

This is a preprint of the following published material:  
Michael Wels, Yefeng Zheng, Martin Huber, Joachim Hornegger and Dorin Comaniciu, "A discriminative model-constrained EM approach to 3D MRI brain tissue classification and intensity non-uniformity correction", Phys. Med. Biol., Vol. 56, No. 11, June 2011.

© IOP Publishing Limited Bristol 2011

# A Discriminative Model-Constrained EM Approach to 3-D MRI Brain Tissue Classification and Intensity Non-Uniformity Correction

Michael Wels<sup>1,3</sup>, Yefeng Zheng<sup>2</sup>, Martin Huber<sup>3</sup>, Joachim Hornegger<sup>1</sup>, and Dorin Comaniciu<sup>2</sup>

<sup>1</sup>Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany

<sup>2</sup>Corporate Research and Technologies, Siemens Corporate Technology, 755 College Road East, Princeton, NJ 08540, USA

<sup>3</sup>Corporate Research and Technologies, Siemens Corporate Technology, Guenther-Scharowsky-Str. 1, 91058 Erlangen, Germany

E-mail: michael.wels@informatik.uni-erlangen.de

**Abstract.** We describe a fully automated method for tissue classification, which is the segmentation into cerebral gray matter (GM), cerebral white matter (WM), and cerebral spinal fluid (CSF), and intensity non-uniformity (INU) correction in brain magnetic resonance imaging (MRI) volumes. It combines supervised MRI modality-specific discriminative modeling and unsupervised statistical expectation maximization (EM) segmentation into an integrated Bayesian framework. While both the parametric observation models as well as the non-parametrically modeled INUs are estimated via EM during segmentation itself, a Markov random field (MRF) prior model regularizes segmentation and parameter estimation. Firstly, the regularization takes into account knowledge about spatial and appearance related homogeneity of segments in terms of pairwise clique potentials of adjacent voxels. Secondly and more importantly, patient-specific knowledge about the global spatial distribution of brain tissue is incorporated into the segmentation process via unary clique potentials. They are based on a strong discriminative model provided by a probabilistic boosting-tree (PBT) for classifying image voxels. It relies on surrounding context and alignment-based features derived from a probabilistic anatomical atlas. The context considered is encoded by 3-D Haar-like features of reduced INU sensitivity. Alignment is carried out fully automatically by means of an affine registration algorithm minimizing cross-correlation. Both types of features do not immediately use the observed intensities provided by the MRI modality but instead rely on specifically transformed features, which are less sensitive to MRI artifacts. Detailed quantitative evaluations on standard phantom scans and standard real world data show the accuracy and robustness of the proposed method. They also demonstrate relative superiority in comparison to other state-of-the-art approaches to this kind of computational task: our method achieves average Dice coefficients of  $0.93 \pm 0.03$  (WM) and  $0.90 \pm 0.05$  (GM) on simulated mono-spectral and  $0.94 \pm 0.02$  (WM) and  $0.92 \pm 0.04$  (GM) on simulated multi-spectral data from the BrainWeb repository. The scores are  $0.81 \pm 0.09$  (WM) and  $0.82 \pm 0.06$  (GM) and  $0.87 \pm 0.05$  (WM) and  $0.83 \pm 0.12$  (GM) for the two collections of real-world data sets—consisting of 20 and 18 volumes, respectively—provided by the Internet Brain Segmentation Repository.

*Keywords:* brain tissue classification, segmentation, intensity non-uniformity correction, magnet resonance imaging, Markov random fields, expectation-maximization, probabilistic boosting-tree, discriminative modeling Submitted to: *Phys. Med. Biol.*

## 1. Introduction

Several inquiries in medical diagnostics, therapy planning and monitoring, as well as in medical research, require highly accurate and reproducible brain tissue segmentation in 3-D magnetic resonance imaging (MRI) data. For instance, studies of neurodegenerative and psychiatric diseases often rely on quantitative measures obtained from MR scans that are segmented into the three common tissue types present in the human brain: cerebral gray matter (GM), cerebral white matter (WM), and cerebral spinal fluid (CSF). There is a need for fully automatic segmentation tools providing reproducible results in this particular context. Automatic tools face a challenging segmentation task due to the characteristic artifacts of the MRI modality, such as, among other things, intra-/inter-scan intensity non-uniformities (INU) (Wells et al., 1996; Jäger and Hornegger, 2009). The human brain's complexity in shape and natural intensity variations additionally complicate the segmentation task at hand. Once a sufficiently good segmentation is achieved it can also be used in enhancing the image quality, as intra-scan INUs can be easily estimated due to the knowledge of the tissue type and the associated image intensities to be observed at a specific spatial site (Wells et al., 1996).

Most approaches in the field of MRI brain tissue segmentation are based on Bayesian modeling, which typically involves providing a prior model and a generative observation model. With these models the most likely tissue class being responsible for the observed intensity values at a certain voxel can be inferred. Offline generated observation models (Held et al., 1997; Wells et al., 1996), that is, models generated from annotated training data, are usually very sensitive to MRI artifacts. (Han and Fischl, 2007) For this reason parametric models are typically estimated online, i.e., simultaneously with an associated segmentation maximizing an a posteriori probability distribution density by means of expectation maximization (EM) (Bricq et al., 2008; Scherrer et al., 2008, 2007; Ashburner and Friston, 2005; Pohl et al., 2002; Fischl et al., 2002; Zhang et al., 2001; van Leemput et al., 1999b; Kapur et al., 1998). Apart from EM, optimization methods comprise max-flow/min-cut computation (Song et al., 2006b,a), segmentation by weighted aggregation (Akselrod-Ballin et al., 2007), and finding the maximizer of the posterior marginals (MPM) in a maximum a posteriori (MAP) setting Marroquin et al. (2002). Also non-parametric (Akselrod-Ballin et al., 2007; Held et al., 1997) approaches for generating observation models within Bayesian frameworks and entirely learning-based (Akselrod-Ballin et al., 2006) approaches to brain tissue classification have been proposed.

Some of them (Akselrod-Ballin et al., 2006, 2007) do not take into account INUs and scanner-specific contrast characteristics present in the data sets used for model generation, which may result in model over-fitting and poor generalization capabilities.

Commonly used prior models comprise, next to the assumption of spatially uniform prior probabilities, spatial interdependencies among neighboring voxels through prior probabilities modeled as hidden Markov random fields (HMRF) (Scherrer et al., 2008, 2007; Marroquin et al., 2002; Zhang et al., 2001; Kapur et al., 1998; Held et al., 1997),

Hidden Markov Chains (HMC) (Bricq et al., 2008), or non-parametric adaptive Markov priors (Awate et al., 2006). They are sometimes combined with prior probabilities derived from probabilistic or anatomical atlases (Awate et al., 2006; Marroquin et al., 2002; Pohl et al., 2002; van Leemput et al., 1999b) or replaced by them (Bazin and Pham, 2008; Ashburner and Friston, 2005) that can also be integrated into the overall MRF-based formulation as external field energies (Scherrer et al., 2008; Pohl et al., 2002). The same holds for prior knowledge encoded by fuzzy localization maps (Scherrer et al., 2007) that can also be integrated into the overall framework via external fields. Alignment of the atlas can be achieved either by rigid (Bazin and Pham, 2008; van Leemput et al., 1999b), affine (Scherrer et al., 2008; Akselrod-Ballin et al., 2007; Awate et al., 2006), or non-rigid (Bricq et al., 2008; Marroquin et al., 2002; Pohl et al., 2002) registration algorithms, either before optimization or simultaneously (Bazin and Pham, 2008; Ashburner and Friston, 2005). Bazin and Pham (Bazin and Pham, 2008) additionally incorporate prior knowledge obtained from a topological atlas into a fuzzy classification technique for topology preservation. Cuadra et al. (2005) compare and validate different statistical non-supervised brain tissue classification techniques in MRI volumes.

While some of the papers mentioned above address further segmentation of cerebral gray matter into individual structures (Scherrer et al., 2008, 2007; Akselrod-Ballin et al., 2007; Bazin and Pham, 2008), which is beyond the scope of this paper, only some of them additionally address INU correction (Bricq et al., 2008; Song et al., 2006a,b; Ashburner and Friston, 2005; Zhang et al., 2001; van Leemput et al., 1999b; Wells et al., 1996; Held et al., 1997). INUs are usually modeled as multiplicative noise corrupting the images in the intensity domain and as additive noise in the log-domain. They can be described either non-parametrically as bias or gain fields in the literal sense (Zhang et al., 2001; Held et al., 1997; Wells et al., 1996) or parametrically by polynomial basis functions (Bricq et al., 2008; van Leemput et al., 1999a), by means of cubic B-splines (Song et al., 2006b,a) or through the exponential of a linear combination of low frequency basis functions (Ashburner and Friston, 2005). Even MRF modeling can be applied to model an a priori probability distribution on a non-parametrically represented gain field (Held et al., 1997).

Conceptually, our approach aligns with the mentioned EM-based approaches using MRF priors and aligned probabilistic atlases. Our method however makes use of more general prior knowledge in terms of a strong machine learning-based discriminative model initializing and continually constraining the segmentation process. We present an extended Hidden Markov Random Field Expectation Maximization (HMRF-EM) approach with simultaneous INU correction. It is, in contrast to the method of Zhang et al. (2001)<sup>†</sup>, consistently formulated to work on multi-spectral 3-D brain MRI data. Further, we demonstrate a mathematically sound integration of prior knowledge encoded by a strong discriminative model into the statistical framework. The learning-based

<sup>†</sup> Although not detailed in the original publication a multi-spectral implementation of Zhang et al.'s method (Zhang et al., 2001) already exists and can be downloaded from [www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl).

component, that is, a probabilistic boosting-tree (PBT) (Tu, 2005), providing the discriminative model exclusively relies on features of reduced sensitivity to INUs and therefore makes this approach MRI modality-specific.

Exhaustive quantitative evaluations of our method on publicly available simulated and real world MRI scans are performed and compared to other state-of-the-art approaches (Scherrer et al., 2008; Bazin and Pham, 2008; Bricq et al., 2008; Scherrer et al., 2007; Akselrod-Ballin et al., 2007, 2006; Awate et al., 2006; Ashburner and Friston, 2005; Marroquin et al., 2002; Zhang et al., 2001; van Leemput et al., 1999b). While other methods may reach particular high values on a particular database we present comparable and mostly better results in terms of segmentation accuracy on a variety of benchmarking databases from different sources. This demonstrates the increased robustness of our approach.

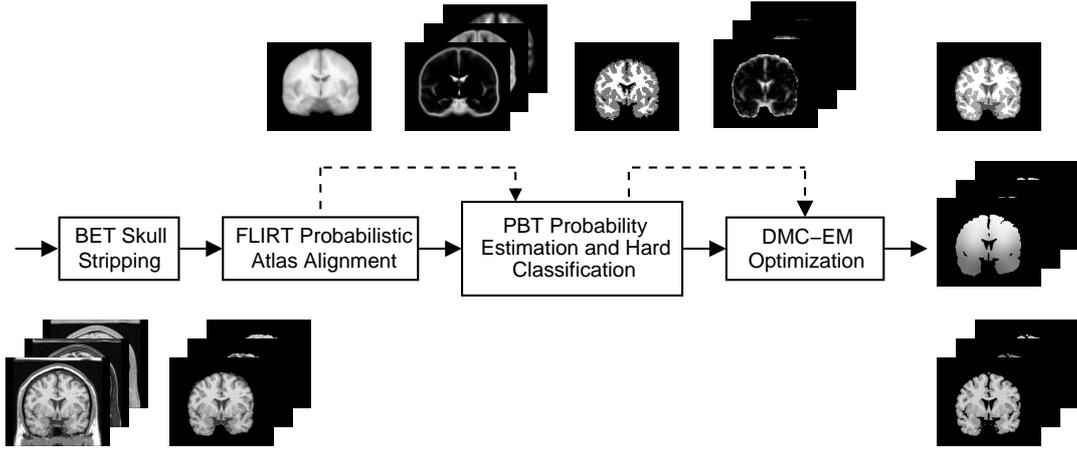
## 2. Method

### 2.1. Pre-Processing and Processing Pipeline

Our method consists of four steps: first, the whole brain is extracted from its surroundings with the Brain Extraction Tool (BET) (Smith, 2002) working on the T1-weighted pulse sequence.† Then, an initial spatially variant prior of the brain soft tissue on different tissue classes is obtained by means of a strong modality specific discriminative model, that is to say, a PBT probability estimator. This also gives an initial segmentation of the brain soft tissue. Subsequently, the final segmentation and the multi-spectral INU fields are estimated via an extended HMRF-EM approach that operates on multi-spectral input data. We will refer to our method as the discriminative model-constrained HMRF-EM approach (DMC-EM). The whole processing pipeline is depicted in Fig. 1.

In the following we will focus our presentation on the last step of this four-step bottom-up processing pipeline as it theoretically links PBT probability estimation and HMRF-EM optimization. The DMC-EM optimization subsumes the main contributions of our work. While the first two steps, BET skull stripping and FLIRT probabilistic atlas alignment, can be considered pre-processing steps the third step, PBT probability estimation and hard classification, serves both as an initialization step as well as a pre-computation step. It initializes the subsequent EM optimization procedure and provides probability estimates that are later on repeatedly used during optimization.

† As BET skull stripping fails on some of the data sets we use for evaluation we extended the original preprocessing tool BET. We introduced thresholding for background exclusion, morphological operations and connected component analysis to generate initializations (center and radius of initial sphere) for the BET main procedure that are closer to the intra-cranial surface to be computed.



**Figure 1.** The processing pipeline of the proposed DMC-EM method for multi-spectral brain tissue segmentation and INU correction.

