

Markov Random Field Segmentation of Brain MR Images

Karsten Held, Elena Rota Kops,* Bernd J. Krause, William M. Wells, III,
Ron Kikinis, and Hans-Wilhelm Müller-Gärtner

Abstract— We describe a fully-automatic three-dimensional (3-D)-segmentation technique for brain magnetic resonance (MR) images. By means of Markov random fields (MRF's) the segmentation algorithm captures three features that are of special importance for MR images, i.e., nonparametric distributions of tissue intensities, neighborhood correlations, and signal inhomogeneities. Detailed simulations and real MR images demonstrate the performance of the segmentation algorithm. In particular, the impact of noise, inhomogeneity, smoothing, and structure thickness are analyzed quantitatively. Even single-echo MR images are well classified into gray matter, white matter, cerebrospinal fluid, scalp-bone, and background. A *simulated annealing* and an *iterated conditional modes* implementation are presented.

Index Terms— Magnetic resonance imaging, Markov random fields, segmentation.

I. INTRODUCTION

EXCELLENT soft-tissue contrast and high spatial resolution make magnetic resonance imaging (MRI) the method for anatomical imaging in brain research. Segmentation of the MR image into different tissues, i.e., gray matter (GM), white matter (WM), cerebrospinal fluid (CSF), scalp-bone and other nonbrain tissues (SB), and background (BG),¹ is an important prerequisite for:

- three-dimensional (3-D) visualization and peeling off nonbrain structures;
- quantitative analysis of brain morphometry;

Manuscript received December 9, 1996; revised July 17, 1997. The co-operation between the Research Centre Jülich, Institute of Medicine and the Harvard Medical School, Brigham and Women's Hospital was supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant Ro 1149/1-1; the work in the Harvard Medical School, Brigham & Women's Hospital was supported by a Whitaker Foundation grant, by the National Institutes of Health under Grant NIH PO1 CA 67165-01A1, and by the National Science Foundation under Grant NSF BES-9 631 710. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was T. M. Taxt. Asterisk indicates corresponding author.

K. Held is with the Institute of Medicine, Research Center Jülich GmbH, D-52425 Jülich, Germany. He is also with the Institute of Theoretical Physics, University Augsburg, 86135 Augsburg, Germany.

*E. Rota Kops is with the Institute of Medicine, Research Center Jülich GmbH, D-52425 Jülich, Germany (e-mail: e.rota.kops@fz-juelich.de).

B. J. Krause is with the Institute of Medicine, Research Center Jülich GmbH, D-52425 Jülich, Germany. He is also with the Department of Nuclear Medicine, Heinrich-Heine-University Hospital, 40225 Düsseldorf, Germany.

W. M. Wells, III and R. Kikinis are with Harvard Medical School and Brigham and Women's Hospital, Department of Radiology, Boston, MA 02115 USA.

H.-W. Müller-Gärtner is with the Institute of Medicine, Research Center Jülich GmbH, Jülich D-52425 Germany. He is also with the Department of Nuclear Medicine, Heinrich-Heine-University Hospital, 40225 Düsseldorf, Germany.

Publisher Item Identifier S 0278-0062(97)09169-6.

¹For clinical applications specific abnormal tissue classes can be added.

- matching MR onto functional images [1], [2];
- partial volume correction [3].

Though several segmentation techniques are present in the literature, the fully-automatic segmentation of MR images remains difficult (for a review see [4]), mainly due to the somewhat noisy MR data, caused by time and equipment limitations. Another persistent difficulty is the spatial inhomogeneity of the MR signal. Some promising algorithms exist, but most of them require multiecho images and some pre- and postprocessing to improve segmentation. To this end, nonlinear filters [5], [6], morphological [7] and connectivity [8] operations, and even snake models [9] are discussed in the literature. Besides additional effort, these processes frequently tend to suppress fine details of high-resolution images. The improvement undoubtedly obtained by the use of multiecho (high-resolution) MR images imposes the penalty of long acquisition times, which are unacceptable for some applications.

In order to address these difficulties, we have developed a new Markov random field (MRF) segmentation algorithm based on an *adaptive segmentation* algorithm described by Wells *et al.* [10]. The new approach uses MRF as a convenient means for introducing context or dependence among neighboring voxels. It incorporates the following important characteristics:

- 1) nonparametric distribution of tissue intensities are described by Parzen-window statistics [11];
- 2) neighborhood tissue correlations are taken into account by MRF to manage the noisy data;
- 3) signal inhomogeneities are also modeled by *a priori* MRF.

Several methods for addressing these issues can be found in the literature, but the algorithm presented here is the first that addresses all three simultaneously.

Wells *et al.* [10] use nonparametric, Parzen-window statistics [11] and adapt a bias field to the inhomogeneities, but do not regard neighborhood dependencies for the tissue segmentation. Additional filtering and connectivity operations are also used.

Geman and Geman [12] were the first to apply the methods of statistical mechanics² to image segmentation. They use an *a priori* probability model for neighboring voxels and some additional hidden-edge elements. But they do not take account of

²The same methods are employed in statistical mechanics for microscopic models, on a $O(10^{-6})$ finer scale, to calculate the T_1 and T_2 relaxation times [13]–[15] used in MR imaging to discriminate between tissues.

nonparametric intensity distributions and the inhomogeneities that are important for MR images.

Pappas' [16] adaptive K-means clustering algorithm uses neighborhood dependencies, but only parametric Gaussian intensity distributions. Inhomogeneities are spatially adapted by adjusting the Gaussian means. In medical imaging this algorithm was applied to computed tomography (CT) images [17]. Related MRF clustering algorithms without inhomogeneity correction were also employed for MR images [18]–[20].

In Section II we explain the MRF segmentation, i.e., the statistical model (capturing the neighborhood MRF, the inhomogeneity MRF, and the Parzen-window intensity distribution) and the optimization [by simulated annealing (SA) or iterated conditional modes (ICM)]. The acquisition and simulation of MR images is described thereafter. The segmentation of these images and a quantitative analysis of the impact of noise, inhomogeneity, smoothing, and structure thickness follow in Section III. In Section IV we discuss the advantages and disadvantages of the SA and two ICM algorithms, differing in the inhomogeneity correction, and compare the MRF segmentation to the *adaptive segmentation* [10].

II. MATERIALS AND METHODS

A. Markov Random Field Segmentation

A natural way of incorporating spatial correlations into a segmentation process is to use MRF's [12], [16], [21], [22] as *a priori* models. The MRF is a stochastic process that specifies the *local* characteristics of an image and is combined with the given data to reconstruct the true image. The MRF of prior contextual information is a powerful method for modeling spatial continuity and other features, and even simple modeling of this type can provide useful information for the segmentation process. The MRF itself is a *conditional* probability model, where the probability of a voxel depends on its neighborhood. It is equivalent to a Gibbs *joint* probability distribution [12] determined by an energy function. This energy function is a more convenient and natural mechanism for modeling contextual information than the local *conditional* probabilities of the MRF. The MRF, on the other hand, is the appropriate method to sample the probability distribution.

