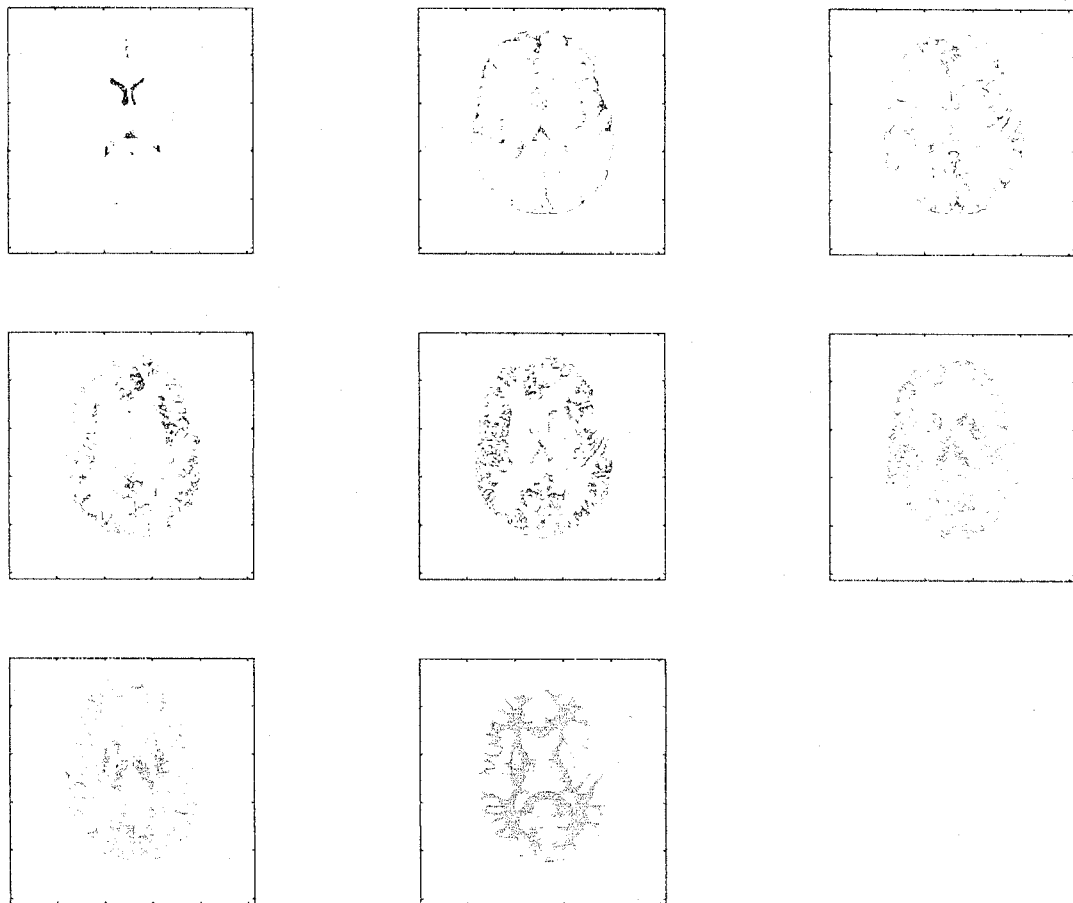


Fig. 10. Result of tissue type decomposition for slice 2 which represent eight types of brain tissues: CSF, CG, CGW, GW, GM, putamen area, caudate area, and WM (left to right, top to bottom).

[23]. It is very interesting to note that, since ergodic theorem is the most fundamental one behind any statistical model-based image analysis approach, postquantification may be a suitable objective criterion for evaluating the quality of image segmentation in a fully unsupervised situation.

V. DISCUSSIONS AND CONCLUSIONS

We have presented a complete procedure for quantifying and segmenting major brain tissue types from MR images,

in which two kinds of probabilistic neural networks: soft and hard classifiers, are employed. The MR brain image is modeled by a standard finite normal mixture model and an extended localized formulation. Information theoretic criteria are applied to detect the number of tissue types thus allowing the corresponding network to adapt its structure for the best representation of the data. The PSOM algorithm is used to quantify the parameters of tissue types leading to a ML estimation. Segmentation of identified tissue components is

then implemented by PCRN through Bayesian decision. The results obtained by using the simulated image and real MR brain images demonstrate the promise and effectiveness of the proposed technique. In particular, the number of tissue types and the associated parameters were consistently estimated. The tissue types were satisfactorily segmented. Although the current algorithms were tested for 2-D images, their application to 3-D situations is straightforward by appropriate neighborhood function in PCRN.

Our main contribution is the complete proposal of a three-step learning strategy for determination of both the modular structure and the components of the network. In this approach, the network structure (in terms of suitability of the statistical model) is justified in the first step. It is followed by soft segmentation of data such that each data point supports all local components simultaneously. The associated probabilistic labels are then realized in the third step by competitive learning of this induced hard classification task.