### 2.2. DMC-EM Brain Tissue Segmentation

Image or volume segmentation by means of the DMC-EM approach, which extends the HMRF-EM approach of Zhang et al. (2001), is closely related to learning finite Gaussian mixtures (FGM) via the EM algorithm. For both cases let  $\mathcal{S} = \{1, 2, \dots, N\}$ ,  $N \in \mathbb{N}$ , be a set of indices to image voxels. At each index  $s \in \mathcal{S}$  there are two random variables  $Y_s$  and  $\mathbf{X}_s$  that take discrete values  $y_s \in \mathcal{Y} = \{1, \dots, K\}$ ,  $K \in \mathbb{N}$ , and  $\mathbf{x}_s \in \mathcal{X} = \{1, \dots, 2^d\}^L$ . The former,  $Y_s$ , denotes the hidden class label, that is, the underlying tissue class, at voxel  $s$ , whereas the latter,  $\mathbf{X}_s$ , states the vector of observed intensity values taken from the  $L \in \mathbb{N}$  aligned input pulse sequences each having a bit depth of  $d \in \mathbb{N}$ . The observable intensities at every voxel  $s$  are assumed to be causally linked to the underlying class labels by parameterized Gaussian distribution densities  $p(\mathbf{x}_s | y_s = k) = N(\mathbf{x}_s; \boldsymbol{\theta}_k)$  with class specific parameters  $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ ,  $\boldsymbol{\mu}_k \in \mathbb{R}^L$ ,  $\boldsymbol{\Sigma}_k \in \mathbb{R}^{L \times L}$  and symmetric positive-definite. Starting from initial values for those parameters and some prior probabilities  $p^{(0)}(k)$  for the occurrence of each class label a proper statistical model in terms of prior probabilities  $p(k)$ ,  $k \in \mathcal{Y}$ , and parameters  $\Theta = (\boldsymbol{\theta}_k)_{k \in \mathcal{Y}}$  can be estimated by means of EM iteratively in an unsupervised manner.

In contrast to the FGM model that considers every voxel’s classification isolated from its local neighborhood the DMC-EM model assumes external influences and spatial interdependencies among neighboring voxels. Both can be incorporated into the existing model by describing the family  $\mathbf{Y} = (Y_s)_{s \in \mathcal{S}}$  of unknown class labels as an MRF. According to Li (2001), within an MRF every voxel at index  $s$  is associated with a subset  $\mathcal{N}_s \subseteq \mathcal{S} \setminus \{s\}$  of neighboring indices having the properties  $s \notin \mathcal{N}_s$  and  $s \in \mathcal{N}_t \Leftrightarrow t \in \mathcal{N}_s$  for all  $s, t \in \mathcal{S}$ .

The graph  $G = (V, E)$  with vertices  $V = \{v_s | s \in \mathcal{S}\}$  and edges  $E = \{(v_s, v_t) | s \in \mathcal{S}, t \in \mathcal{N}_s\}$  associated with an MRF contains multiple sets of cliques, which are sets of complete sub-graphs,  $\mathcal{C}_n$  denoting all the sets of vertices’ indices within cliques of size

$n \in \{1, \dots, |V|\}$ .

Under these circumstances, according to the Hammersley-Clifford theorem, the joint probability density function (PDF)  $p(\mathbf{y})$  can equivalently be described by a Gibbs distribution  $p(\mathbf{y}) = \frac{1}{Z} \exp(-U(\mathbf{y}))$ . Here  $U(\mathbf{y}) = \sum_n \sum_{c_n \in \mathcal{C}_n} V_{c_n}(\mathbf{y})$  denotes the energy function, which is a sum of clique potentials  $V_{c_n}$ , and  $Z = \sum_{\mathbf{y}} \exp(-U(\mathbf{y}))$  denotes the partition function, which is a normalization constant.

In contrast to Zhang et al. (2001) our model considers both unary ( $n = 1$ ) as well as pairwise ( $n = 2$ ) clique potentials as we want to introduce an MRF prior that constrains segmentation by an external field, provided by a strong discriminative model, and by mutual spatial dependencies among pairs of neighboring voxels. In this case the energy function can be stated as

$$U(\mathbf{y}) = \sum_{s \in \mathcal{S}} \left( V_s(y_s) + \frac{\beta}{2} \sum_{t \in \mathcal{N}_s} V_{st}(y_s, y_t) \right). \quad (1)$$

Thus, by applying Bayes' rule and by marginalizing over the possible class labels, we have

$$\begin{aligned} p(y_s | \mathbf{y}_{\mathcal{N}_s}) &= p(y_s | (y_t)_{t \in \mathcal{S} \setminus \{s\}}) \\ &= \frac{p(y_s, (y_t)_{t \in \mathcal{S} \setminus \{s\}})}{\sum_{k \in \mathcal{Y}} p(y_s = k, (y_t)_{t \in \mathcal{S} \setminus \{s\}})} \\ &= \frac{\exp(-V_s(y_s) - \sum_{t \in \mathcal{N}_s} V_{st}(y_s, y_t))}{\sum_{k \in \mathcal{Y}} \exp(-V_s(y_s = k) - \sum_{t \in \mathcal{N}_s} V_{st}(y_s = k, y_t))} \end{aligned} \quad (2)$$

with the labels  $\mathbf{y}_{\mathcal{N}_s}$  understood as observable evidence.

Due to the fact that Equation (2) can be formulated dependent on unary and pairwise clique potentials it is possible to introduce prior knowledge into the classification process. In order to make a strong discriminative model constrain expectation maximization we will later define unary clique potentials based on tissue class probability estimations from PBT classifiers. With regards to the pair-wise clique potentials, which are defined on fully labeled data, the best segmentation  $\arg \max_{\mathbf{y}} p(\mathbf{y} | \mathbf{x}; \Theta^{(i-1)})$  that is needed to properly evaluate  $V_{st}(y_s, y_t)$  in iteration  $i$  is not available during iterative expectation maximization. This means, in accordance with Zhang et al. (2001), a currently best segmentation using the MAP

$$\mathbf{y}^* = \arg \max_{\mathbf{y}} p(\mathbf{y} | \mathbf{x}; \Theta^{(i-1)}) \quad (3)$$

where

$$\begin{aligned} p(\mathbf{y} | \mathbf{x}; \Theta^{(i-1)}) &= \frac{p(\mathbf{x} | \mathbf{y}; \Theta^{(i-1)}) p(\mathbf{y})}{p(\mathbf{x})} \\ &= \frac{1}{Z} \prod_s \frac{p(\mathbf{x}_s | y_s; \theta^{(i-1)}) \cdot \exp(-V_s(y_s) - \sum_{t \in \mathcal{N}_s} V_{st}(y_s, y_t))}{p(\mathbf{x}_s)} \\ &\propto \prod_s N(\mathbf{x}_s | \theta^{(i-1)}) \cdot \exp(-V_s(y_s) - \sum_{t \in \mathcal{N}_s} V_{st}(y_s, y_t)) \end{aligned} \quad (4)$$

has to be found in every iteration  $i$  of the overall expectation maximization procedure to form the complete dataset where we assume the intensities  $\mathbf{X}_s$  to be i.i.d. In our method forming the complete dataset is done by iterated conditional modes (ICM) as proposed by Besag (1986) and adapted for brain tissue segmentation by Zhang et al. (2001) and Held et al. (1997).

Once a sufficiently good approximation of the currently best segmentation is computed the parameters of the intensity model can be updated by

$$\boldsymbol{\mu}_k^{(t)} = \frac{\sum_{s \in \mathcal{S}} p(y_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \boldsymbol{\theta}_k^{(t-1)}) \mathbf{x}_s}{\sum_{s \in \mathcal{S}} p(y_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \boldsymbol{\theta}_k^{(t-1)})} \quad (5)$$

and

$$\boldsymbol{\Sigma}_k^{(t)} = \frac{\sum_{s \in \mathcal{S}} p(y_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \boldsymbol{\theta}_k^{(t-1)}) (\mathbf{x}_s - \boldsymbol{\mu}_k^{(t)}) (\mathbf{x}_s - \boldsymbol{\mu}_k^{(t)})^T}{\sum_{s \in \mathcal{S}} p(y_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \boldsymbol{\theta}_k^{(t-1)})}. \quad (6)$$

The complete DMC-EM procedure can be summarized as follows: starting from initial values  $\mathbf{y}^{(0)}$  and  $\boldsymbol{\Theta}^{(0)}$ , in each iteration  $i$  the current segmentation  $\mathbf{y}^{(i)}$  is approximated and used to compute the posterior probabilities  $p(y_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \boldsymbol{\theta}_k^{(i-1)})$  for each voxel  $s \in \mathcal{S}$ . Subsequently, the parameters  $\boldsymbol{\Theta}^{(i)}$  are updated.

At this current point our method equals the HMRF-EM approach (Zhang et al., 2001). In the following sections we will derive unary and pairwise clique potentials that take into account probability estimations from a strong MRI modality-specific discriminative model, i.e. a PBT, and spatial coherence in terms of observed intensities and current classification labels, respectively. This combination of discriminative modeling via the PBT algorithm and MAP tissue classification via the EM algorithm through the formulation of appropriate unary clique potentials is what we consider the major contribution of our work. It is also what makes the difference between our DMC-EM algorithm and the HMRF-EM algorithm (Zhang et al., 2001). Further we will extend the approach from its theoretical point of view in order to simultaneously estimate multi-spectral INUs similarly to Zhang et al. (2001) who presented a mono-spectral extension of their method for this purpose.

### 2.3. MRI INU Estimation

As shown by Zhang et al. (2001) the HMRF-EM as well as our DMC-EM method can be extended to simultaneously estimate the INU field according to the method of Wells et al. (1996). The INUs are modeled by a multiplicative gain field  $\mathbf{g} = (g_s)_{s \in \mathcal{S}}$  that disturbs the true intensities  $\mathbf{i}^* = (i_s^*)_{s \in \mathcal{S}}$ . That is

$$i_s = i_s^* \cdot g_s \quad (7)$$

for one of the MRI channels at voxel  $s \in \mathcal{S}$  where  $\mathbf{i} = (i_s)_{s \in \mathcal{S}}$  are the disturbed and observed intensities. Although less appropriate for modeling INUs caused by induced currents and inhomogeneous excitation within the acquisition device the multiplicative model adequately describes the inhomogeneous sensitivity of the reception coil. (Sled et

al., 1998) After logarithmic transformation of intensities the gain field can be treated as an additive bias field  $\mathbf{b} = (b_s)_{s \in \mathcal{S}}$  and

$$x_s = x_s^* + b_s, \quad (8)$$

where  $x_s = \log(i_s)$ ,  $x_s^* = \log(i_s^*)$ , and  $b_s = \log(g_s)$ . In the case of multi-spectral images we have

$$\mathbf{x}_s = \mathbf{x}_s^* + \mathbf{b}_s. \quad (9)$$

For DMC-EM this means that the class-conditional probabilities are no longer only dependent on the parameters  $\Theta$  of the Gaussian distributions but also of the bias field  $\mathbf{b}$ , that is,

$$p(\mathbf{x}_s | y_s, \mathbf{b}_s) = N(\mathbf{x}_s - \mathbf{b}_s; \theta_k). \quad (10)$$

Following Wells et al. (1996) the joint probability of intensities and tissue class conditioned on the bias field can be stated as

$$p(\mathbf{x}_s, y_s | \mathbf{b}_s) = p(\mathbf{x}_s | y_s, \mathbf{b}_s) p(y_s). \quad (11)$$

Marginalization over  $\mathcal{Y}$  yields

$$p(\mathbf{x}_s | \mathbf{b}_s) = \sum_{k \in \mathcal{Y}} p(\mathbf{x}_s | y_s = k, \mathbf{b}_s) p(y_s = k), \quad (12)$$

which is a class-independent PDF consisting of a mixture of Gaussian populations. By applying the MAP principle to the posterior probability of the bias field, which can be derived from Equation (12), an initial expression for the bias field estimate can be formulated. Then, a zero-gradient condition with respect to  $\mathbf{b}$  leads to a non-linear bias field estimator fulfilling a necessary condition for optimality:

$$\mathbf{b} = \left[ \overline{\Sigma}^{-1} + \Sigma_{\mathbf{b}}^{-1} \right]^{-1} \bar{\mathbf{r}}, \quad (13)$$

where  $\bar{\mathbf{r}} = (\bar{r}_s)_{s \in \mathcal{S}}$  are the mean residuals

$$\bar{r}_s = \sum_{k \in \mathcal{Y}} p(y_s = k | \mathbf{x}_s, \mathbf{b}_s) (\mathbf{x}_s - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x}_s - \boldsymbol{\mu}_k) \quad (14)$$

and  $\overline{\Sigma}^{-1} = (\overline{\Sigma}^{-1}_s)_{s \in \mathcal{S}}$  are the mean inverse covariances with entries

$$\overline{\Sigma}^{-1}_s = \sum_{k \in \mathcal{Y}} p(y_s = k | \mathbf{x}_s, \mathbf{b}_s) \Sigma_k^{-1} \quad (15)$$

written down as a family of  $L \times L$  matrices. Please refer to Wells et al. (1996) for a detailed description of the mathematical assumptions and derivation steps involved.

Using an approximation instead of the optimal estimator the bias field at every voxel  $s \in \mathcal{S}$  is given by

$$\mathbf{b}_s = \left( \mathbf{F}[\overline{\Sigma}^{-1}] \right)_s^{-1} \cdot (\mathbf{F}[\bar{\mathbf{r}}])_s \quad (16)$$

where  $\mathbf{F}$  is a low-pass filter working component-wise on the matrix- or vector-valued, in our case, volumes  $\bar{\mathbf{r}}$  and  $\overline{\Sigma}^{-1}$  (Wells, 1986).