In the following we denote the observed MR echo intensities as \mathbf{z} , a vector containing an individual intensity z_i for every voxel i in the 3-D MR image. The segmentation task is to classify the tissue of every voxel i , i.e., to determine the segmentation vector \mathbf{x} with discrete values $x_i \in \{\text{BG, WM, GM, CSF, and SB}\}$. To model intensity inhomogeneities an additional vector \mathbf{y} has to be calculated. By multiplying y_i with the echo intensity z_i , the inhomogeneity corrected MR echo is obtained [10].

As pointed out in Section I, the MR segmentation algorithm includes nonparametric statistics, neighborhood correlations, and signal inhomogeneities. A MRF *a priori* probability $p(\mathbf{x})$ for the segmented image is used to model the spatial correlations within the image. A smooth inhomogeneity field \mathbf{y} is provided by a restrictive *a priori* MRF distribution $p(\mathbf{y})$. Given the *a priori* probabilities for the tissue x_i and the

inhomogeneity y_i at a voxel i , the conditional probability of the observed echo intensity z_i is calculated by a Parzen-window distribution $p(z_i | x_i, y_i)$.

For given MR intensities \mathbf{z} , Bayes rule is used to calculate the *a posteriori* probability of the segmentation \mathbf{x} and the inhomogeneity \mathbf{y}

$$p(\mathbf{x}, \mathbf{y} | \mathbf{z}) \propto p(\mathbf{z} | \mathbf{x}, \mathbf{y})p(\mathbf{x})p(\mathbf{y}). \quad (1)$$

This probability is maximized either by SA or ICM. The corresponding segmentation \mathbf{x} is the *maximum a posteriori* (MAP) estimator of the true image. The three terms on the right side of (1) are now explained.

1) *Neighborhood MRF*: The discrete Fourier transformation to reconstruct MR image has the effect that neighboring voxels contribute to echo intensities by a *sinc* function. Partial volume effects additionally smooth fine anatomical structures. These effects, and especially the anatomic arrangement of the tissues, lead to spatial correlations in the MR intensities. These correlations are exploited to manage with the somewhat noisy MR data and, hence, to improve the segmentation.

Balancing computing time versus *a priori* information, the first-order neighborhood of six nearest neighbors was chosen. Including dependencies between more voxels will dramatically increase the computing time, e.g., by a factor of 26/6 if all third-order neighbors of a voxel are included. On the contrary, information gain is small since strong dependencies between distant voxels cannot model the complex, convoluted boundaries between brain tissues. With a neighborhood system of six nearest neighbors in three dimensions, we employ the probability distribution

$$p(\mathbf{x}) \propto e^{-E(\mathbf{x})} \quad (2)$$

with the Gibbs energy

$$E(\mathbf{x}) = \sum_{\langle i, j \rangle} e_{x_i x_j} \quad (3)$$

where $\langle i, j \rangle$ denotes the sum over all voxels i and their six nearest, first-order neighbors j . The local potential $e_{x_i x_j}$ is the most general potential between two voxels. It must be minimal for neighboring voxels of the same tissue ($x_i = x_j$) to prefer equal segmentation at neighboring voxels, i.e., to incorporate the spatial correlations within the MR image.

The additional *a priori* information—that the scalp-bone tissue is not connected with cerebral tissues—is used by setting a high potential $e_{x_i x_j}$ for $x_i = \text{SB}$ and $x_j \in \{\text{GM, WM, CSF}\}$, or vice versa. This potential facilitates the correct classification of scalp-bone even with strongly varying intensities.

2) *Inhomogeneity MRF*: A main difficulty in MR segmentation is the variation of the MR signal caused by inhomogeneities of the magnetic fields, especially of the excitation (B_1) field,³ and the sensitivity profile of the receiving coil. Therefore, a segmentation that is correct in one part of the volume will likely fail in another part.

³The tissue, the patient's head, may have significant impact on the inhomogeneity of the MR signal. In our approach we do not take into account this possible dependence, i.e., details on a possible origin of the inhomogeneity. Furthermore, the inhomogeneity is assumed to have a multiplicative effect on the MR echo intensities that is the same for all tissues.

This is corrected by adapting an additional, continuous inhomogeneity field \mathbf{y} , also called the bias field [10]. As the magnetic field is slowly varying, the *a priori* probability $p(\mathbf{y})$ must provide a smooth inhomogeneity field \mathbf{y} . We present a new approach using a MRF with a local potential connecting six nearest neighbors

$$p(\mathbf{y}) \propto e^{-U(\mathbf{y})} \quad (4)$$

with the Gibbs energy

$$U(\mathbf{y}) = \alpha \sum_{\langle i,j \rangle} (y_i - y_j)^2 + \beta \sum_i y_i^2. \quad (5)$$

At first sight it seems to be insufficient to regard a local neighborhood of six neighbors only to get a smooth inhomogeneity field. But it is well known from statistical mechanics that such local potentials yield high spatial correlations, as two distant points are connected by a number of local potentials.⁴ Indeed, similar local potentials are used in physics to model ferromagnetic phases with infinite correlation length. The Gaussian MRF was chosen since the quadratic potential grows quickly and will, therefore, yield a smooth bias field. In image segmentation, MRF's with a linear potential for large neighbor differences are frequently used (e.g., see [17]). These potentials are of interest to preserve edges that are absent in the inhomogeneity field.

Further *a priori* information of the inhomogeneity field \mathbf{y} is modeled by β , i.e., small inhomogeneity corrections are more probable than large. In Section III we compare this approach to implementation of the inhomogeneity field used in the *adaptive segmentation* algorithm by Wells *et al.* [10]. While the theory of the *adaptive segmentation* is based on a *a priori* bias field models, in the implementation the bias field is smoothed by a linear lowpass filter [10, eq. 21].

3) *Parzen-Window Intensity Distribution*: Physical and chemical differences between tissues lead to different T_1 and T_2 Blochian relaxation times visualized in T_1 - and T_2 -weighted MR echoes. Noise and in-tissue variations yield non-Gaussian, $T_1 - T_2$ correlated distributions of the MR echo intensities.

Nonparametric statistics are used to describe the insufficiently separated tissue intensity distributions as accurately as possible. To this end, the conditional probability of observing the echo intensity z_i for given segmentation x_i and inhomogeneity y_i is modeled by a Parzen-window distribution [11]. For every tissue class \bar{x} a set of $n_{\bar{x}}$ training points $\bar{z}_{k,\bar{x}}$ ($k = 1, \dots, n_{\bar{x}}$) is selected by a supervisor (see Kikinis *et al.* [23]). The Parzen-window distribution is obtained by centering a small Gaussian of width $\sqrt{\Sigma}$ around each training point

$$p(\mathbf{z} | \mathbf{x}, \mathbf{y}) = \frac{1}{n_{x_i}} \sum_{k=1}^{n_{x_i}} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(y_i z_i - \bar{z}_{k,x_i})^t \Sigma^{-1} (y_i z_i - \bar{z}_{k,x_i})} \quad (6)$$

⁴ At n iterations the inhomogeneity field is influenced by all voxels in a n th-order neighborhood.

where the covariance matrix Σ is equal to σ^2 in the case of single-echo MR images ($d = 1$) and to σ^2 times the two-dimensional (2-D) unit matrix for double-echo MR images ($d = 2$). In the latter case y_i , z_i , and \bar{z}_{k,x_i} have two components, one for each echo.