We introduced a model selection scheme that explicitly incorporates the bias and variance dilemma in finite data training. When tested with synthetic and actual data, the results show that the number of hidden nodes in PSOM should be adjusted to match the data, and hence order selection may be important to consider. Theory is developed showing that ML quantification and Bayesian classification have distinct objectives, and both soft and hard classification problems are studied which describe performance differences. The quantification results from the presegmentation and the postsegmentation stages generated the evidence. However, the results of tissue segmentation that includes probabilistic constraints, indicate that the use of local context information can provide better results that is often consistent with the recurrent network structure.

The main limitations of the current approach are that i) it requires the testing of all possible network structure candidates during the model fitting procedure, hence is not efficient especially for processing MR sequence images where an on-line learning might be preferred, and ii) applications to real MR data indicates the possibility of being trapped in a local maximum in ML estimation by the PSOM since there is no guarantee of attaining the global maximum.

There are possible ways to mitigate these problems: Since one possible contribution to the local minima problem is imperfect initialization, we use a simple automated threshold selection, based on Lloyd–Max histogram quantization [43], to systematically initialize the algorithm during model selection and quantification. Experimental results suggested that the method is quite effective in a variety of situations with different data structures [16], [21], [43]. To address the first limitation mentioned above, we tested an adaptive model selection procedure by incorporating the correlation between slices in a given MR sequence. More precisely, model selection starts from a slice in the middle of the sequence and moves in each direction, such that for slice $i + 1$, we set $K_{\max}^{i+1} = K_0^i + 2$ and $K_{\min}^{i+1} = K_0^i - 2$ where K_0^i is the optimal number of tissue types for slice i given by the information theoretic criteria. It should be addressed, however, that they are by no means the only, or the best, possible solutions; in

fact, it will be interesting to compare the effect of random and systematic algorithm initialization on the final performance, and further study is needed for interpretation of the results of these information theoretic criteria: AIC, MDL, and MCBV.

To summarize, the results of the experiments we have performed indicate the plausibility of our approach for brain tissue analysis from MRI scans, and show that it can be applied to clinical problems such as those encountered in tissue segmentation and quantitative diagnosis.

ACKNOWLEDGMENT

The authors would like to thank T. Lei of the University of Maryland, Baltimore, S. K. Mun and M. T. Freedman of the Georgetown University Medical Center, and R. F. Wagner of the U.S. Food and Drug Administration, for their valuable input and guidance on this work.

REFERENCES

- [1] P. Santago and H. D. Gage, "Quantification of MR brain images by mixture density and partial volume modeling," *IEEE Trans. Med. Imag.*, vol. 12, pp. 566–574, Sept. 1993.
- [2] A. J. Worth and D. N. Kennedy, "Segmentation of magnetic resonance brain images using analog constraint satisfaction neural networks," in *Information Processing in Medical Imaging*, H. H. Bartlett and A. F. Gmitro, Eds. Berlin, Germany: Springer-Verlag, 1993, pp. 225–243.
- [3] H. S. Choi, D. R. Haynor, and Y. Kim, "Partial volume tissue classification of multichannel magnetic resonance images—A mixture model," *IEEE Trans. Med. Imag.*, vol. 10, pp. 395–407, Sept. 1994.
- [4] H. E. Cline, W. E. Lorensen, R. Kikinis, and R. Jolesz, "Three-dimensional segmentation of MR images of the head using probability and connectivity," *J. Comp. Assist. Tomog.*, vol. 14, pp. 1037–1045, 1990.
- [5] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter estimation and tissue segmentation from multispectral MR images," *IEEE Trans. Med. Imag.*, vol. 13, pp. 441–449, Sept. 1994.
- [6] K. S. Cheng, J. S. Lin, and C. W. Mao, "The application of competitive Hopfield neural network to medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 15, pp. 560–567, Aug. 1996.
- [7] M. Morrison and Y. Attikiouzel, "A probabilistic neural network based image segmentation network for magnetic resonance images," in *Proc. Conf. Neural Networks*, Baltimore, MD, 1992, vol. 3, pp. 60–65.
- [8] A. P. Dhawan and L. Arata, "Segmentation of medical images through competitive learning," *Comput. Meth. Prog. Biomed.*, vol. 40, pp. 203–215, 1993.
- [9] L. O. Hall *et al.*, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," *IEEE Trans. Neural Networks*, vol. 3, pp. 672–682, 1992.
- [10] Y. Wang, T. Adali, C. M. Lau, and Z. Szabo, "Quantification of MR brain images by a probabilistic self-organizing map," *Radiology*, vol. 197, pp. 252–253, Nov. 1995.
- [11] Y. Wang, T. Adali, C.-M. Lau, and S. Y. Kung, "Quantitative analysis of MR brain image sequences by adaptive self-organizing finite mixtures," *J. VLSI Tech. Signal Process.*, vol. 8, pp. 219–239, 1998.
- [12] T. Adali, Y. Wang, and N. Gupta, "A block-wise relaxation labeling scheme and its application to edge detection in cardiac MR image sequences," to be published.
- [13] W. C. Lin, E. C. K. Tsao, and C. T. Chen, "Constraint satisfaction neural networks for image segmentation," *Pattern Recognit.*, vol. 25, pp. 679–693, 1992.
- [14] Y. Wang and T. Adali, "Probabilistic neural networks for parameter quantification in medical image analysis," in *Biomedical Engineering Recent Development*, J. Vossoughi, Ed., 1994.
- [15] Y. Wang, "MR Imaging statistics and model-based MR image analysis," Ph.D. dissertation, Univ. Maryland, College Park, MD, May 1995.
- [16] Y. Wang, T. Adali, M. T. Freedman, and S. K. Mun, "MR brain image analysis by distribution learning and relaxation labeling," in *Proc. 15th South. Biomedical Engineering Conf.*, Dayton, OH, Mar. 1996, pp. 133–136.
- [17] T. Adali, X. Liu, and M. K. Sönmez, "Conditional distribution learning with neural networks and its application to channel equalization," *IEEE Trans. Signal Processing*, vol. 45, pp. 1051–1064, Apr. 1997.