---

**Algorithm 1:** DMC-EM algorithm
 

---

**Input:** (Multi-spectral) MRI volume  $\mathbf{x}$ , parameters  $\Theta^{(0)} = (\theta_k^{(0)} = (\mu_k^{(0)}, \Sigma_k^{(0)}))_{k \in \mathcal{Y}}$ , initial segmentation  $\mathbf{y}^{(0)}$   
**Output:** Parameters  $\Theta^{(t)}$ , segmentation  $\mathbf{y}^{(t)}$ , and bias field  $\mathbf{b}^{(t)}$

```

begin
     $t \leftarrow 0$ ;
     $\Theta^{(t)} \leftarrow \Theta^{(0)}$ ;
     $\mathbf{b}^{(t)} \leftarrow \mathbf{0}$ ;
    repeat
         $t \leftarrow t + 1$ ;

        // 1. Estimate the class labels by MRF-MAP estimation (see Equation (3))
         $\mathbf{y}^{(t)} \leftarrow \arg \max_{\mathbf{y}} p(\mathbf{y} | \mathbf{x}; \Theta^{(t-1)}, \mathbf{b})$ ;

        // 2. Calculate the posterior distributions for the corrected and non-corrected
        intensities
        forall voxels  $s$  do
            forall class labels  $k$  do
                 $p(\mathbf{y}_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \theta_k^{(t-1)}, \mathbf{b}_s) \leftarrow \frac{N(\mathbf{x}_s - \mathbf{b}_s; \theta_k^{(t-1)}) p(\mathbf{y}_s = k | \mathbf{y}_{\mathcal{N}_s})}{\sum_{l \in \mathcal{Y}} N(\mathbf{x}_s - \mathbf{b}_s; \theta_l^{(t-1)}) p(\mathbf{y}_s = l | \mathbf{y}_{\mathcal{N}_s})}$ ;
                 $p(\mathbf{y}_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \theta_k^{(t-1)}) \leftarrow \frac{N(\mathbf{x}_s; \theta_k^{(t-1)}) p(\mathbf{y}_s = k | \mathbf{y}_{\mathcal{N}_s})}{\sum_{l \in \mathcal{Y}} N(\mathbf{x}_s; \theta_l^{(t-1)}) p(\mathbf{y}_s = l | \mathbf{y}_{\mathcal{N}_s})}$ ;
            end
        end

        // 3. Update the parameters of the observation model (see Equations (5) and (6))
         $\mu_k^{(t)} \leftarrow \frac{\sum_{s \in \mathcal{S}} p(\mathbf{y}_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \theta_k^{(t-1)}) \mathbf{x}_s}{\sum_{s \in \mathcal{S}} p(\mathbf{y}_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \theta_k^{(t-1)})}$ ;
         $\Sigma_k^{(t)} \leftarrow \frac{\sum_{s \in \mathcal{S}} p(\mathbf{y}_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \theta_k^{(t-1)}) (\mathbf{x}_s - \mu_k^{(t)}) (\mathbf{x}_s - \mu_k^{(t)})^T}{\sum_{s \in \mathcal{S}} p(\mathbf{y}_s = k | \mathbf{x}_s, \mathbf{y}_{\mathcal{N}_s}; \theta_k^{(t-1)})}$ ;

        // 4. Estimate the bias field (see Equation (16))
        forall voxels  $s$  do
             $\mathbf{b}_s = [\mathbf{F}(\overline{\Sigma^{-1}})]_s^{-1} \cdot [\mathbf{F}\bar{\mathbf{r}}]_s$ ;
        end
    until  $t = T$ ;
end
    
```

---

The DMC-EM algorithm for simultaneous brain tissue segmentation and INU correction of multi-spectral data with a predefined number  $T$  of iterations can be stated as depicted in Algorithm 1.

As pointed out by Zhang et al. (2001) and originally discovered by Guillemaud and Brady (1997) the method of Wells et al. (1996), which serves as the base of our INU correction system, does not adequately work on image segments whose actual intensity distribution is not Gaussian. Such a tissue class usually has a large variance, which prevents the mean from being representative. In our system this is the case for the CSF tissue class that does not only include the ventricular system inside but also around the brain. Especially at the outer bounds of the automatically generated brain mask,

this class may include several other non-brain structures introducing intensity values different from the ones expected from true CSF, which correspondingly increases intra-class variance.

Inspired by Wells et al. (1996), where everything but GM and WM is excluded both from the INU estimation as well as from the segmentation, we therefore estimate the bias field only on the current GM and WM segments assuming the current CSF segment to be part of the background. However, in contrast to Wells et al. (1996), we do address CSF segmentation, together with GM and WM segmentation, during iterative tissue classification.

In the following we will derive appropriate higher dimensional feature vectors  $\mathbf{z}$  for PBT training and PBT probability estimation. In order to keep the discriminative models MRI modality-specific we have to make sure that the features  $\mathbf{z}$  used are not sensitive to inter- and intra-scan INUs as probability estimation will be performed on the non-corrected input data. We will therefore rely on 3-D Haar-like (Tu et al., 2006a) features of reduced INU sensitivity and probabilistic atlas-based whole brain anatomy features. Both types of features are the result of specific transformations and do not immediately use the observed intensities provided by the MRI modality.

#### 2.4. MRI Modality-Specific Discriminative Model-Based Unary Clique Potentials

*2.4.1. Probabilistic Boosting-Tree* The discriminative classifier PBT (Tu, 2005) recursively groups boosted ensembles of weak classifiers to a tree structure during learning from expert annotated data. (see Appendix Appendix A) For every tissue class we learn a voxel-wise discriminative PBT probability estimator relying on higher dimensional feature vectors  $\mathbf{z}_s$ , which are derived from the surrounding 3-D context of a voxel of interest  $s$ . We use the class-wise probability estimates  $\tilde{p}^k(+1|\mathbf{z}_s)$ ,  $k \in \{1, \dots, K\}$ , for the  $K$  tissue classes to define the unary clique potentials

$$V_s(y_s = k) = -\log \tilde{p}^k(+1|\mathbf{z}_s) \quad (17)$$

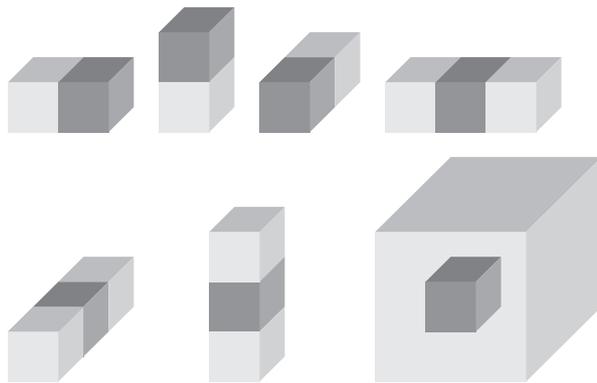
used in our system.

*2.4.2. Haar-like Features of Reduced INU Sensitivity* In the case of a 1-D signal  $f(t)$ ,  $t \in \mathbb{R}$ , as well as for any higher dimensional signal Haar-like filters can be interpreted as non-normalized child wavelets  $\psi(\frac{t-\tau}{\alpha})$  of the classical Haar mother wavelet

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leq t < \frac{1}{2}, \\ -1 & \text{if } \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \quad (18)$$

As normalization does not affect linear independence the family of non-normalized child wavelets spans the same infinite-dimensional vector space as their normalized counterparts. Feature responses, which are comparable to wavelet coefficients, typically are only computed for discrete  $-\tau_{max} \leq \tau \leq +\tau_{max}$  and  $0 < \alpha \leq \alpha_{max}$ . This equals projecting a transformed signal to a finite-dimensional subspace where only certain

position and frequency characteristics are taken into account. As seen above, MRI inter-scan intensity inhomogeneities can be modeled as gain fields (Wells et al., 1996) where a spatially varying factor multiplicatively disturbs the observed intensities  $i_s$  at voxel  $s \in \mathcal{S}$ . After logarithmic transformation it can be seen as an additive bias field of low frequency and zero mean. The parameter  $\alpha_{max}$  can be chosen sufficiently low such that low frequencies of that kind are attenuated and do not significantly affect the signal’s projection onto the subspace. The obtained coefficients are therefore of reduced bias field sensitivity when considering the log-transformed signal and of reduced gain field, that is, INU field, sensitivity in the original domain.



**Figure 2.** The 3-D Haar-like feature prototypes used in the DMC-EM algorithm’s discriminative model.

This is perfectly accompanied by the intuition that small neighboring areas should have an almost identical additive bias in the log-domain, which disappears after subtraction when computing the Haar-like features.

Fig. 2 depicts the 3-D Haar-like feature prototypes used in our system. The associated features are computed at different anisotropic scales of the prototypes with a fixed offset centered at the voxel of interest. For every feature prototype the average of the log-transformed intensities within the white cuboids is subtracted from the average of the log-transformed intensities within the black cuboids.

*2.4.3. Probabilistic Atlas-Based Whole Brain Anatomy Features* The second category of features contributing to the feature vectors  $\mathbf{z}_s$  for PBT training and probability estimation encode the voxel’s probability to be either part of the CSF, the GM, or the WM. They are taken from a probabilistic anatomical atlas (Rex et al., 2003), which is affinely registered (Scherrer et al., 2008; Akselrod-Ballin et al., 2007; Awate et al., 2006) with the current data set by means of the publicly available registration software FLIRT (Jenkinson and Smith, 2001). The objective function for the registration step is based on the correlation ratio metric, which is suited for inter-modality registration purposes by design. It ensures robustness of the registration procedure in the case of inter- and intra-scan INUs. The choice in favor for a 12-parameter affine registration

algorithm is motivated by the trade-off between maximum flexibility and computational demand of the underlying registration procedure. Non-rigid registration algorithms may lead to more discriminative atlas-based features. (Bricq et al., 2008; Ashburner and Friston, 2005; Marroquin et al., 2002)

### 2.5. Coherence Preserving Pairwise Clique Potentials

Inspired by Boykov and Funka-Lea (2006) the interaction potentials used in our system are

$$V_{st}(y_s, y_t) \propto \exp\left(-\frac{1}{2L} \sum_{l=1}^L \frac{(x_{s_l} - x_{t_l})^2}{(\Sigma_{l,l})_{y_s}^{(t)}}\right) \cdot \frac{\delta(y_s, y_t)}{\text{dist}(s, t)} \quad (19)$$

where vectors  $(x_{s_1}, \dots, x_{s_L})^T$  and  $(x_{t_1}, \dots, x_{t_L})^T$  denote the observed intensities at voxels  $s$  and  $t$  taken from  $L \in \mathbb{N}$  aligned input pulse sequences and

$$\delta(y_s, y_t) = \begin{cases} 1 & \text{if } y_s \neq y_t, \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

The function  $\text{dist}(s, t)$  denotes the physical distance between voxels  $s$  and  $t$ , which varies when working on image volumes with anisotropic voxel spacing. The model emphasizes homogeneous classifications among neighboring voxels but weights penalties for heterogeneity according to intensity similarities of the voxels involved. It assumes the noise among neighboring voxels of an input volume to be distributed in a multivariate Gaussian manner without taking into account dependencies among the spectral channels.

### 2.6. Summary

Reconsidering the processing pipeline of our DMC-EM approach depicted in Fig. 1 we make use of the results from the PBT probability estimation and classification step in the subsequent DMC-EM optimization step in two ways: first, we use the PBT hard classification as initial segmentation  $\mathbf{y}^{(0)}$  where  $y_s^{(0)} = \arg \max_k \tilde{p}^k(+1|\mathbf{z}_s)$  at the beginning of the EM iterations. Based on this initial hard classification the parameters  $\Theta^{(0)}$  are initialized via class-wise maximum likelihood estimation. Second, the probability estimates serve as constraints for the maximization of Equation (3) via ICM within every iteration  $t$  as well as for the parameter updates given by Equations (5) and (6). This is achieved by defining the unary clique potentials as functions of the PBT probability estimates in Equation (17). We therefore utilize the discriminative model involved not only as a preprocessing step but also throughout the whole optimization procedure to repeatedly regularize model adaptation.

**Table 1.** Summary of the publicly available standard databases from the BrainWeb repository used for evaluation purposes.

	Multi-spectral BrainWeb	Mono-spectral BrainWeb
<b>Source</b>	www.bic.mni.mcgill.ca/brainweb	www.bic.mni.mcgill.ca/brainweb
<b>Volume Size</b>	181 × 217 × 181	181 × 217 × 181
<b>Voxel Spacing</b>	1.0 × 1.0 × 1.0 mm <sup>3</sup>	1.0 × 1.0 × 1.0 mm <sup>3</sup>
<b>Spectral Channels</b>	T1, T2, PD	T1
<b>Number of Scans</b>	10	10

### 3. Validation

#### 3.1. Experimental Setup

For quantitative evaluation of the proposed method we carried out experiments both on mono-spectral as well as on multi-spectral (T1-weighted, T2-weighted, PD-weighted) publicly available simulated MRI scans from Cocosco et al. (1997) (see Table 1). All the simulated MRI volume sequences share resolution and size of  $1.0 \times 1.0 \times 1.0 \text{ mm}^3$  and  $181 \times 217 \times 181$ , respectively. INU and noise levels vary among 20% and 40%, and 1%, 3%, 5%, 7%, and 9%, correspondingly. The noise in the simulated images follows a Rayleigh distribution in the background and a Rician distribution in the signal regions. The noise level represents the percent ratio of the standard deviation of the white Gaussian noise added to the real and imaginary channels during simulation versus a reference tissue intensity.