This distribution deals with correlations between the echo signals. Even nonconnected probability distributions can be described, e.g., SB tissue shows nonconnected small and high echo intensities. For computational efficiency this probability distribution is stored in a look-up table as proposed in [23].

4) *Simulated Annealing (SA)*: A Markov process is used to generate random configurations $\{\mathbf{x}, \mathbf{y}\}$ according to the probability distribution (1). A new configuration $\{\mathbf{x}', \mathbf{y}'\}$ is accepted with a probability $W(\{\mathbf{x}, \mathbf{y}\} \rightarrow \{\mathbf{x}', \mathbf{y}'\})$ preserving detailed balance

$$\begin{aligned} p(\mathbf{x}', \mathbf{y}' | \mathbf{z}) W(\{\mathbf{x}', \mathbf{y}'\} \rightarrow \{\mathbf{x}, \mathbf{y}\}) \\ = p(\mathbf{x}, \mathbf{y} | \mathbf{z}) W(\{\mathbf{x}, \mathbf{y}\} \rightarrow \{\mathbf{x}', \mathbf{y}'\}). \end{aligned} \quad (7)$$

According to Metropolis *et al.* [14] a prototype acceptance probability is

$$W(\{\mathbf{x}, \mathbf{y}\} \rightarrow \{\mathbf{x}', \mathbf{y}'\}) = \min \left\{ 1, \frac{p(\mathbf{x}', \mathbf{y}' | \mathbf{z})}{p(\mathbf{x}, \mathbf{y} | \mathbf{z})} \right\}. \quad (8)$$

Due to the local potential chosen, this probability can be calculated effectively when $(\mathbf{x}', \mathbf{y}')$ differs from (\mathbf{x}, \mathbf{y}) at a single voxel. Since detailed balance is fulfilled and every configuration is accessible from any other configuration, ergodicity is established [12] and the Markov process will yield segmentations \mathbf{x} and inhomogeneities \mathbf{y} obeying (1). Testing one new segmentation and inhomogeneity for every voxel is called a Monte-Carlo sweep. From a set of sweeps one may calculate mean values as well as standard deviations and correlations within the segmentation. We follow a slightly different approach to obtain the MAP estimate instead of these mean values, i.e., SA.

SA was introduced by Kirkpatrick *et al.* [24] and refers to the slow decrease of a control parameter T that corresponds to temperature in physical systems. We modify (1) in the following way:

$$p(\mathbf{x}, \mathbf{y} | \mathbf{z}) \rightarrow \exp \left(\frac{1}{T} \log p(\mathbf{x}, \mathbf{y} | \mathbf{z}) \right). \quad (9)$$

As T decreases, samples from the *a posteriori* distribution are forced toward the minimal energy, i.e., toward the MAP. If $T(p)$ satisfies

$$T(p) \geq \frac{\text{const}}{\log(1+p)} \quad (10)$$

at the p th Monte-Carlo sweep, it can be proved that this annealing schedule theoretically guarantees convergence to the global MAP [12]. The starting configuration for the SA algorithm is zero inhomogeneity $\mathbf{y} = 0$ and the segmentation \mathbf{x} , that would be chosen by a point-by-point segmentation solely due to the Parzen-window distribution, i.e., this \mathbf{x} maximizes (6) with $\mathbf{y} = 0$ [this is equivalent to (14) if the neighborhood information, not available at this time, is disregarded]. The temperature was reduced according to $T(p) = 1/\log(1+p)$.

5) *Iterated Conditional Modes (ICM)*: Especially for clinical applications the 3-D SA-segmentation algorithm requires too much computing time, at least for current single processor systems. Therefore, we present an alternative method, i.e., ICM proposed by Besag [21]. Following this concept, the algorithm maximizes the *a posteriori* probability with respect to the segmentation \mathbf{x} and the inhomogeneity \mathbf{y} iteratively.

In a first step for every voxel i , the most probable discrete segmentation x_i , maximizing (1) at fixed \mathbf{y} and neighboring segmentation \mathbf{x}_{∂_i} is chosen⁵

$$\max_{x_i} \{p(\mathbf{x}, \mathbf{y} | \mathbf{z})\} = \max_{x_i} \left\{ p(z_i | x_i, y_i) e^{-\sum_{j \in \partial_i} e_{x_i x_j}} \right\}. \quad (11)$$

In a second step the continuous inhomogeneity y_i is maximized for every voxel i at fixed \mathbf{x} and neighboring inhomogeneity \mathbf{y}_{∂_i}

$$\frac{\partial}{\partial y_i} p(\mathbf{x}, \mathbf{y} | \mathbf{z}) = 0. \quad (12)$$

To solve this equation with respect to y_i the Parzen-window distribution (6) is too complicated. Therefore, we follow Wells *et al.* [10] and heuristically replace (6) by a Gaussian distribution for the logarithmic echo intensities $\tilde{z}_i = \ln z_i$ with tissue dependent mean \bar{z}_i and covariance matrix Σ_{x_i}

$$p(z_i | x_i, y_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_{x_i}|^{1/2}} e^{-\frac{1}{2}(\tilde{z}_i + y_i - \bar{z}_{x_i})^t \Sigma_{x_i}^{-1} (\tilde{z}_i + y_i - \bar{z}_{x_i})}, \quad (13)$$

Differentiation of (12) yields

$$\frac{\partial}{\partial y_i} \left\{ \frac{1}{2} (\tilde{z}_i + y_i - \bar{z}_{x_i})^t \Sigma_{x_i}^{-1} (\tilde{z}_i + y_i - \bar{z}_{x_i}) + \alpha \sum_{j \in \partial_i} (y_i - y_j)^2 + \beta y_i^2 \right\} = 0.$$

Solving this equation gives the most probable inhomogeneity y_i for fixed segmentation and neighborhood inhomogeneities

$$y_i = A^{-1} \left\{ 2\alpha \Sigma_{x_i} \sum_{j \in \partial_i} y_j + (\bar{z}_{x_i} - \tilde{z}_i) \right\} \quad (14)$$

with

$$A = (2\alpha n + 2\beta) \Sigma_{x_i} + 1, \quad n = 6.$$

Following (11) and (14) an iterated conditional modes algorithm (ICM1) is implemented. The algorithm ICM1 begins, as SA, with zero inhomogeneity $\mathbf{y} = 0$ and the segmentation \mathbf{x} maximizing (6) for $\mathbf{y} = 0$. From this segmentation an inhomogeneity field is calculated according to (14). The inhomogeneity field on the other hand, together with the old neighborhood segmentation, allows determination of the new segmentation according to (11). Continuing the iteration of

(14) and (11) will converge to some local, not necessary global maximum of the *a posteriori* probability.