- [18] Y. Wang, "Image quantification and the minimum conditional bias/variance criterion," in *Proc. 30th Conf. Information Science and Systems*, Princeton, NJ, Mar. 20-22, 1996, pp. 1061-1064.
- [19] Y. Wang and T. Lei, "A new stochastic model-based image segmentation technique for MR images," *Proc. 1st IEEE Int. Conf. Image Processing*, Austin, TX, 1994, pp. 182-185.
- [20] M. Fuderer, "The information content of MR images," *IEEE Trans. Med. Imag.*, vol. 7, pp. 368-380, 1988.
- [21] Y. Wang and T. Adali, "Efficient learning of finite normal mixtures for image quantification," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Atlanta, GA, 1996, pp. 3422-3425.
- [22] C. Bouman and B. Liu, "Multiple resolution segmentation of texture images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 99-113, Feb. 1991.
- [23] D. M. Titterton, "Comments on 'application of the conditional population-mixture model to image segmentation'," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 656-658, Sept. 1984.
- [24] J. Rissanen, "Minimax entropy estimation of models for vector processes," *System Identification, Advances and Case Studies*. New York: Academic, 1976, pp. 97-119.
- [25] D. M. Titterton, A. F. M. Smith, and U. E. Markov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.
- [26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
- [27] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: Macmillan, 1994.
- [28] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, May 1983.
- [29] J. L. Marroquin and F. Girosi, "Some extensions of the K-means algorithm for image segmentation and pattern classification," Tech. Rep., MIT Artif. Intell. Lab., Cambridge, MA, Jan. 1993.
- [30] T. Adali, M. K. Sönmez, and K. Patel, "On the dynamics of the LRE Algorithm: A distribution learning approach to adaptive equalization," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Detroit, MI, 1995, pp. 929-932.
- [31] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, Dec. 1974.
- [32] J. Zhang and J. M. Modestino, "A model-fitting approach to cluster validation with application to stochastic model-based image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 1009-1017, Oct. 1990.
- [33] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Stat.*, vol. 11, 1983.
- [34] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Comput.*, vol. 4, pp. 1-52, 1992.
- [35] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 108, pp. 620-630/171-190, May 1957.
- [36] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Berlin, Germany: Springer-Verlag, 1988.
- [37] L. I. Perlovsky, "Cramer-Rao Bounds for the estimation of normal mixtures," *Pattern Recognit. Lett.*, vol. 10, pp. 141-148, 1989.
- [38] Y. Wang, S.-H. Lin, H. Li, and S.-Y. Kung, "Data mapping by probabilistic modular network and information theoretic criteria," *IEEE Trans. Signal Processing*, 1998, to be published.
- [39] L. Perlovsky and M. McManus, "Maximum likelihood neural networks for sensor fusion and adaptive classification," *Neural Networks*, vol. 4, pp. 89-102, 1991.
- [40] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixture," Tech. Rep., MIT Artif. Intell. Lab., Cambridge, MA, Jan. 1995.
- [41] E. Weinstein, M. Feder, and A. V. Oppenheim, "Sequential algorithms for parameter estimation based on the Kullback-Leibler information measure," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1652-1654, 1990.
- [42] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks*, vol. 1, pp. 295-307, 1988.
- [43] Y. Wang, T. Adali, and B. Lo, "Automatic threshold selection by histogram quantization," *SPIE J. Biomed. Opt.*, vol. 2, pp. 211-217, Apr. 1997.
- [44] R. A. Redner and N. M. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, pp. 195-239, 1984.
- [45] H. Li, Y. Wang, K.-J. Liu, and S.-H. Lo, "Morphological filtering and model-based segmentation of masses on mammographic images," *IEEE Trans. Med. Imaging*, to be published.
- [46] H. Li, et al., "Detection of masses on mammograms using advanced segmentation techniques and an HMOE classifier," in *Digital Mammography*. Amsterdam, The Netherlands: Elsevier, 1996.
- [47] A. P. 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.*, vol. 13, pp. 716-724, Dec. 1994.
- [48] L. Perlovsky, W. Schoendorf, B. Burdick, and D. M. Tye, "Model-based neural network for target detection in SAR images," *IEEE Trans. Image Processing*, vol. 6, pp. 203-216, Jan. 1997.
- [49] H. Gish, "A probabilistic approach to the understanding and training of neural network classifiers," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, 1990, pp. 1361-1364.
- [50] J. L. Marroquin, "Measure fields for function approximation," *IEEE Trans. Neural Networks*, vol. 6, pp. 1081-1090, 1995.
- [51] H. Li, *Model-Based Image Processing Techniques for Breast Cancer Detection in Digital Mammography*, Ph.D. dissertation, Univ. Maryland, College Park, May 1997.
- [52] J. H. Friedman, "On bias, variance, 0/1-loss, and the curse-of-dimensionality," Tech. Rep., Stanford Univ., Stanford, CA, 1996.
- [53] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 33, no. 2, Apr. 1985.
- [54] Z. Szabo, Dept. Nuclear Med., Johns Hopkins Medical Institutions, Baltimore, MD, private communication.
- [55] M. Freedman, M.D., Dept. Radiology, ISIS Ctr., Georgetown Univ. Med. Ctr., Washington, DC, private communication.