Furthermore, our system was quantitatively evaluated on two sets of real T1-weighted MRI scans provided by the Center of Morphometric Analysis at the Massachusetts General Hospital (see Table 2), which are publicly available on the Internet Brain Segmentation Repository (IBSR). One of the data sets consists of 20 coronal T1-weighted MRI volumes ( $256 \times 65 \times 256$ ) of normal subjects with a resolution of  $1.0 \times 3.1 \times 1.0 \text{ mm}^3$  (IBSR 20). The other one (IBSR 18) consists of 18 scans ( $256 \times 256 \times 128$ ) of normal subjects with varying resolutions ( $0.84 \times 0.84 \times 1.5 \text{ mm}^3$ ,  $0.94 \times 0.94 \times 1.5 \text{ mm}^3$ , and  $1.0 \times 1.0 \times 1.5 \text{ mm}^3$ ). Both the sets are accompanied by ground-truth segmentations of the three tissue types of interest (CSF, GM, and WM). All the scans had been subject to a specific preprocessing including spatial normalization before they were released in the IBSR. However, our system does not make use of the additional spatial information provided herewith and the scans are treated as if they were native scans according to the common quality standards of radiological image acquisition.

All the images were re-oriented to a uniform orientation (“RAI”; right-to-left, anterior-to-posterior, inferior-to-superior). The discriminative model involved was trained on one volumetric scan of the IBSR 20 data set, which is therefore excluded from the quantitative evaluations. In order to keep our system as general as possible, we use the same model for multi-spectral data and carry out PBT probability estimation and hard classification based on the T1-weighted pulse sequences. We measure

**Table 2.** Summary of the publicly available standard databases from the IBSR used for evaluation purposes.

	IBSR 18	IBSR 20
<b>Source</b>	www.cma.mgh.harvard.edu/ibsr	www.cma.mgh.harvard.edu/ibsr
<b>Volume Size</b>	256 × 256 × 128	256 × 65 × 256
<b>Voxel Spacing</b>	0.84 × 0.84 × 1.5 mm <sup>3</sup> , 0.94 × 0.94 × 1.5 mm <sup>3</sup> , 1.0 × 1.0 × 1.5 mm <sup>3</sup>	1.0 × 3.1 × 1.0 mm <sup>3</sup>
<b>Spectral Channels</b>	T1	T1
<b>Number of Scans</b>	18	20

segmentation accuracy by means of the Dice coefficient and the Jaccard coefficient to ensure comparability to other work (see Tables 4–7). The quality of INU correction is quantified by the class-wise coefficient of variation (COV = standard deviation/average) achieved.

Table 3 summarizes the methods whose accuracy will be compared one against the other later. All of them were evaluated on at least one of the publicly available standard databases mentioned above.

Due to the larger amount of free parameters involved, especially with regards to the PBT model, we did not have the ambition to evaluate every possible choice of parameter settings throughout the processing pipeline. For every processing step design choices were based on what can be found in the literature, e.g., (Tu et al., 2006a). For example, we set the weight of the pair-wise clique potentials  $\beta = 1.2$  in accordance with Cuadra et al. (2005) whose Potts model-based pair-wise clique potentials have approximately the same range as ours. The PBT voxel classifiers were built from approximately one million samples randomly selected from one training volume (data set 1 from IBSR 20). The samples are voxels within the brain of the subject and are uniformly distributed over all the input slices of the training scan. Due to the high number of voxels within the brain of the subjects we restricted training sample generation to only one data set. It provides a sufficiently large training subset with enough variance to entirely capture the classification tasks we are interested in. The maximum number of features selected by AdaBoost in each tree node was set to 8. The maximum depth of the trees learned was restricted to 10 and a soft thresholding parameter of  $\epsilon = 0.05$  was used. Initial experiments showed that different parameter choices for the PBT classifier are of negligible impact on the final segmentation results. Furthermore, we consider the choice for the PBT algorithm to generate discriminative models for our DMC-EM approach to be only one of a multitude of alternatives, such as pure AdaBoost (Freund and Shapire, 1995) or Random Forests (Breiman, 2001). We expect such alternatives to yield comparable results. The 3-D voxel context chosen for computing the 747 Haar-like features used per individual voxel sample was of size  $30 \times 30 \times 30 \text{mm}^3$  centered at the voxel of interest. For PBT probability estimation and classifier training the scans were re-sampled to a voxel spacing of  $2.0 \times 2.0 \times 2.0 \text{mm}^3$ . This is done to be able to capture a larger context surrounding the voxels of interest while keeping training feasible

**Table 3.** Summary of the methods used for benchmarking.

Method	Characteristics	INU correction	Multi-spectral
<b>DMC-EM</b>	Parametric EM-based approach with MRF prior and integrated discriminative model relying on MRI-specific Haar-like features and rigidly aligned probabilistic atlas-based features	Yes	Yes
Awate et al. (2006)	Iterative approach with adaptive, non-parametric MRF prior and affinely aligned probabilistic atlas-based initialization and regularization	No	Yes
van Leemput et al. (1999b)	Parametric EM-based approach with MRF prior, rigidly aligned probabilistic atlas-based initialization and regularization	Yes	Yes
Bazin and Pham (2008)	Fuzzy classification approach with rigidly aligned probabilistic and topological atlas-based initialization and simultaneous rigid re-alignment and topology preservation	Yes	Yes
LOCUS-T (Scherrer et al., 2007)	Parametric EM-based approach with MRF prior and integrated FLM-based regularization, Fuzzy C-Means initialization and regular image volume decomposition	No	No
FBM-T (Scherrer et al., 2008)	Parametric EM-based approach with MRF prior with integrated affinely aligned probabilistic atlas-based initialization and regularization and integrated parameter regularization across image sub-volumes	No	No
Akselrod-Ballin et al. (2006)	Support vector machine-based voxel classification relying on intensity, texture, shape, and rigidly aligned probabilistic atlas-based features	No	No
Akselrod-Ballin et al. (2007)	Bayesian multiscale segmentation framework with affinely aligned probabilistic atlas-based initialization and regularization and non-parametric tissue class modeling	No	No
HMRf-EM (Zhang et al., 2001)	Parametric EM-based approach with MRF prior with thresholding-based initialization	Yes	No
Bricq et al. (2008)	Parametric EM-based approach with HMC prior and non-rigidly aligned probabilistic atlas-based initialization and regularization	Yes	No
Ashburner and Friston (2005)	Parametric EM-based approach with simultaneous non-rigid alignment of probabilistic atlas priors for regularization	Yes	No
Marroquin et al. (2002)	Parametric MPM-MAP-based approach with MRF prior and non-rigidly aligned probabilistic atlas-based initialization and regularization	Yes	No

on standard hardware and to establish a uniform compromise between the different resolutions of the benchmarking data sets used. After initially assigning probability estimates to the down-sampled voxels the data is up-sampled again and DMC-EM is carried out on the original resolution.

In a standard C++ implementation of our segmentation framework, it takes about 12 minutes to process one mono-spectral MRI volume ( $181 \times 217 \times 181$ ) without brain extraction and affine alignment on a Fujitsu Siemens notebook equipped with an Intel Core 2 Duo CPU (2.20 GHz) and 3 GB of memory. Pre-processing in total takes about 3 minutes (brain extraction: 1 minute, atlas alignment: 2 minutes) on the same hardware. Training one PBT classifier takes about 1 hour and 15 minutes. During all our experiments, mono-spectral and multi-spectral, we keep a uniform parameter setting for all the free parameters involved both for PBT training and probability estimation as well as for DMC-EM optimization including pre-processing. We can therefore exclude

over-adaptation to one particular set of MRI scans.

**Table 4.** Average segmentation accuracy for multi-spectral (T1-weighted, T2-weighted, and PD-weighted) simulated BrainWeb data of varying noise and INU levels. From left to right the columns contain the tissue class and the segmentation accuracy in terms of the average Dice coefficient achieved by the alternative method and by DMC-EM. For a reliable comparison the average values for DMC-EM are computed for exactly the same noise and INU levels as the ones for the other methods.

Method	Tissue Class	Accuracy	DMC-EM
<b>Marroquin et al. (2002)</b>	WM	$0.95 \pm 0.02$	$0.94 \pm 0.02$
	GM	$0.94 \pm 0.02$	$0.92 \pm 0.04$
	CSF	-	$0.77 \pm 0.03$
<b>van Leemput et al. (1999b)</b>	WM	$0.92 \pm 0.03$	$0.94 \pm 0.02$
	GM	$0.93 \pm 0.02$	$0.92 \pm 0.04$
	CSF	-	$0.77 \pm 0.03$
<b>Bazin and Pham (2008)</b>	WM	$0.94 \pm 0.02$	$0.93 \pm 0.02$
	GM	$0.92 \pm 0.02$	$0.91 \pm 0.04$
	CSF	$0.92 \pm 0.01$	$0.80 \pm 0.06$
<b>Awate et al. (2006)</b>	WM	$0.95 \pm 0.01$	$0.94 \pm 0.02$
	GM	$0.91 \pm 0.01$	$0.92 \pm 0.04$
	CSF	-	$0.77 \pm 0.03$

### 3.2. Quantitative Results on Multi-Spectral Simulated BrainWeb

#### Data

Results on multi-spectral BrainWeb data obtained by DMC-EM are comparable to those of Bazin and Pham (Bazin and Pham, 2008)<sup>†</sup> and van Leemput et al. (1999b)<sup>‡</sup> as depicted in Table 4. They are close to those of Awate et al. (2006)<sup>§</sup> and worse than those of Marroquin et al. (2002)<sup>||</sup>.

Fig. 4 shows that INU, measured by the average coefficient of variation, is reduced for all the spectral channels. It also gives a visual impression of the obtained INU fields in comparison to the ground-truth bias fields.

### 3.3. Quantitative Results on Mono-Spectral Simulated BrainWeb

#### Data

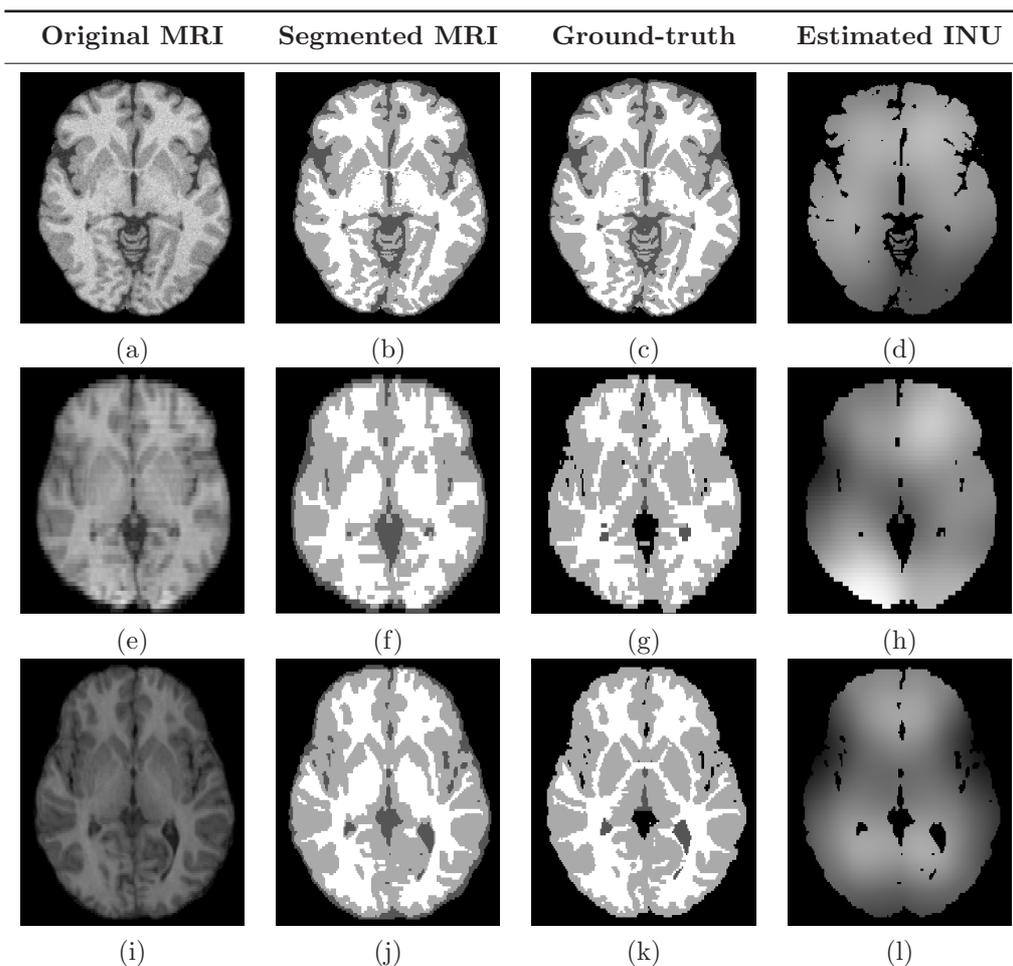
As Table 5 shows, the results achieved for mono-spectral BrainWeb data are comparable to those of other state-of-the-art approaches to brain tissue classification (Awate et al., 2006; Scherrer et al., 2007, 2008; Bazin and Pham, 2008; Marroquin et al., 2002;

<sup>†</sup> Averaged over “varying levels of noise and inhomogeneity” (Bazin and Pham, 2008); we assume all possible noise (0–9%) and INU (0–40%) levels.