Another algorithm (ICM2) using the inhomogeneity approach of Wells *et al.* [10] is presented. Here, step (14) is replaced by calculating the inhomogeneity y_i locally

$$y_i = \frac{R_i}{N_i} \quad (15)$$

with residual

$$R_i = \sum_{x_i} p(z_i | x_i, y_{i_{\text{old}}}) p(x_i) \Sigma_{x_i}^{-1} (\bar{z}_{x_i} - \tilde{z}_i) \quad (16)$$

and normalizer

$$N_i = \sum_{x_i} p(z_i | x_i, y_{i_{\text{old}}}) p(x_i) \Sigma_{x_i}^{-1}. \quad (17)$$

In addition, all possible segmentations x_i are evaluated according to their probability $p(z_i | x_i, y_{i_{\text{old}}}) p(x_i)$ at the old inhomogeneity field $y_{i_{\text{old}}}$. Instead of the *a priori* inhomogeneity distribution (4), a linear lowpass filter is used to obtain a smooth residual R_i and normalizer N_i before calculating the inhomogeneity \mathbf{y} according to (15). Further details can be found in [10, Section 2.3]. ICM2 is initialized in the same way as ICM1 and SA, and then iterates (15) and (11).

6) *Parameter Choice*: The strength of the MRF is that simple modeling of this type yields additional information that considerably helps to segment otherwise ambiguous voxels. The MRF probability should not be oversized since a cluster of voxels is first of all classified according to its MR signal. Using faithful anatomical potentials $e_{x_i x_j}$ that were calculated from the statistics of a segmented image did not always improve the segmentation since some probabilities were extremely low, e.g., to have a CSF voxel in a GM/WM neighborhood. Consequently, in parts of the image, CSF voxels were misclassified as WM/GM. This led us to the following considerations for setting suitable parameters.

To take all three features into account (i.e., neighborhood correlations, signal inhomogeneities, and Parzen-window distribution of tissue intensities), the energies (3), (5), and (6) must have the same order of magnitude. As the tissue intensities typically differ from their mean value by the standard deviation, the corresponding energy is 1/2 for (6). The neighborhood contribution (3) must have the same order of magnitude. Two different energies $e_{x_i x_j}$ for neighboring voxels of the same (set to zero) and of different tissues (set to ε) are used. With six neighbors contributing, ε should be about $1/2 \times 1/6$. As the contribution due to the MR intensity is more important than this *a priori* knowledge, we actually choose $\varepsilon = 0.05$.

The *a priori* information on the inhomogeneity is described by (5). The parameter β corresponds to the expected inverse variance of the inhomogeneity field, e.g., if the inhomogeneities have a standard deviation of $\Delta I = 0.1$ then $\beta = 1/(2\Delta I^2) = 50$. In order to allow larger inhomogeneities and not be too severely restricted by this *a priori* information, we used $\beta = 20$. The parameter α describes the standard deviation of the inhomogeneity gradient. To obtain a smooth inhomogeneity field α is set to 100 allowing an inhomogeneity variation of $I = 0.1$ within about five voxels.

⁵ If the maximal probability is below some fixed value the voxel is labeled as unclassified. Unclassified voxels are indifferent in the potential $e_{x_i x_j}$.

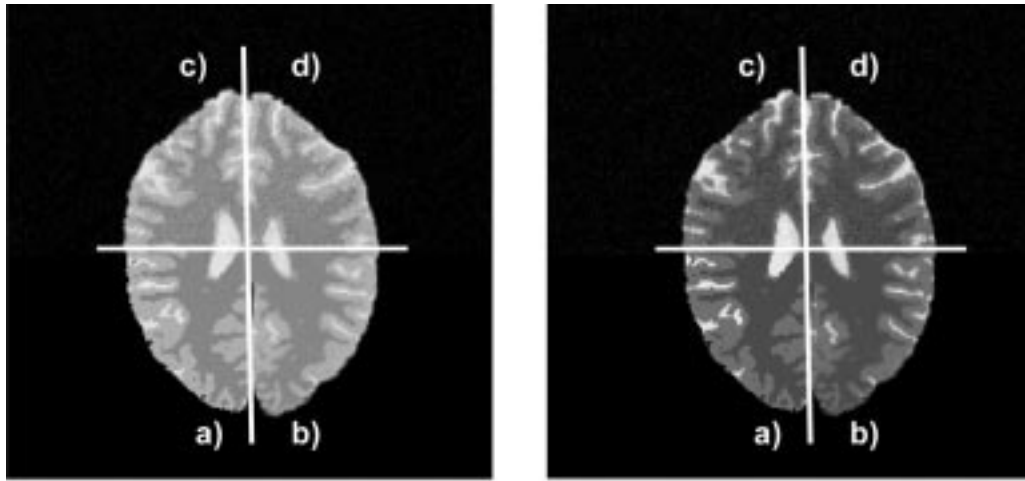


Fig. 1. The four tasks performed to simulate double-echo [(left) pd and (right) T_2] MR images: The intensities are initialized (a) according to their mean intensities, (b) afterwards smoothed ($S = 0.2$), (c) noised ($N = 50$), and, finally, (d) an inhomogeneity is simulated ($I = 0.1$).

B. Acquisition of MR Images

Brain MR images of six volunteers were acquired on a Siemens Vision, 1.5 Tesla. The sequence used was a turbo spin echo pd/T_2 -weighted double-echo with $T_R = 6536$ ms, $T_{E1} = 20$ ms, and $T_{E2} = 120$ ms. At an image resolution of $1 \times 1 \times 1$ mm³ a field of view of $256 \times 160 \times 32$ mm³ (32 slices for each dataset) was scanned within 10.5 min. For this sequence the equipment specific noise, as estimated by the mean BG signal, was about 30 (relative units) for both echoes. Mean tissue signals (standard deviations) were calculated from a segmented MR slice as follows:

pd -echo: SB = 456(120), WM = 823(70), GM = 1059(95), CSF = 1363(177);

T_2 -echo: SB = 167(69), WM = 426(59), GM = 602(102), CSF = 1223(307).

These two echoes are correlated and a criterion that takes this into account is the Mahalanobis distance D between two tissues (\bar{x} and \bar{x}'), i.e., $D^2 = (\bar{x} - \bar{x}')^T \Sigma^{-1} (\bar{x} - \bar{x}')$ where Σ is the mean value of the two covariance matrices of the two tissues. As one can already guess from the mean echo signals and their standard deviations, the tissues that are most difficult to distinguish are WM-GM ($D = 2.9$) and GM-CSF ($D = 3.0$).

For a direct qualitative comparison, the single-echo algorithm was tested using the pd -weighted image of these double echo MR images which has the better WM-GM contrast of the two MR echoes.

C. Simulation of MR Images

To validate the segmentation algorithms, as well as to compare them quantitatively, we have used simulated MR images. A template segmentation of a $256 \times 256 \times 16$ MR image with 1 mm isotropic resolution was used. Smoothing, noise and inhomogeneity were modeled as follows (see Fig. 1).