Yue Wang (S'93-M'95) received the B.S. and M.S. degrees from Shanghai Jiao Tong University, Shanghai, China, in 1984 and 1987, respectively, and the Ph.D. degree from University of Maryland-Baltimore County, in 1995, all in electrical engineering.

During 1995-1996, he was a Post-doctoral Fellow at the Georgetown University Medical Center, Washington, DC. In 1996, he joined the Department of Electrical Engineering and Computer Science, Catholic University of America, Washington, DC, as

an Assistant Professor. He is also affiliated with the Department of Radiology, Georgetown University School of Medicine, as an Adjunct Assistant Professor. He is the Guest Editor for the *Journal of Intelligent Automation and Soft Computing*. He is currently on the National Cancer Institute working group on lung cancer imaging. His research interests include image analysis, medical imaging, information visualization, data base mapping, volumetric display, visual explanation, and their applications in biomedicine and multimedia informatics.

Dr. Wang is the recipient of a 1998 U.S. Army Medical Research and Material Command CDA Award.



Tülay Adalı (S'88-M'92) received the B.S. degree from Middle East Technical University, Ankara, Turkey, in 1987 and the M.S. and Ph.D. degrees from North Carolina State University, Raleigh, in 1988 and 1992, respectively, all in electrical engineering.

In 1992, she joined the Department of Electrical Engineering, University of Maryland-Baltimore County, as an Assistant Professor. Her research interests are in neural computation, adaptive signal processing, estimation theory, and their applications

in channel equalization, biomedical image analysis, soliton communications, and time-series prediction. She has authored or co-authored more than 75 publications.

Dr. Adalı is the Guest Editor of a special issue of the *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology on Neural Networks* on neural networks for biomedical image processing. She currently serves on the IEEE Technical Committee on Neural Networks for Signal Processing and the IEEE Signal Processing Conference Board. She is the recipient of a 1997 National Science Foundation CAREER Award.

Sun-Yuan Kung (S'74-M'78-SM'84-F'88), for a photograph and biography, see this issue, p. 1095.



Zsolt Szabo received the M.D. degree from the University of Belgrade and the Ph.D. degree from the University of Dusseldorf. He received postgraduate training at the University of Dusseldorf, Germany, and the Johns Hopkins Medical Institutions, Baltimore, MD, and special training in positron emission tomography (PET) at the Nuclear Research Center, Julich, Germany.

He is an Associate Professor of Radiology at the Johns Hopkins Medical Institutions. His research activities involve PET/MRI coregistration, partial volume correction, investigation of serotonin and dopamine transporters and investigation of angiotensin II receptors with PET. His special interests include nuclear medicine and PET with special emphasis on receptor imaging and neuroscience.

Dr. Szabo is member of the Society of Nuclear Medicine and of the German Association of Nuclear Medicine.