<sup>‡</sup> Average over noise levels 1–9% and INU level 40%

<sup>§</sup> Average over noise level 0–9% and INU level 40%

<sup>||</sup> Average over noise levels 1–9% and INU levels 0% and 40%

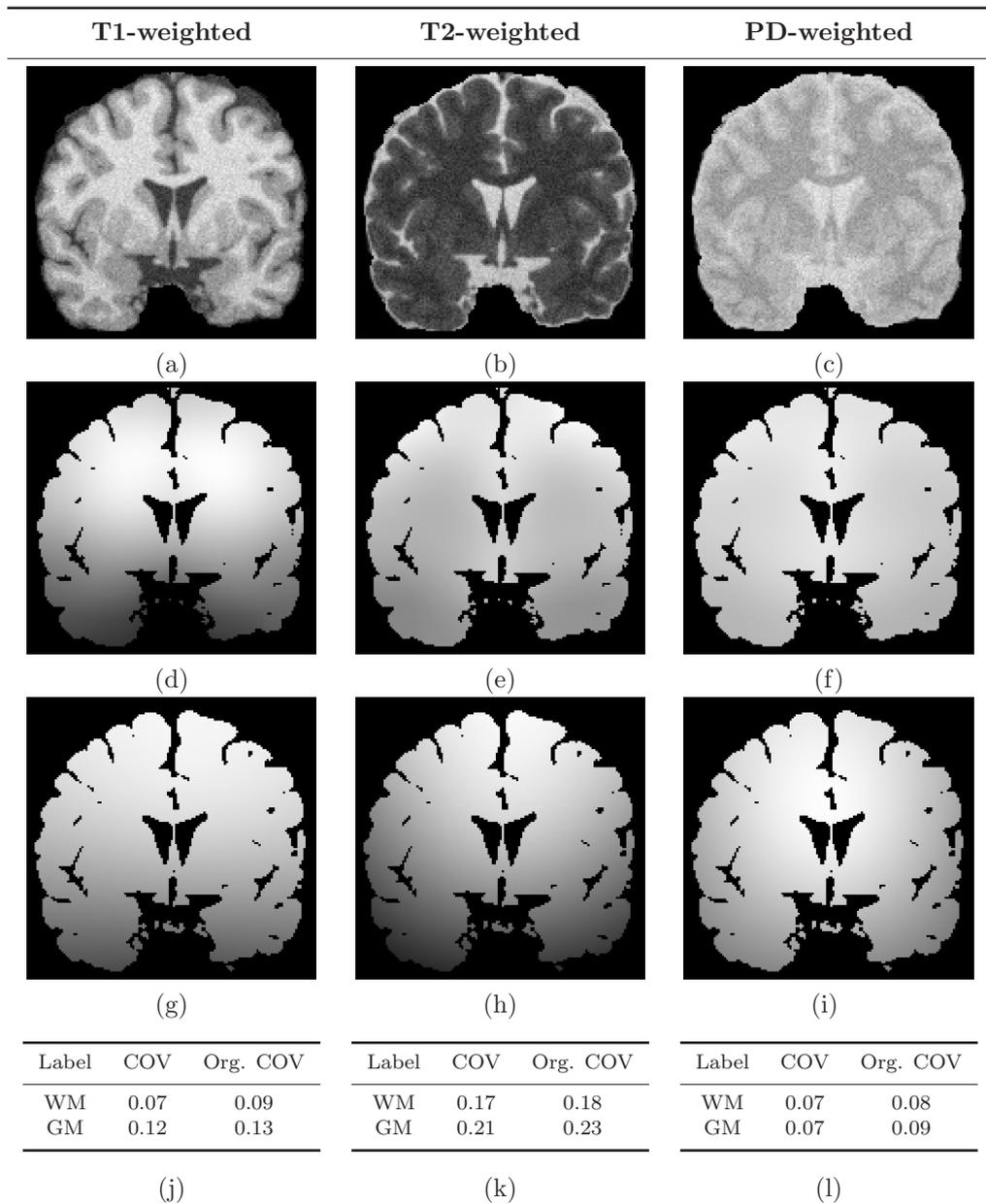


**Figure 3.** Axial slices of original images, the segmentation results, the ground-truth and the estimated INU field for one mono-spectral T1-weighted BrainWeb volume (5% noise, 20% INU) (a–d), one volume of the IBSR 20 Normal Subjects data set (e–h), and one volume of the IBSR 18 Subjects data set (i–l) .

Ashburner and Friston, 2005). The results of Ashburner and Friston (2005) are reported by Tsang et al. (2008). The results are better than those of van Leemput et al. (1999b) and the original HMRF-EM approach (Zhang et al., 2001)¶. Awate et al. (2006), and Bazin and Pham (2008) average over the same BrainWeb data sets as mentioned above for their experimental results. The methods of Awate et al. (2006) and van Leemput et al. (1999b) are evaluated on BrainWeb data sets corrupted by noise levels 0%, 1%, 3%, 5%, 7%, and 9% and an INU level of 40%. Scherrer et al. (2007, 2008) present average values for noise levels 1%, 3%, 5%, 7%, and 9% and INU levels 20% and 40%. For Ashburner and Friston (2005) and Zhang et al.’s HMRF-EM (Zhang et al., 2001) the values are averaged over noise levels 1%, 3%, 5%, and 7% and INU level 20%. For Bricq et al. (2008) the values are averaged over noise levels 0%, 1%, 3%, 5%, 7%, and 9% and INU level 20%, for Marroquin et al. (2002) over noise levels 1%, 3%, 5%, 7%,

¶ Reported by Tsang et al. (2008)





**Figure 4.** Coronal slices of original multi-spectral (T1-weighted, T2-weighted, and PD-weighted) BrainWeb images of 5% noise and 20% INU (a–c), estimated INU fields (d–f), and ground-truth INU fields (g–i). Average INU correction accuracy on multi-spectral BrainWeb data in terms of the coefficient of variation before and after INU correction (j–l).

#### 3.4. Quantitative Results on Normal Subjects Mono-Spectral Scans

With regards to experimental comparison our method shows better results in terms of segmentation accuracy (Jaccard coefficient) than the methods of Akselrod-Ballin et al. (2006) and Marroquin et al. (2002) (see Table 7) for the IBSR 20 data set.

In terms of the Dice coefficient DMC-EM reaches a higher accuracy for GM

**Table 7.** Average segmentation accuracy for IBSR 20 with exclusion of data set no. 1 that has been used for training. From left to right the columns contain the tissue class and the achieved average Dice and Jaccard coefficients.

Method	Tissue Class	Dice Coeff.	Jaccard Coeff.
<b>DMC-EM</b>	WM	$0.81 \pm 0.09$	$0.69 \pm 0.12$
	GM	$0.82 \pm 0.06$	$0.71 \pm 0.08$
	CSF	$0.83 \pm 0.05$	$0.71 \pm 0.07$
<b>Akselrod-Ballin et al. (2006)</b>	WM	-	0.67
	GM	-	0.68
	CSF	-	-
<b>Marroquin et al. (2002)</b>	WM	-	0.68
	GM	-	0.66
	CSF	-	0.23
<b>Ashburner and Friston (2005)</b>	WM	-	-
	GM	0.79	-
	CSF	-	-
<b>HMRF-EM (Zhang et al., 2001)</b>	WM	-	-
	GM	0.76	-
	CSF	-	-

segmentation than the method of Ashburner and Friston (2005)<sup>†</sup> and the original HMRF-EM (Zhang et al., 2001)<sup>‡</sup>. Table 6(b) shows that all the data sets were, on average, successfully corrected for INU.

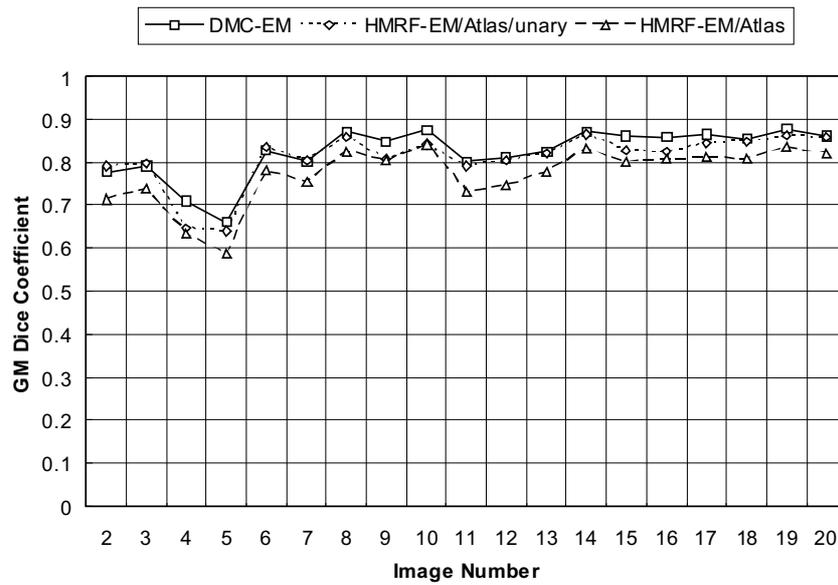
As depicted in Figs. 5 and 6 DMC-EM constantly gives better results than pure HMRF-EM with zero-valued unary clique potentials and probabilistic atlas-based initialization. Except for a few cases it also gives better results than the HMRF-EM approach with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization.

On the IBSR 18 data set our method performs comparably to other state-of-the-art approaches with regards to segmentation accuracy (see Table 8). The statistics are mostly better than the ones of the competing methods when we exclude outlier data set 10 from the evaluation. In this case DMC-EM segmentation suffers from a poor result of the initial brain extraction procedure: while the brain is sufficiently well extracted from the upper part of the skull the whole neck and face area remained in the brain mask after skull stripping.

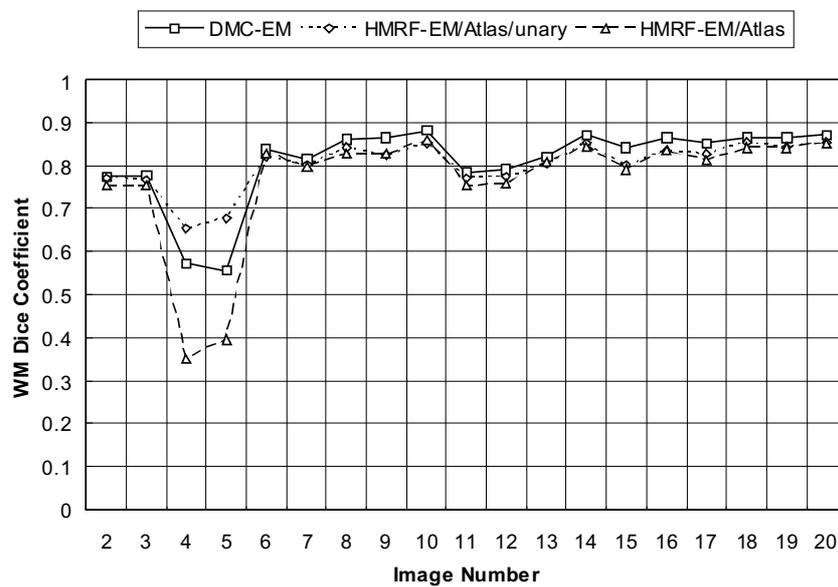
Figs. 7 and 8 show that the introduction of discriminative model dependent unary clique potentials and PBT initialization improves segmentation accuracy for the IBSR 18 data set. In comparison to the HMRF-EM approach with zero-valued unary clique potentials and probabilistic atlas-based initialization and to the HMRF-EM approach with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization DMC-EM usually reaches a higher segmentation accuracy in terms of the

<sup>†</sup> Reported by Tsang et al. (2008)

<sup>‡</sup> Reported by Tsang et al. (2008)



**Figure 5.** Achieved accuracy for GM segmentation in terms of the Dice coefficient for the IBSR 20 data set by the DMC-EM algorithm, the HMRf-EM algorithm with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization, and the HMRf-EM algorithm with zero-valued unary clique potentials and probabilistic atlas-based initialization.



**Figure 6.** Achieved accuracy for WM segmentation in terms of the Dice coefficient for the IBSR 20 data set by the DMC-EM algorithm, the HMRf-EM algorithm with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization, and the HMRf-EM algorithm with zero-valued unary clique potentials and probabilistic atlas-based initialization.

Dice coefficient for GM and WM.

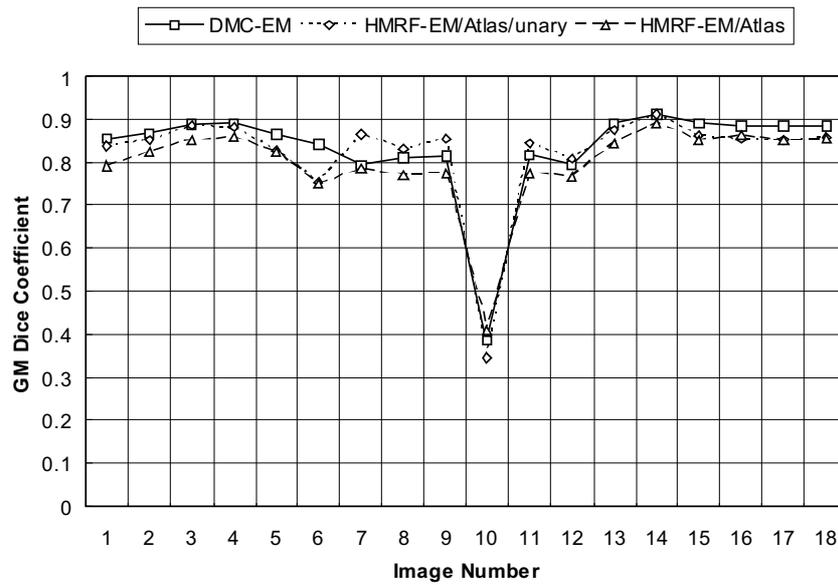
In our last experiments we intend to illustrate the accuracy of our method in particular brain regions on real-world data. For the IBSR 18 data collection there are more detailed ground-truth annotations available that identify individual brain structures, which together form the three common segments (WM, GM, and CSF) our method aims to segment. These annotations allow computing region specific false negative rates (FNR) achieved by our trinary segmentation algorithm. In Figs. 9–12 we studied regional misclassifications for the caudate nucleus, the hippocampus, the whole cerebral cortex, and the internal ventricular system. In all the scenarios, except for the caudate nucleus, DMC-EM achieves a lower FNR than the HMRF-EM approach with zero-valued unary clique potentials and probabilistic atlas-based initialization and to the HMRF-EM approach with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization.