- 1) The template segmentation was defined as the correct segmentation. The mean values were calculated for all tissues from the corresponding double-echo MR image (see Section II-B). Each voxel intensity was set as specified in Section II-B.

- 2) This image was smoothed with six nearest, first-order, neighbors in three dimensions. The weight of every neighboring voxel, relative to the weight of the center voxel, was denoted as S . At a value of $S = 1/6$, the echo intensity of one voxel was 50% due to its tissue and 50% due to the neighboring tissues.
- 3) Two-dimensional Gaussian noise of standard deviation N was added. A relative measure for this noise was the contrast-to-noise ratio (CNR) between two tissues (\bar{x} and \bar{x}') $CNR_{\bar{x}, \bar{x}'} = (\bar{z}_{\bar{x}'} - \bar{z}_{\bar{x}})/N$. At the above mean pd -echo signals a typical noise level of $N = 50$ corresponded to a contrast-to-noise ratio $CNR_{WM, GM} = 4.7$ and to a signal-to-noise ratio $SNR_{GM} = 21$ for GM.
- 4) The image was multiplied by an inhomogeneity factor which changed linearly with Euclidean distance from the point $(x, y, z) = (0, 0, 15/2)$. The maximal inhomogeneity factor at a nonbackground tissue was set to $1 + I$, whereas the minimal factor was set to $1 - I$. In this way I was a measure of the strength of the inhomogeneity. At $I = 0.125$ the local GM pd -echo mean varied from 927 in one part of the simulated image to 1191 in another. The CSF variation was between 1192 and 1533. Therefore, at $I = 0.125$ the inhomogeneity was so large that CSF had the same mean intensity in one part of the image as GM in the other.

III. RESULTS

We simulated MR images with different noise (N), inhomogeneity (I), and smoothing (S). These images were segmented and compared to the correct segmentation, i.e., the one the simulation was based on. The error was calculated as the ratio of misclassified voxels to nonbackground voxels within the entire $256 \times 256 \times 16$ volume.

First we analyzed the convergence of the algorithms at a typical noise level $N = 50$, inhomogeneity $I = 0.1$, and smoothing $S = 0.2$. In Fig. 2 no SB tissue was simulated. It shows that at six iterations an adequate convergence was obtained, only the ICM1 algorithm slightly improved with further iterations. Therefore, further comparisons were done

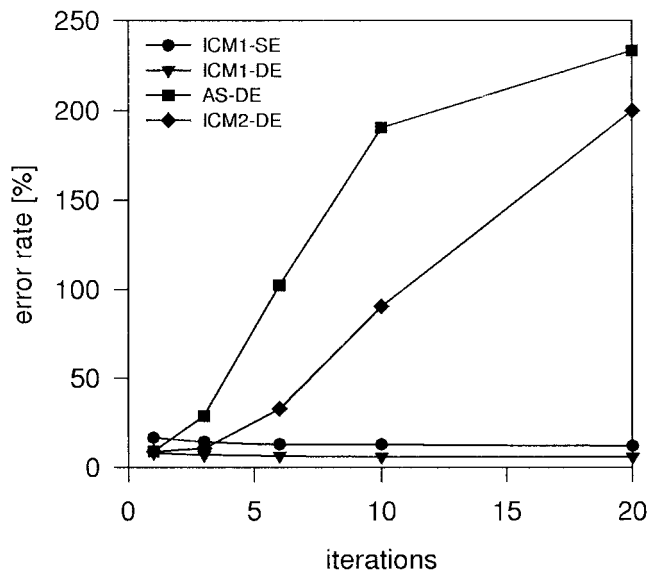


Fig. 3. Convergence of the segmentation algorithms with simulation of scalp-bone tissue. A noise level of $N = 50$, additional inhomogeneity $I = 0.1$ and smoothing $S = 0.2$ is simulated. The ICM2 and the *adaptive segmentation* algorithm fail to correct the inhomogeneity.

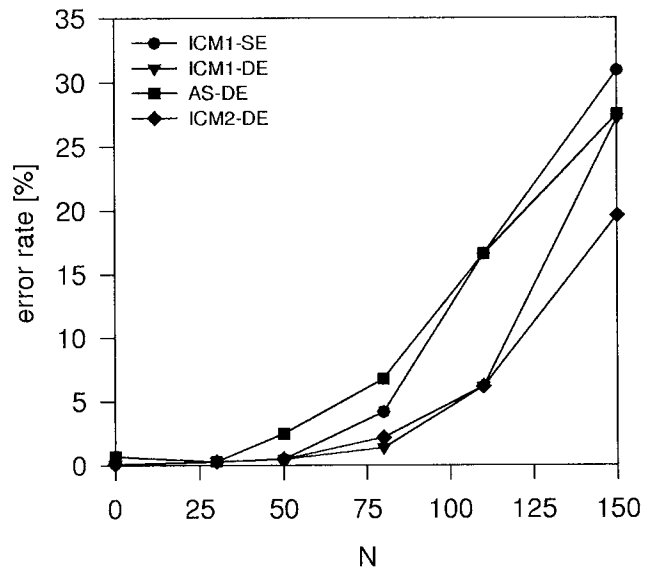


Fig. 4. Error rate versus noise level at $I = 0$, $S = 0$, and six iterations. The error rate increases exponentially with the Gaussian noise N . The ICM algorithms show no significant error up to $N = 50$ or $\text{CNR}_{\text{WM,GM}} = 4.7$.

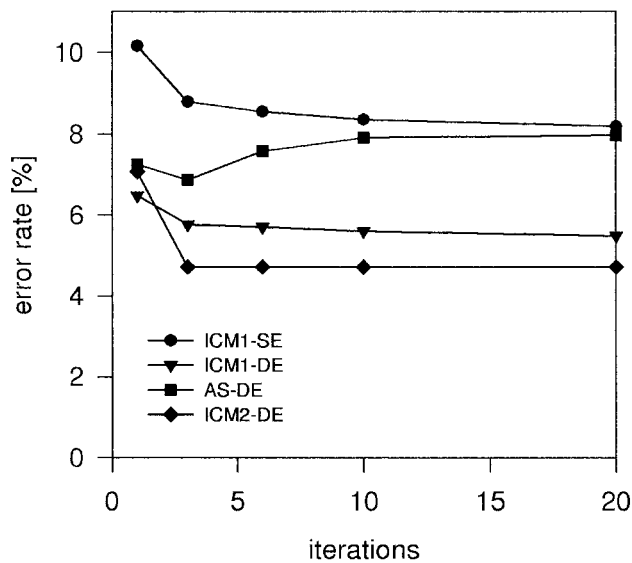


Fig. 2. Convergence of the segmentation algorithms without the simulation of the scalp-bone tissue. A noise level of $N = 50$, additional inhomogeneity $I = 0.1$, and smoothing $S = 0.2$ is simulated. Plotting the error rate versus number of iterations shows that at six iterations the convergence is adequate. (ICM = *iterated conditional mode*; AS = *adaptive segmentation*; SE = *single-echo*; DE = *double-echo*).

at six iterations. In Fig. 3, the SB tissue was simulated additionally. In this case, the ICM2 and the *adaptive segmentation* (AS) algorithms had huge errors of more than the number of brain voxels after convergence. The reason for this was that many background voxels were misclassified as SB. In practice, one can reduce these errors by pre- and postfiltering or by attaching a low *a priori* probability for SB. For an appropriate comparison we did not further simulate the SB structure.

The impact of noise was analyzed in Fig. 4. Up to $N = 50$, the ICM1 and ICM2 algorithms show low error rates of

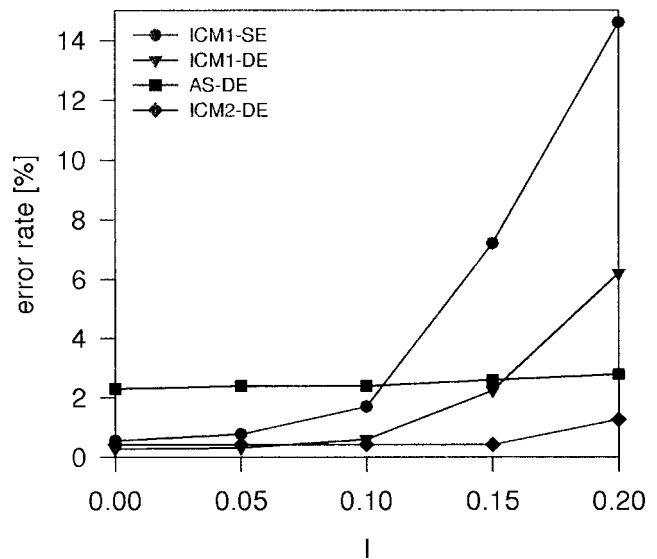


Fig. 5. Influence of inhomogeneity I at fixed $N = 50$, $S = 0$, and six iterations. Up to $I = 0.10$, all algorithms correct the inhomogeneity well. For higher inhomogeneities the ICM1 algorithm performs worse.

less than 0.5%, whereas the *adaptive segmentation* algorithm misclassified 2.3% of the voxels. For noise levels of $N = 30$ – 100 , even the single-echo ICM1 yielded better results than the *adaptive segmentation* algorithm. For all algorithms the error increased exponentially with the standard deviation N of the Gaussian noise.

Fig. 8 compares the SA to the ICM1 implementation for two different sets of parameters. The SB tissue was simulated again. It was demonstrated, that SA reduced the error, e.g., by 4% for $N = 50$, $I = 0.1$, and $S = 0.2$. This reduction was smaller if the SB tissue was not simulated.

The influence of inhomogeneity was simulated by different values of I ; see Fig. 5. For the whole range of tested inho-

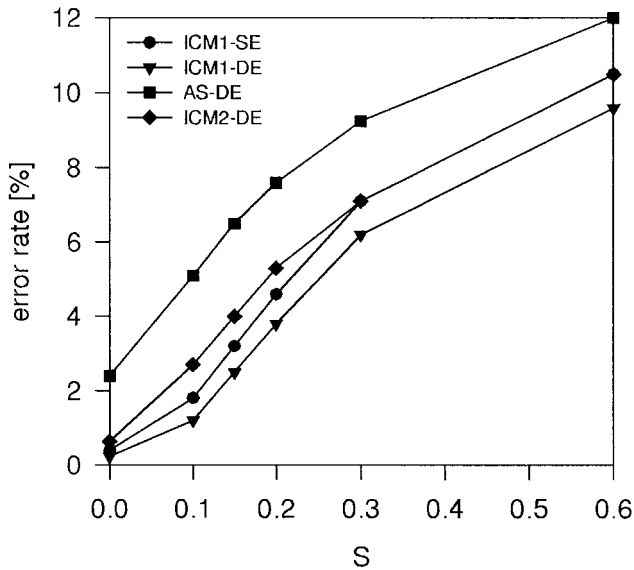


Fig. 6. Error rate versus smoothing at $N = 50$, $I = 0$, and six iterations. The information loss due to the smoothing S cannot be recovered and, thus, the error rate increased drastically.

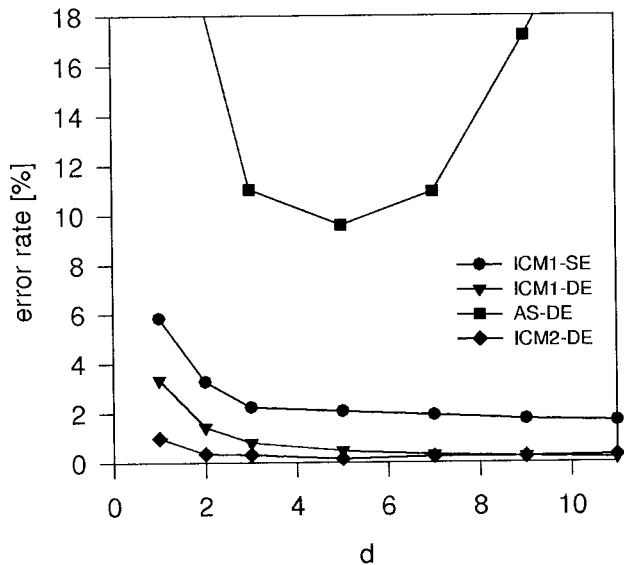


Fig. 7. Error rate versus GM thickness d . When using MRF's, fine structures can be resolved at a low error that is only somewhat enhanced for fine GM structures of one voxel thickness ($d = 1$).

mogeneities the *adaptive segmentation* algorithm yielded an almost constant error of 2.3%. Whereas at inhomogeneities $I < 0.10$ both ICM algorithms performed better than the *adaptive segmentation* algorithm, at inhomogeneities larger than $I = 0.10$, i.e., larger than the tissue contrast, the ICM1 algorithm performed worse.

Smoothing, controlled by the parameter S , wipes out fine details of the simulated MR image. Fig. 6 shows that all algorithms had difficulty to reconstruct these details, the error considerably grew with S . However, ICM1 performed slightly better with smoothing.