**Table 8.** Average segmentation accuracy for IBSR 18. From left to right the columns contain the tissue label and the achieved average Dice and Jaccard coefficients for all the data sets and for data sets 1–9 and 11–18 with outlier data set 10 removed in brackets.

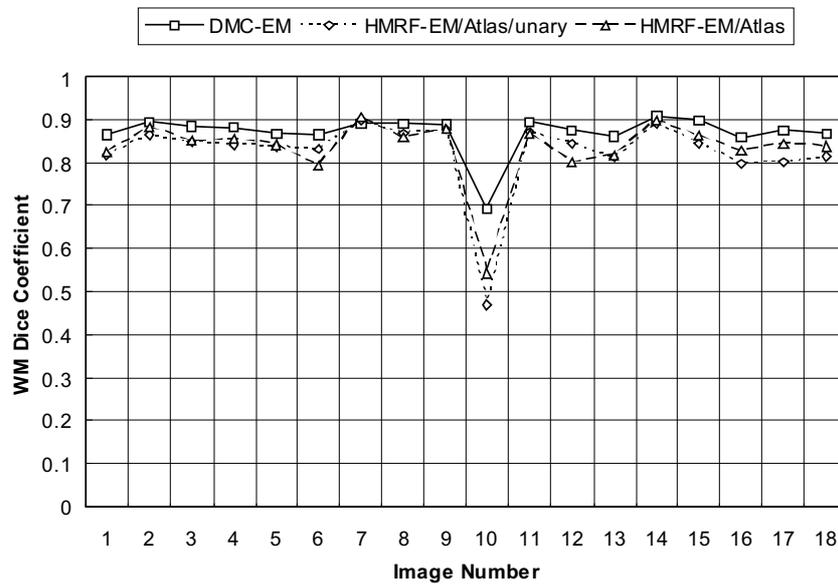
Method	Label	Dice Coeff.	Jaccard Coeff.
<b>DMC-EM</b>	WM	$0.87 \pm 0.05$ ( $0.88 \pm 0.01$ )	$0.77 \pm 0.06$ ( $0.79 \pm 0.02$ )
	GM	$0.83 \pm 0.12$ ( $0.86 \pm 0.04$ )	$0.73 \pm 0.13$ ( $0.76 \pm 0.06$ )
	CSF	$0.76 \pm 0.09$ ( $0.77 \pm 0.08$ )	$0.62 \pm 0.11$ ( $0.63 \pm 0.10$ )
<b>Bazin and Pham (Bazin and Pham, 2008)</b>	WM	$0.82 \pm 0.04$	-
	GM	$0.88 \pm 0.01$	-
	CSF	-	-
<b>Akselrod-Ballin et al. (2007)</b>	WM	0.87	-
	GM	0.86	-
	CSF	0.83	-
<b>Awate et al. (2006)</b>	WM	$0.89 \pm 0.02$	-
	GM	$0.81 \pm 0.04$	-
	CSF	-	-
<b>Bricq et al. (2008)</b>	WM	$0.87 \pm 0.02$	-
	GM	$0.80 \pm 0.06$	-
	CSF	-	-

## 4. Discussion

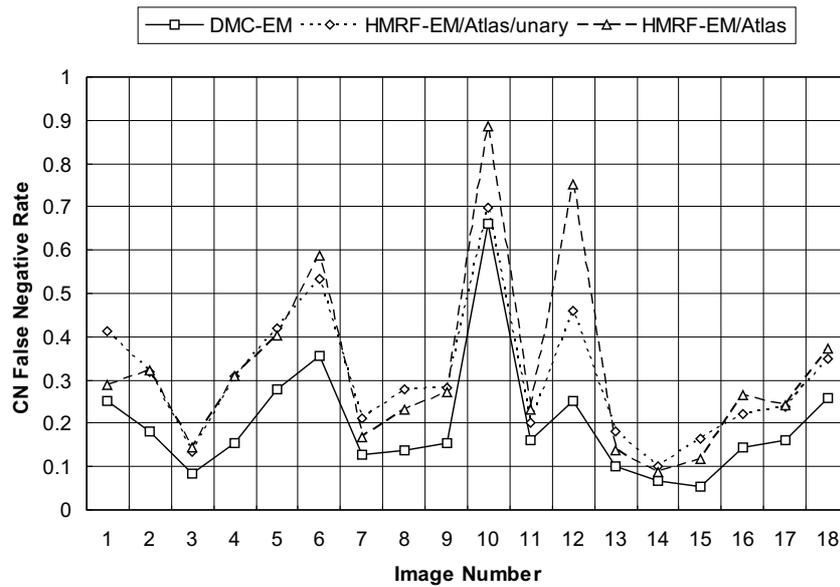
Our newly proposed DMC-EM approach to fully automated 3-D Brain MRI tissue classification and INU correction makes use of two different types of spatial priors: the first one, which contributes the unary clique potentials of the hidden Markov random field’s Gibbs distribution, is derived from a strong discriminative model, in our case a PBT classifier, that has been built from annotated training data. It only makes use of features of reduced INU sensitivity and therefore prevents the model from over-fitting to scanner specific tissue contrast characteristics, which is experimentally validated by detailed evaluations on publicly available data sets from different sources and scanners. Usually, if the set of features is not carefully chosen, using supervised learning for MRI brain tissue classification ties a method to the exact acquisition protocol the classifier is



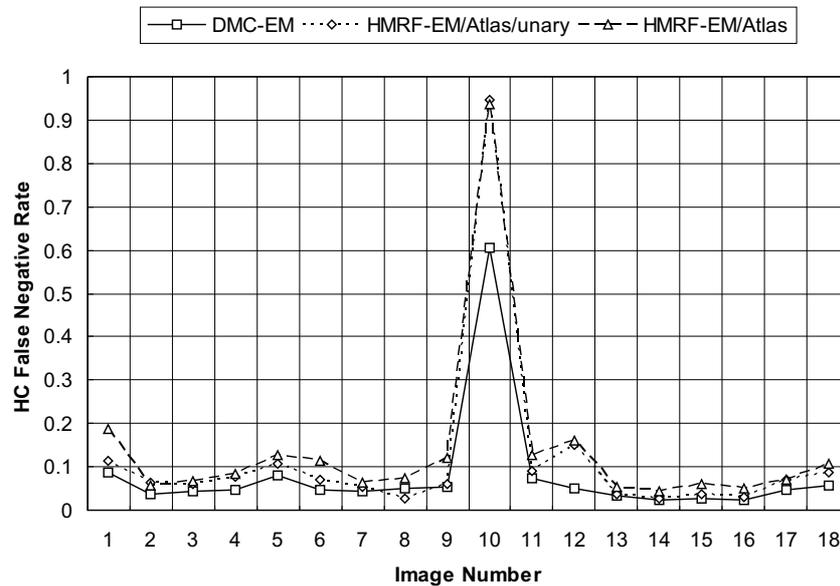
**Figure 7.** Achieved accuracy for GM segmentation in terms of the Dice coefficient for the IBSR 18 data set by the DMC-EM algorithm, the HMRf-EM algorithm with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization, and the HMRf-EM algorithm with zero-valued unary clique potentials and probabilistic atlas-based initialization.



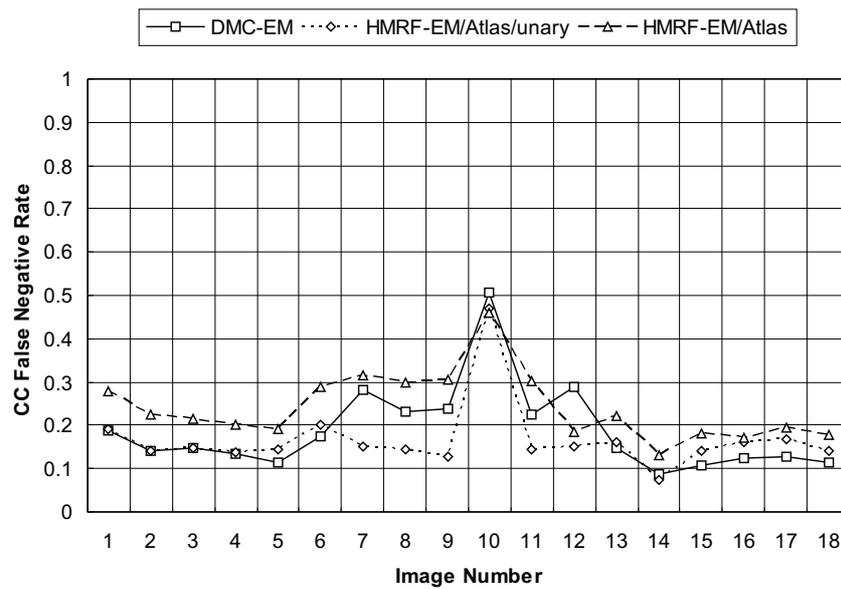
**Figure 8.** Achieved accuracy for WM segmentation in terms of the Dice coefficient for the IBSR 18 data set by the DMC-EM algorithm, the HMRf-EM algorithm with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization, and the HMRf-EM algorithm with zero-valued unary clique potentials and probabilistic atlas-based initialization.



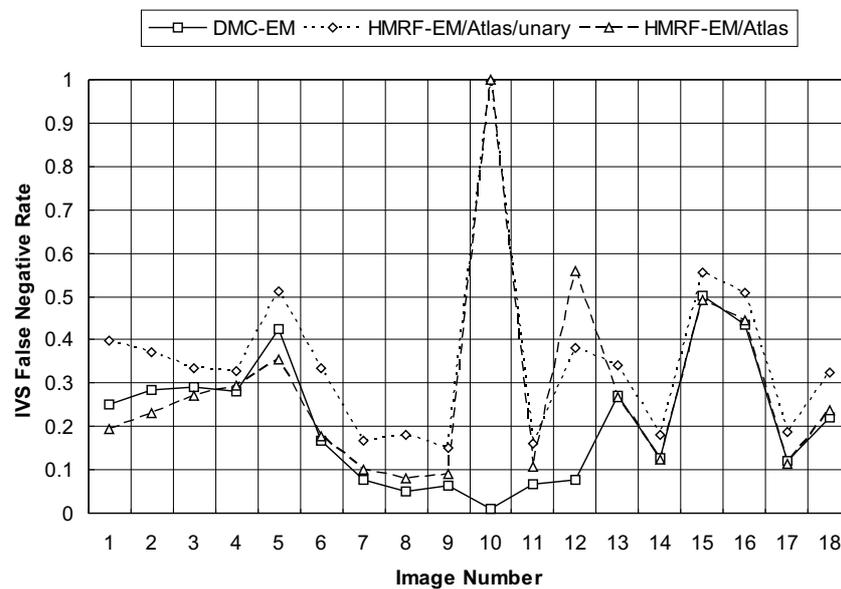
**Figure 9.** The FNR for the segmentation of the caudate nucleus in IBSR 18 by the DMC-EM algorithm, the HMRF-EM algorithm with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization, and the HMRF-EM algorithm with zero-valued unary clique potentials and probabilistic atlas-based initialization.



**Figure 10.** The FNR for the segmentation of the hippocampus in IBSR 18 by the DMC-EM algorithm, the HMRF-EM algorithm with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization, and the HMRF-EM algorithm with zero-valued unary clique potentials and probabilistic atlas-based initialization.



**Figure 11.** The FNR for the segmentation of the cerebral cortex in IBSR 18 by the DMC-EM algorithm, the HMRF-EM algorithm with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization, and the HMRF-EM algorithm with zero-valued unary clique potentials and probabilistic atlas-based initialization.



**Figure 12.** The FNR for the segmentation of the internal ventricular system in IBSR 18 by the DMC-EM algorithm, the HMRF-EM algorithm with probabilistic atlas-based unary clique potentials and probabilistic atlas-based initialization, and the HMRF-EM algorithm with zero-valued unary clique potentials and probabilistic atlas-based initialization.

trained for: the method of Akselrod-Ballin et al. (2007), for instance, relies on stationary observation models that have been derived in a cross-validation setting from separate training volumes, which all origin from the same source of data. Their method might therefore be highly biased to uniform contrast characteristics present in the IBSR 18 data collection and the results may not necessarily adequately reflect the performance of the method when applied to a larger variety of data sets in clinical practice. Han and Fischl (2007) try to weaken this effect by introducing an intensity renormalization procedure into the method of Fischl et al. (2002, 2004). As seen in our experiments an appropriate choice of features can help to circumvent this dependency without the need for additional pre-processing.

Our experimental setup did not allow specially adapted parameter settings for any of the data sets. All free parameters were kept fixed during experimentation. By not only including prior knowledge from an affinely preregistered probabilistic atlas our discriminative model is capable of producing external fields that are more specific to the data at hand. The second prior used, constituting the pairwise clique potentials, is a smoothing prior that penalizes certain configurations in local neighborhoods depending on similarity of observed intensities, physical distance between image voxels, and estimated image noise. This makes the approach robust against different levels of noise, which is also shown by quantitative experimental evaluation.

From the theoretical point of view, in contrast to Zhang et al. (2001), a consistent multi-spectral formulation of our DMC-EM framework both for brain tissue segmentation as well as for INU correction is presented. Accordingly, evaluation is carried out on mono- and multi-spectral data. On all the data sets our method achieves a segmentation accuracy that is either higher or comparable to the state-of-the-art even though progress in this highly investigated branch of research is difficult due to the well-established competitiveness of the methods available.