To analyze the influence of tissue thickness on the segmentation error we did not use the brain template, but an artificial

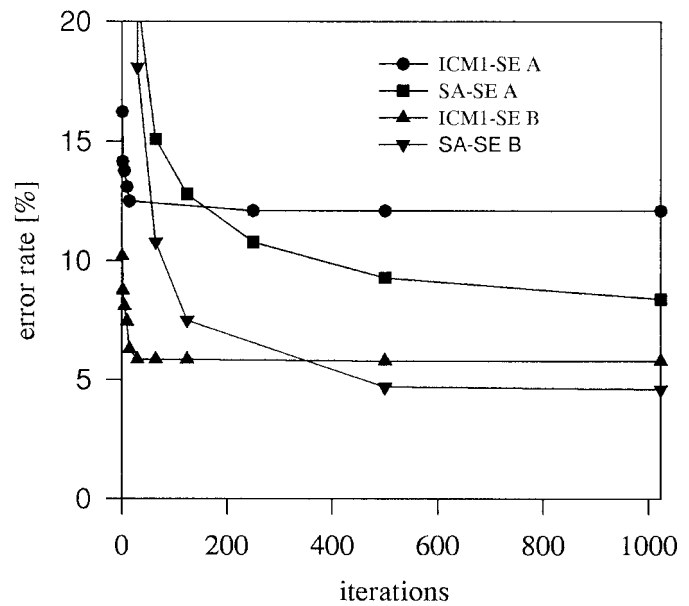


Fig. 8. SA compared to ICM1 for single-echo simulations with scalp-bone structure at two parameter sets (A: $N = 50$, $I = 0.1$, and $S = 0.2$; B: $N = 80$, $I = 0$, and $S = 0$).

image with a sinusoidal GM structure between WM and BG. The GM "gyrus" was d -voxels thick. The error was calculated as misclassified voxels in an 11-voxel neighborhood of the GM structure divided by the total number of GM voxels. At $N = 50$, $I = 0.1$, and $S = 0.2$ the ICM algorithms yielded good results (see Fig. 7). The error did not improve significantly for $d > 3$, whereas it was moderately higher for a fine GM structure of 1-voxel thickness ($d = 1$).

For real brain MR images, only a qualitative comparison of the algorithms was possible since the correct segmentation is unknown. Therefore, a quantitative comparison of the two algorithms did not reveal which one was wrong. Consequently, we show only the segmentations obtained and qualitatively point out some obvious failures. Fig. 9 displays one slice of a 3-D double-echo MR image. Fig. 10 shows the corresponding segmentations. For this slice the *adaptive segmentation* algorithm performed poorly at six iterations [Fig. 10(a)] misclassifying many WM voxels as GM. Therefore, the result for the first iteration is displayed, too [Fig. 10(b)]. At one iteration no inhomogeneity field was adapted and the segmentation was slightly worse in the occipital cortex [see bottom of Fig. 10(b)]. The MRF neighborhood correlations improved the segmentation of the nonbrain scalp-bone structure and reduced single-voxel misclassifications [Fig. 10(c) and (d)]. Based on the pd -weighted image, the single-echo ICM1 [Fig. 10(e)] and SA algorithms [Fig. 10(f)] were able to classify the MR brain correctly into GM, WM, CSF, and SB.

IV. DISCUSSION

The three new segmentation algorithms (ICM1, ICM2, and SA) presented differ in the method to handle the inhomogeneity field (ICM2 versus ICM1/SA) and to maximize the *a posteriori* probability (SA versus ICM1/ICM2).

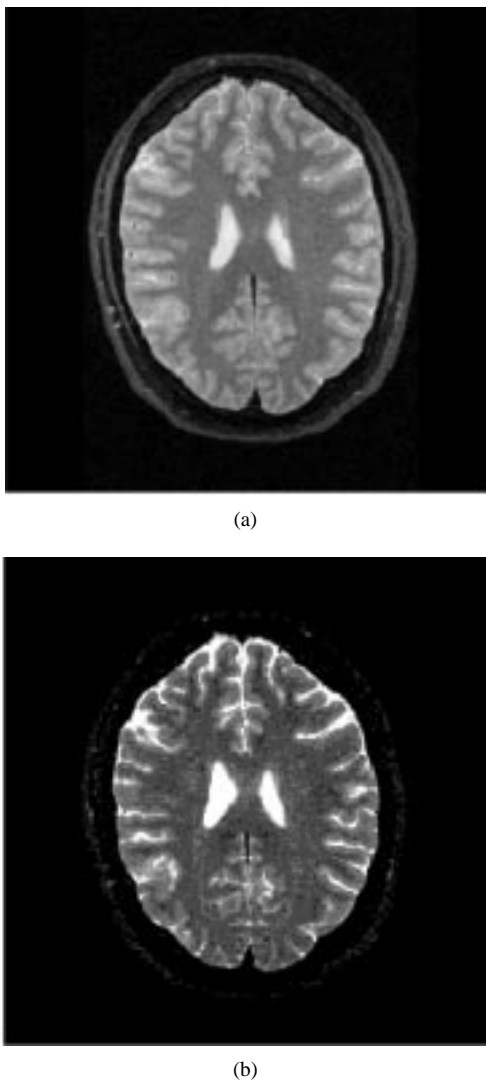


Fig. 9. Double-echo (a) pd - and (b) T_2 -weighted transversal slices of a 3-D brain MR image.

The ICM2 algorithm captures the intensity distribution and the inhomogeneity field in the same way as the *adaptive segmentation* [10] algorithm does, but ICM2 models neighborhood correlations as additional information to improve the segmentation. **The main improvement of the use of the MRF is better noise resistance.** This can be seen in Fig. 4, where ICM2 shows an error rate of less than 0.5% up to $CNR = 2.4$. **Smoothing and even inhomogeneities are better captured** (see Figs. 5 and 6).

The heuristic smoothing of the inhomogeneity field, as do the *adaptive segmentation* and ICM2 algorithms, does not take into account that small inhomogeneity corrections are more probable than large ones. Therefore, the ICM2 and the AS algorithms have the tendency to overestimate the inhomogeneity. Upon more iterations, the segmentation becomes worse; see Fig. 2 and especially Fig. 3 where the additional simulation of scalp-bone structure drastically increased the error. In practice, this has the consequence that the operator often has to try different numbers of iterations to yield the best segmentation result; see Fig. 10. Additionally, at fewer iterations the inhomogeneities at the outer slices are not

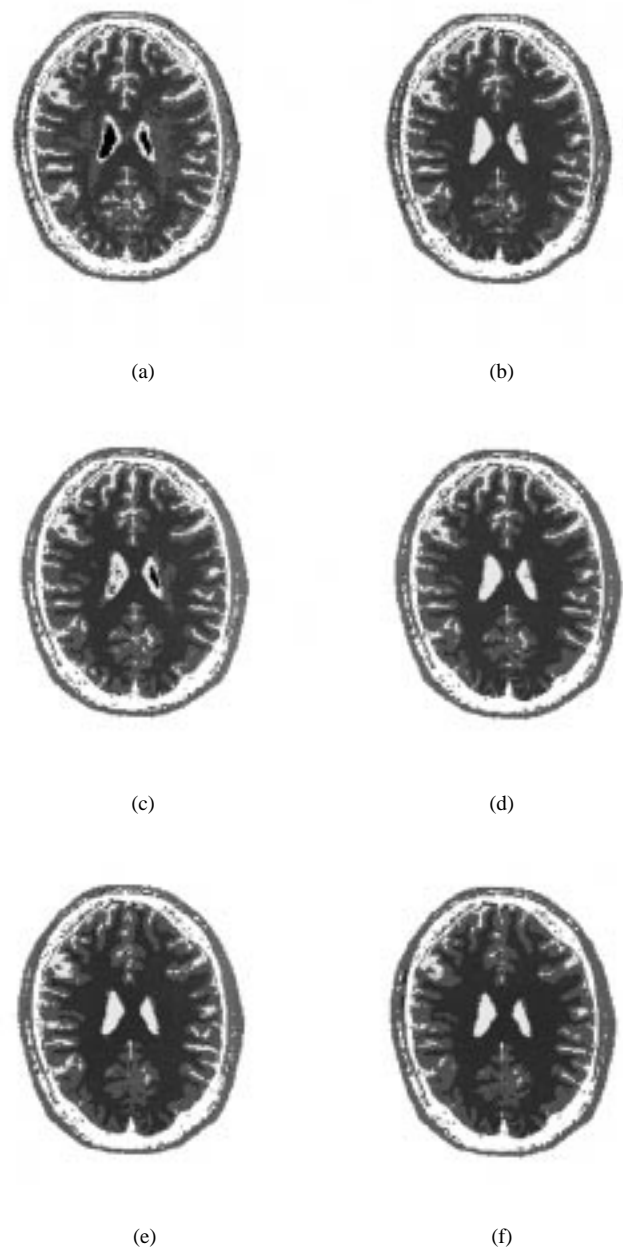


Fig. 10. Segmentations of the MR images from Fig. 9 (white: background, pink: scalp-bone, yellow: CSF, red: gray matter, blue: white matter, and brown: unclassified) as determined by the AS algorithm at (a) six and (b) one iteration, (c) double-echo ICM2, and (d) ICM1, and finally by the single-echo versions of (e) ICM1 and (f) SA.

captured correctly. The ICM1, as well as the SA algorithm, achieved further improvement by the use of an *a priori* Gibbs distribution for the inhomogeneity field. The ICM2 algorithm is superior to ICM1, see Fig. 4, only at inhomogeneities larger than the tissue contrast which were not observed in our MR images.

The large error of the *adaptive segmentation* algorithm for the artificial simulated sinusoidal image to test the GM thickness (Fig. 7) is based on the fact that at small GM structures the smoothing lowers the GM intensity. Therefore, many GM voxels were misclassified as WM when using the

adaptive segmentation algorithm. At large GM structures the inhomogeneity was not captured correctly, leading to a great error.

The SA algorithm (see Fig. 8) improves the segmentation, with the penalty being long computing times. One needs 1000 SA sweeps instead of ten ICM iterations to obtain accurate results.⁶ This shortcoming can be overcome by parallelizing the SA algorithm. Herein lies an important advantage of the local structure of the MRF: test and update of new configurations are possible at the same time and are only based on the information within a small neighborhood of a voxel.

V. CONCLUSION

We have presented a *simulated annealing* and an *iterated conditional modes* version of a new MRF 3-D segmentation algorithm. After the training of typical echo intensities and setting one MRF parameter (β) according to the expected inhomogeneity (see Section III), the algorithm fully automatically segments the entire 3-D MR volume, as well as different MR images that are acquired with the same MR sequence. For the first time nonparametric intensity distributions, neighborhood correlations, and inhomogeneities are combined in one segmentation algorithm. Volunteer MR images and detailed, quantitative simulations demonstrated that only the combination of these three features leads to an accurate and robust segmentation with respect to noise, inhomogeneity, and structure thickness. SA allows an additional improvement compared to ICM at the expense of long running times.

ACKNOWLEDGMENT

The authors would like to thank S. Posse, H. Herzog, and L. Tellmann for their advice and technical assistance. They would also like to thank J. Shah for assistance in preparing the manuscript.

REFERENCES

- [1] R. P. Woods, S. R. Cherry, and J. C. Mazziotta, "Rapid automated algorithm for aligning and reslicing PET images," *J. Comput. Assist. Tomogr.*, vol. 16, pp. 620–633, 1992.
- [2] K. J. Friston, J. Ashburner, C. D. Frith, J.-B. Poline, J. D. Heather, and R. S. J. Frackowiak, "Spatial registration and normalization of images," *Human Brain Mapping*, vol. 2, pp. 165–189, 1995.
- [3] H.-W. Müller-Gärtner, J. M. Links, J. L. Prince, R. N. Bryan, E. McVeigh, J. P. Leal, Ch. Davatzikos, and J. J. Frost, "Measurements of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects," *J. Cereb. Blood Flow Metab.*, vol. 12, pp. 571–583, 1992.
- [4] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, pp. 1277–1294, 1993.
- [5] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Pattern Anal. Machine Intell.*, vol. 12, pp. 629–639, 1990.
- [6] G. Gerig, O. Kübler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Trans. Med. Imag.*, vol. 11, pp. 221–232, 1992.
- [7] B. Jähne, *Digital Image Processing*, Berlin Heidelberg New York: Springer Verlag, 1993, pp. 200–208.
- [8] H. E. Cline, C. L. Dumoulin, H. R. Hart, W. E. Lorensen, and S. Ludke, "3-D reconstruction of the brain from magnetic resonance imaging using a connectivity algorithm," *Magn. Res. Imag.*, vol. 5, pp. 345–352, 1987.
- [9] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vision*, vol. 1, pp. 321–331, 1988.
- [10] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz, "Adaptive segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 15, pp. 429–442, 1996.
- [11] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.
- [12] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, 1984.
- [13] E. Ising, *Zeitschrift für Physik*, vol. 31, p. 253, 1925.
- [14] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machine," *J. Chem. Phys.*, vol. 21, pp. 1087–1091, 1953.
- [15] A. W. Sandvik, "NMR relaxation rates for the spin-1/2 Heisenberg chain," *Phys. Rev. B*, vol. 52, no. 14, pp. R9831–R9834, 1995.
- [16] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans. Signal Processing*, vol. 40, pp. 901–914, 1992.
- [17] J. Luo, C. W. Chen, and K. J. Parker, "On the application of Gibbs random field in image processing: From segmentation to enhancement," *J. Electron. Imag.*, vol. 4, pp. 187–198, 1995.
- [18] T. Taxt and A. Lundervold, "Multispectral analysis of the brain using magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 13, pp. 470–481, 1994.
- [19] A. Lundervold, T. Taxt, L. Ersland, T. Bech, and K.-J. Gjesdal, "Cerebrospinal fluid quantification and visualization using 3-D multispectral magnetic resonance images," presented at 9th Scand. Conf. Image Anal., Uppsala, Sweden, 1995.
- [20] J.-K. Fwu and P. M. Djuric, "EM algorithm for image segmentation initialized by a tree structure," *IEEE Trans. Imag. Processing*, vol. 6, pp. 349–352, 1997.
- [21] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B*, vol. 48, pp. 259–302, 1986.
- [22] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, pp. 39–55, 1987.
- [23] R. Kikinis, M. Shenton, F. A. Jolesz, G. Gerig, J. Martin, M. Anderson, D. Metcalf, C. Guttmann, R. W. McCarter, W. Lorensen, and H. Cline, "Quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging," *J. Magn. Res. Imag.*, pp. 619–629, 1992.
- [24] S. A. Kirkpatrick, C. D. Gellatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983; S. A. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," *J. Stat. Phys.* vol. 34, pp. 975–986, 1984.

⁶This corresponds to 2000 min instead of about 20 min for the segmentation of a $256 \times 256 \times 64$ MR volume on a SPARC20 workstation.