In accordance with Marroquin et al. (2002) and Bricq et al. (2008) we only observe a limited gain in segmentation accuracy when going from mono-spectral to multi-spectral data. This effect may be due to the fact that the three tissue types of interest can already be almost perfectly separated from the mono-spectral T1-weighted pulse sequences only. Therefore, the additional information about the phantom's true composition provided by further pulse sequences may be rather redundant than of any additive value.

From visually inspecting our segmentation results and from our experiments on individual brain regions we observe that our method seems to reveal weaknesses when it comes to individual GM structure segmentation. Even though the FNRs are low for the caudate nucleus and the hippocampus in Figs. 9 and 10 and the GM area covering, for instance, the caudate nucleus and the putamen could to most parts be successfully segmented in all the images depicted in Fig. 3, the globus pallidus and the thalamus were misclassified in all the three image volumes. As both structures appear brighter than most of the other GM structures our observation model that models tissue classes as single Gaussian distributions seems too restrictive in this case. The problem may be solved by trying to model individual tissue classes, and not only the whole brain,

by mixtures of Gaussians. In addition, more complex discriminative models could be considered that further decompose cervical GM into individual structures (Wels et al., 2009). By doing so the dominance of the prior model over the observation model could be steered separately for individual anatomical entities. Fig. 11 shows that the positive effect of DMC-EM is less apparent at the boundary of the brain where the cerebral cortex is located.

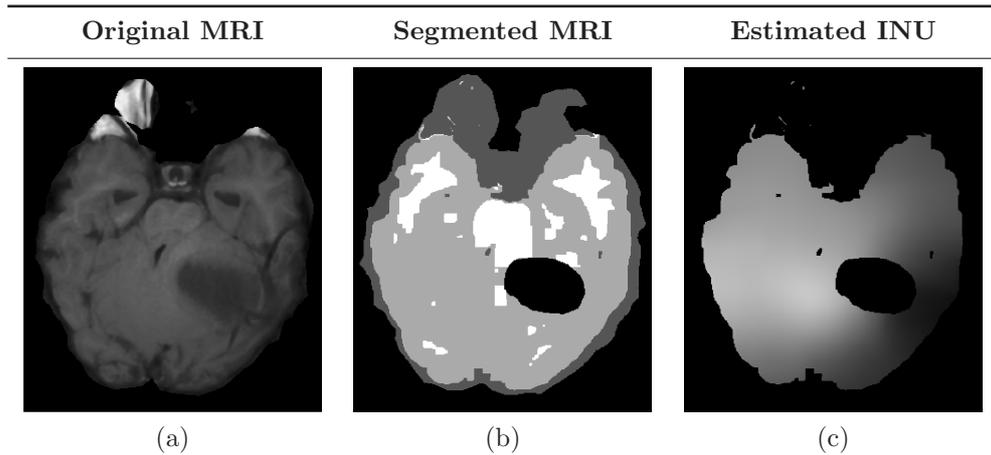
It has to be mentioned also that particular high values for segmentation accuracy on the BrainWeb data sets (see Tables 4 and 5) do not necessarily mean a particular method is giving anatomically correct segmentation results. As depicted in Fig. 3(c) the associated ground-truth annotation suffers from obvious weaknesses in the area of the globus pallidus and the thalamus.

Concerning PVEs our method is conceptually predisposed to explicitly handle the effect that individual voxels may be composed of different tissue types due to the limited resolution of the acquisition devices. The inherent mixture model estimation of our algorithm provides an insight on how or to which degree different tissue types contribute to a certain voxel. However, we decided not to focus on handling PVEs and rather transform our results into hard classifications for evaluation purposes after algorithmic processing.

Similarly to, for instance, the method of Marroquin et al. (2002) our method seems not to be of high accuracy with respect to CSF estimation. This may be caused by the fact that we consider the complete fluid filled space outside and inside the brain to be the CSF segment. It includes both the ventricular system as well as the subarachnoid space. Especially the segmentation of the latter may be subject to errors originating from imperfections of the initial skull stripping procedure. However, our method is carried out completely automatically without any user interaction. Results for CSF segmentation might be better if a “perfect” initial skull stripping was assumed. When considering the internal ventricular system segment only, Fig. 12 shows that DMC-EM performs mostly better than comparable methods for this particular part of the CSF segment.

In principle, our method is also able to be applied to pathologic data. Fig. 13 shows the results obtained when segmenting T1-weighted data (512x512x20) acquired from a patient suffering from a pediatric brain tumor. As the pathologic tissue types have not been modeled explicitly in our approach—we assume three tissue classes to be present in the image region to be segmented—they need to be excluded from the GM, WM, and CSF segmentation procedure. This can be done by fast brain tumor segmentation techniques like, for example, the one to be found in reference (Wels et al., 2008).

With regards to INU correction our method suffers from the same limitations as the method of Wells et al. (1996) does due to the fact that it forms the base of our approach. In a broader context, focusing on the method of Wells et al. (1996) can be seen as an exemplary choice. Other more robust techniques that parametrically constrain estimated INU fields might in fact benefit in an equal manner if they were embedded in our modality-specific discriminative model-constrained HMRF-EM approach. DMC-EM



**Figure 13.** Axial slices of the original image (a), the segmentation result (b), and the estimated INU field (b) for one mono-spectral T1-weighted volume data set where the patient suffers from a pediatric brain tumor.

is comparable fast when compared to other state-of-the-art approaches and it takes only a few minutes to process a data volume. We did not address sub-cortical segmentation as it is beyond the scope of this paper. On the other hand, any generic state-of-the-art approach to organ segmentation will profit significantly from classwise intensity standardized and INU corrected MRI input volumes.

## 5. Conclusions

We have presented an MRI modality-specific discriminative model-constrained HMRF-EM (DMC-EM) approach to brain tissue segmentation and INU correction in multi-spectral 3-D MRI. The major contribution of our work is a strong discriminative model obtained by a PBT classifier that is integrated into the framework by means of unary clique potentials in a mathematically sound manner. The discriminative model used is MRI modality specific as it only relies on features of reduced INU sensitivity taking into account the particularities of the MRI modality.

As experimentally validated the choice of features prevents our method from being tied to a particular acquisition protocol at a specific site or scanner. Detailed quantitative evaluations on publicly available benchmarking databases demonstrate this increased robustness of our approach. At the same time the segmentation accuracy achieved is comparable to those of other state-of-the-art approaches to brain tissue classification in MRI data.

## Appendix A. Discriminative Modeling

### Appendix A.1. Probabilistic Boosting-Trees

Training a PBT resembles inducing a multivariate binary regression tree from a set of weighted labeled training examples  $\mathcal{T} = \{(\mathbf{z}_n, y_n, w_n) \mid n = 1, \dots, N\} \in \mathcal{T}$ ,  $N \in \mathbb{N}$ , with feature vectors  $\mathbf{z}_n \in \mathcal{Z} = \mathbb{R}^M$ ,  $M \in \mathbb{N}$ , labels  $y_n \in \{-1, +1\}$ , and weights  $w_n \in [0, 1]$  with  $\sum_{n=1}^N w_n = 1$ . Within each node  $v$  of the tree a strong discriminative model  $H_v(\mathbf{z}) \in (-1, +1)$  for feature vectors  $\mathbf{z} \in \mathbb{R}^M$ ,  $M \in \mathbb{N}$ , is generated. By construction, all those models  $H(\mathbf{z})$  asymptotically approach an additive logistic regression model (Friedman et al., 1998)

$$H(\mathbf{z}) \approx \frac{1}{2} \ln \frac{p(y = +1|\mathbf{z})}{p(y = -1|\mathbf{z})} \quad (\text{A.1})$$

where  $y \in \{-1, +1\}$  denotes the outcome of the associated binary classification task. Accordingly, at each node  $v$  of the resulting PBT there are current approximations of the posterior probabilities  $\tilde{p}_v(+1|\mathbf{z}) = q_v(\mathbf{z}) = \exp(2H(\mathbf{z})) / (1 + \exp(2H(\mathbf{z})))$  and  $\tilde{p}_v(-1|\mathbf{z}) = 1 - q_v(\mathbf{z})$ . During classification those values are used to guide tree traversing and combined propagation of posteriors in order to get a final approximation  $\tilde{p}(y|\mathbf{z})$  of the true posterior probability  $p(y|\mathbf{z})$  at the tree's root node.

While training the classifier, those probabilities are used to successively split the set of training data relative to the prior probability  $p_v(y = +1)$  associated with the current training (sub-)set in node  $v$  into two new subsets. We write  $p_v$  instead of  $p_v(y = +1)$  in the following for simplicity. The soft thresholding parameter  $\epsilon > 0$  sees to pass on training samples  $\mathbf{z}$  that are close to the current node's decision boundary, that is to say, if  $q_v(\mathbf{z}) \in [(1 - \epsilon)p_v; (1 + \epsilon)p_v]$ , to both of the resulting subsets and associated subtrees. See Algorithm 2 for details on how a PBT is built.

During classification the values for  $q_v(\mathbf{z})$  are used to guide tree traversing and combined propagation of posteriors in order to get final approximations  $\tilde{p}_v(y|\mathbf{z})$  of the true posterior probabilities  $p_v(y|\mathbf{z})$  at each tree node  $v$ : for outgoing edges  $r_v^{-1}$  and  $r_v^1$  associated with the possible classifications the approximation  $\tilde{p}_v(y|\mathbf{z})$  can be computed via the recursive formula

$$\tilde{p}_v(y|\mathbf{z}) = \begin{cases} \tilde{p}_{\beta(r_v^{-1})}(y|\mathbf{z}) & \text{if } q_v(\mathbf{z}) < (1 - \epsilon)p_v, \\ \tilde{p}_{\beta(r_v^1)}(y|\mathbf{z}) & \text{if } q_v(\mathbf{z}) > (1 + \epsilon)p_v, \\ \sum_i \tilde{p}_{\beta(r_i)}(y|\mathbf{z}) \cdot q_v(i|\mathbf{z}) & \text{otherwise,} \end{cases} \quad (\text{A.2})$$

where  $\beta(r)$  denotes the vertex where edge  $r$  ends and  $q_v(+1|\mathbf{z}) = q_v(\mathbf{z})$  and  $q_v(-1|\mathbf{z}) = 1 - q_v(\mathbf{z})$ .

### Appendix A.2. AdaBoost

Probabilistic boosting-trees can be built in combination with several strong learning algorithms providing the strong classifier within each tree node. In the following we give a concise description of the most commonly used one, which is AdaBoost (Freund

and Shapire, 1995). It is called Discrete AdaBoost by Friedman et al. (1998). In the two-class classification setting we have a set  $\mathcal{T} = \{(\mathbf{z}_n, y_n, w_n) \mid n = 1, \dots, N\} \in \mathcal{T}$  of weighted labeled training data,  $N \in \mathbb{N}$ , with feature vectors  $\mathbf{z}_n \in \mathcal{Z} = \mathbb{R}^M$ ,  $M \in \mathbb{N}$ , labels  $y_n \in \{-1, +1\}$ , and weights  $w_n = 1/N$ . The purpose of Discrete AdaBoost is to find a strong classifier

$$H(\mathbf{z}) = \sum_{t=1}^T \alpha_t h_t(\mathbf{z}), \quad (\text{A.3})$$

that is, a linear combination of  $T \in \mathbb{N}$  weak classifiers  $h_t(\mathbf{z})$  giving hard classification outcomes with weights  $\alpha_t \in \mathbb{R}$ ; the corresponding prediction of this strong classifier is  $\text{sgn}(H(\mathbf{z}))$ . The procedure builds weak classifiers on weighted training samples in turn giving higher weight to those that are currently misclassified. A detailed description of Discrete AdaBoost is given in Algorithm 3.

For the purpose of discriminative brain tissue modeling, we use a generalized version of AdaBoost, which is called Real AdaBoost (Friedman et al., 1998) (see Algorithm 4). One of the major differences to Discrete AdaBoost is the fact that the weak learners return class probability estimates  $f(\mathbf{z}) = p(y = +1 \mid \mathbf{z})$  instead of hard classifications. We generate class probability estimates by means of decision stumps, which are inductively learned decision trees of depth 1, returning the probability distributions of  $y \in \{-1, +1\}$  after only one split of a training data set  $\mathcal{T}$ . A split is found by choosing a feature vector component  $z_m$  and an accompanying threshold  $\theta_m$  that “best” separates the positive from the negative samples. We refer to Quinlan (1986) for details on this.

**Algorithm 2: PBT**


---

**Input:** set of weighted labeled training examples  $\mathcal{T} = \{(\mathbf{z}_n, y_n, w_n) \mid n = 1, \dots, N\} \in \mathcal{T}$ ,  $N \in \mathbb{N}$ , with feature vectors  $\mathbf{z}_n \in \mathcal{Z} = \mathbb{R}^M$ ,  $M \in \mathbb{N}$ , labels  $y_n \in \{-1, +1\}$ , and weights  $w_n \in [0, 1]$ ,  $\sum_{n=1}^N w_n = 1$ , a strong discriminative probability estimator  $\mathbf{L} : \mathcal{T} \times \mathbb{N} \rightarrow \{f : \mathcal{Z} \rightarrow (0, 1) \text{ with } f(\mathbf{z}) = p(y = +1 | \mathbf{z})\}$ , the number of weak classifiers  $S \in \mathbb{N}$  per tree node, the current tree depth  $d \in \mathbb{N}$  (initially  $d = 0$ ), and the maximum tree depth  $D \in \mathbb{N}$

**Output:** Probabilistic Boosting-Tree node

**begin**

    Let  $v$  be the current tree node;

    // Compute the empirical distribution  
     $p_v \leftarrow \sum_{n=1}^N w_n \delta(+1, y_n)$ ;

    // Train a strong discriminative model  
     $q_v \leftarrow \mathbf{L}[\mathcal{T}, S]$ ;

    // Initialize subsets

**if**  $d=D$  **then**

        | return  $v$

**else**

        Add new tree nodes  $\beta(r_v^{-1})$  and  $\beta(r_v^{+1})$ ;

$\mathcal{T}^{-1} = \emptyset$ ;

$\mathcal{T}^{+1} = \emptyset$ ;

**for**  $n = 1, \dots, N$  **do**

**if**  $q_v(\mathbf{z}_n) < (1 - \epsilon)p_v$  **then**

                |  $\mathcal{T}^{-1} \leftarrow \mathcal{T}^{-1} \cup \{(\mathbf{z}_n, y_n, w_n)\}$ ;

**else**

**if**  $q_v(\mathbf{z}_n) > (1 + \epsilon)p_v$  **then**

                    |  $\mathcal{T}^{+1} \leftarrow \mathcal{T}^{+1} \cup \{(\mathbf{z}_n, y_n, w_n)\}$ ;

**else**

                    |  $\mathcal{T}^{-1} \leftarrow \mathcal{T}^{-1} \cup \{(\mathbf{z}_n, y_n, w_n)\}$ ;

                    |  $\mathcal{T}^{+1} \leftarrow \mathcal{T}^{+1} \cup \{(\mathbf{z}_n, y_n, w_n)\}$ ;

**end**

**end**

**end**

**end**

        // Increase tree depth and normalize  
         $d \leftarrow d + 1$ ;

**for**  $n = 1, \dots, |\mathcal{T}^{-1}|$  **do**

            |  $w_n \leftarrow w_n / (\sum_{n=1}^{|\mathcal{T}^{-1}|} w_n)$ ;

**end**

**for**  $n = 1, \dots, |\mathcal{T}^{+1}|$  **do**

            |  $w_n \leftarrow w_n / (\sum_{n=1}^{|\mathcal{T}^{+1}|} w_n)$ ;

**end**

        // Repeat procedure recursively  
         $\beta(r_v^{-1}) \leftarrow PBT[\mathcal{T}^{-1}, \mathbf{L}, S, d, D]$ ;

$\beta(r_v^{+1}) \leftarrow PBT[\mathcal{T}^{+1}, \mathbf{L}, S, d, D]$ ;

        return  $v$ ;

**end**

**end**

---

**Algorithm 3:** Discrete AdaBoost

**Input:** set of weighted labeled training examples  $\mathcal{T} = \{(\mathbf{z}_n, y_n, w_n) \mid n = 1, \dots, N\} \in \mathcal{T}$ ,  $N \in \mathbb{N}$ , with feature vectors  $\mathbf{z}_n \in \mathcal{Z} = \mathbb{R}^M$ ,  $M \in \mathbb{N}$ , labels  $y_n \in \{-1, +1\}$ , and weights  $w_n = 1/N$ , a weak learning algorithm  $\mathbf{L} : \mathcal{T} \rightarrow \{h : \mathcal{Z} \rightarrow \{-1, +1\}\}$ , and the number of weak classifiers  $T \in \mathbb{N}$

**Output:** strong classifier  $H : \mathcal{Z} \rightarrow \mathbb{R}$  with  $H(\mathbf{z}) = \sum_{t=1}^T \alpha_t h_t(\mathbf{z})$

**begin**

**for**  $t = 1, \dots, T$  **do**

    // Build weak classifier

$h_t \leftarrow \mathbf{L}[\mathcal{T}]$ ;

    // Compute error rate

$\epsilon \leftarrow 0$ ;

**for**  $n = 1, \dots, N$  **do**

**if**  $h_t(\mathbf{x}_n) \neq y_n$  **then**

$\epsilon \leftarrow \epsilon + w_n$ ;

**end**

**end**

    // Adapt sample weights

**for**  $n = 1, \dots, N$  **do**

**if**  $h_t(\mathbf{x}_n) = y_n$  **then**

$w_n \leftarrow w_n \cdot \epsilon / (1 - \epsilon)$ ;

**end**

**end**

**for**  $n = 1, \dots, N$  **do**

$w_n \leftarrow w_n / (\sum_{n=1}^N w_n)$ ;

**end**

    // Compute weights of weak classifiers

$\alpha_m \leftarrow \log \frac{1-\epsilon}{\epsilon}$ ;

**end**

  return  $H(\mathbf{z}) = \sum_{t=1}^T \alpha_t h_t(\mathbf{z})$ ;

**end**

**Algorithm 4:** Real AdaBoost

---

**Input:** set of weighted labeled training examples  $\mathcal{T} = \{(\mathbf{z}_n, y_n, w_n) \mid n = 1, \dots, N\} \in \mathcal{T}$ ,  
 $N \in \mathbb{N}$ , with feature vectors  $\mathbf{z}_n \in \mathcal{Z} = \mathbb{R}^M$ ,  $M \in \mathbb{N}$ , labels  $y_n \in \{-1, +1\}$ , and  
weights  $w_n = 1/N$ , a discriminative probability distribution estimator  
 $\mathbf{L} : \mathcal{T} \rightarrow \{f : \mathcal{Z} \rightarrow (0, 1) \text{ with } f(\mathbf{z}) = p(y = +1 | \mathbf{z})\}$ , and the number of weak  
classifiers  $T \in \mathbb{N}$

**Output:** strong classifier  $H : \mathcal{Z} \rightarrow \mathbb{R}$  with  $H(\mathbf{z}) = \sum_{t=1}^T h_t(\mathbf{z})$

**begin**

**for**  $t = 1, \dots, T$  **do**

        // Build probability estimator

$f_t \leftarrow \mathbf{L}[\mathcal{T}]$ ;

$\forall_{\mathbf{z}} h_t(\mathbf{z}) \leftarrow 0.5 \cdot \log \frac{f_t(\mathbf{z})}{1-f_t(\mathbf{z})}$ ;

        // Adapt sample weights

**for**  $n = 1, \dots, N$  **do**

$w_n \leftarrow w_n \cdot \exp(-y_n h_t(\mathbf{z}))$ ;

**end**

**for**  $n = 1, \dots, N$  **do**

$w_n \leftarrow w_n / (\sum_{n=1}^N w_n)$ ;

**end**

**end**

    return  $H(\mathbf{z}) = \sum_{t=1}^T h_t(\mathbf{z})$ ;

**end**

---

**References**

- A. Akselrod-Ballin, M. Galun, J. M. Gomori, R. Basri, and A. Brandt. Atlas guided identification of brain structures by combining 3D segmentation and SVM classification. In R. Larsen, M. Nielsen, and J. Sporring, editors, *Int. Conf. Med. Image Comput. Comput.-Assist. Interv., Copenhagen, Denmark*, pages 209–216, Oct. 2006.
- A. Akselrod-Ballin, M. Galun, J. M. Gomori, A. Brandt, and R. Basri. Prior knowledge driven multiscale segmentation of brain MRI. In N. Ayache, S. Ourselin, and A. Maeder, editors, *Int. Conf. Med. Image Comput. Comput.-Assist. Interv., Brisbane, Australia*, pages 118–126, Oct. 2007.
- J. Ashburner and K. J. Friston. Unified segmentation. *NeuroImage*, 26(3):839–851, Apr. 2005.
- S. P. Awate, T. Tasdizen, N. Foster, and R. T. Whitaker. Adaptive Markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification. *Med. Image Anal.*, 10(5):726–739, Oct. 2006.
- P.-L. Bazin and D. L. Pham. Homeomorphic brain image segmentation with topological and statistical atlases. *Med. Image Anal.*, 12(5):616–625, Oct. 2008.
- J. Besag. On the statistical analysis of dirty pictures. *J. Roy. Stat. Soc. B Stat. Meth.*, 48(3):259–302, 1986.
- Y. Boykov and G. Funka-Lea. Graph cuts and efficient N-D image segmentation. *Int. J. Comput. Vis.*, 70(2):109–131, Nov. 2006.
- L. Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, Oct. 2001.
- S. Bricq, Ch. Collet, and J. P. Armspach. Unifying framework for multimodal brain MRI segmentation based on hidden Markov chains. *Med. Image Anal.*, 12(6):639–652, Dec. 2008.
- C. A. Cocosco, V. Kollokian, R. K.-S. Kwan, and A. C. Evans. BrainWeb: Online interface to a 3D MRI simulated brain database. In *Int. Conf. Func. Mapp. Hum. Brain, Copenhagen, Denmark*, page 425, May 1997.
- M. Bach Cuadra, L. Cammoun, T. Butz, O. Cuisenaire, and J.-P. Thiran. Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. *IEEE Trans. Med. Imag.*, 24(12):1548–1565, Dec. 2005.
- B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, Jan. 2002.
- B. Fischl, D. H. Salat, A. J. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23(1):69–84, Sept. 2004.

- Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *J. Comput. Syst. Sci.*, 55(1):119–39, 1997.
- J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Ann. Stat.*, 28(2):337–407, Apr. 2000.
- R. Guillemaud and J. M. Brady. Estimating the bias field of MR images. *IEEE Trans. Med. Imag.*, 16(3):238–251, June 1997.
- X. Han and B. Fischl. Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE Trans. Med. Imag.*, 26(4):479–486, Apr. 2007.
- K. Held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis, and H.-W. Mller-Grtner. Markov random field segmentation of brain MR images *IEEE Trans. Med. Imag.*, 16(6):878–886, Dec. 1997.
- F. Jäger and J. Hornegger. Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging. *IEEE Trans. Med. Imag.*, 28(1):137–150, Jan. 2009.
- M. Jenkinson and S. M. Smith. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.*, 5(2):143–156, June 2001.
- T. Kapur, W. E. L. Grimson, R. Kikinis, and W. M. Wells. Enhanced spatial priors for segmentation of magnetic resonance imagery. In *Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Cambridge, MA, USA, pages 457–468, Oct. 1998.
- Stan Z. Li. *Markov Random Field Modeling in Image Analysis*. Computer Science Workbench. Springer-Verlag Tokyo, Japan, July 2001.
- J. L. Marroquin, B. C. Vemuri, S. Botello, F. Calderon, and A. Fernandez-Bouzas. An accurate and efficient bayesian method for automatic segmentation of brain MRI. *IEEE Trans. Med. Imag.*, 21(8):934–945, Aug. 2002.
- K. M. Pohl, W. M. Wells, A. Guimond, K. Kasai, M. E. Shenton, R. Kikinis, W. E. L. Grimson, and S. K. Warfield. Incorporating non-rigid registration into expectation maximization algorithm to segment MR images. In *Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Tokyo, Japan, pages 564–571, Sept. 2002.
- J. R. Quinlan. Induction of decision trees. *Mach. Learn.*, 1(1):81–106, Mar. 1986.
- D. E. Rex, J. Q. Ma, and A. W. Toga. The LONI pipeline processing environment. *NeuroImage*, 19(3):1033–1048, July 2003.
- B. Scherrer, M. Dojat, F. Forbes, and C. Garbay. LOCUS: LOcal Cooperative Unified Segmentation of MRI brain scans. In *Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Brisbane, Australia, pages 219–227, Nov. 2007.
- B. Scherrer, F. Forbes, C. Garbay, and M. Dojat. Fully Bayesian joint model for MR brain scan and structure segmentation. In *Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, New York, NY, USA, pages 1066–1074, Sept. 2008.
- J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imag.*, 17(1):87–97, Feb. 1998.

- S. M. Smith. Fast robust automated brain extraction. *Hum. Brain Mapp.*, 17(3):143–155, Nov. 2002.
- Z. Song, N. Tustison, B. Avants, and J. Gee. Adaptive graph cuts with tissue priors for brain MRI segmentation. In *IEEE Int. Symp. Biomed. Imag.: Nano To Macro, Arlington, VA*, pages 762–765, Apr. 2006.
- Z. Song, N. Tustison, B. Avants, and J. C. Gee. Integrated graph cuts for brain MRI segmentation. In *Int. Conf. Med. Image Comput. Comput.-Assist. Interv., Copenhagen, Denmark*, pages 831–838, Oct. 2006.
- O. Tsang, A. Gholipour, N. Kehtarnavaz, K. Gopinath, R. Briggs, and I. Panahi. Comparison of tissue segmentation algorithms in neuroimage analysis software tools. In *Int. IEEE EMBS Conf., Vancouver, BC, Canada*, pages 3924–3928, Aug. 2008.
- Z. Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In *IEEE Int. Conf. Comput. Vis., Beijing, China*, pages 1589–1596, Oct. 2005.
- Z. Tu, X. Zhou, D. Comaniciu, and L. Bogoni. A learning based approach for 3D segmentation and colon detagging. In *Eur. Conf. Comput. Vis., Graz, Austria*, pages 436–448, May 2006.
- K. van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based bias field correction of MR images of the brain. *IEEE Trans. Med. Imag.*, 18(10):885–896, Oct. 1999.
- K. van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imag.*, 18(10):897–908, Oct. 1999.
- W. M. Wells. Efficient synthesis of Gaussian filters by cascaded uniform filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-8(2):234–239, March 1986.
- W. M. Wells, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imag.*, 15(4):429–442, Aug. 1996.
- M. Wels, G. Carneiro, A. Aplas, M. Huber, J. Hornegger, and D. Comaniciu. A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3-D MRI. In *Int. Conf. Med. Image Comput. Comput.-Assist. Interv., New York, NY, USA*, pages 67–75, Sept. 2008.
- M. Wels, Y. Zheng, G. Carneiro, M. Huber, J. Hornegger, and D. Comaniciu. Fast and robust 3-D MRI brain structure segmentation. In *Int. Conf. Med. Image Comput. Comput.-Assist. Interv., London, UK*, pages 575–583, Sept. 2009.
- Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imag.*, 20(1):45–57, Jan. 2001.