Parametric estimate of intensity inhomogeneities applied to MRI

Martin Styner, Christian Brechbühler , Gábor Székely and Guido Gerig

Abstract— **This paper presents a new approach to the correction of intensity inhomogeneities in Magnetic Resonance Imaging (MRI) that significantly improves intensity-based tissue segmentation. The distortion of the image brightness values by a low-frequency bias field impedes visual inspection and segmentation. The new correction method called PABIC (PArametric BIas field Correction) is based on a simplified model of the imaging process, a parametric model of tissue class statistics, and a polynomial model of the inhomogeneity field. We assume that the image is composed of pixels assigned to a small number of categories with a priori known statistics. Further we assume that the image is corrupted by noise and a low-frequency inhomogeneity field. The estimation of the parametric bias field is formulated as a non-linear energy minimization problem using an Evolution Strategy. The resulting bias field is independent of the image region configurations and thus overcomes limitations of methods based on homomorphic filtering. Further, PABIC can correct bias distortions much larger than the image contrast. Input parameters are the intensity statistics of the classes and the degree of the polynomial function. The polynomial approach combines bias correction with histogram adjustment, making it well-suited for normalizing the intensity histogram of datasets from serial studies.**

We present simulations and a quantitative validation with phantom and test images. A large number of MR image data acquired with breast, surface and head coils, both in 2D and 3D, have been processed and demonstrate the versatility and robustness of this new bias correction scheme.

Keywords— **MRI, Intensity inhomogeneity, RF field inhomogeneity, Brain tissue segmentation, Nonlinear optimization**

I. Introduction

A major obstacle to segmentation by multiple thresholding or by multivariate statistical classification [1], [2], [3], [4], [5], [6], [7], [8] is insufficient data quality. Besides corruption by noise, the brightness of MR image data is often deficient due to Radio Frequency (RF) field inhomogeneities. This variability of tissue intensity values with respect to image location can severely affect visual evaluation as well as segmentation based on absolute pixel intensities. The effects of such inhomogeneities on the subsequent segmentation in early applications have been discussed by Kohn et al. [9]. Our approach of removing inhomogeneities is image-based and is not applied during scanning, making it suitable for the correction of existing image data in retrospective studies.

The segmentation experiments presented by Jungke et al. [10] clearly illustrate that the major problem was not noise but rather inhomogeneity. Vannier et al. [11] described a correction method based on line-by-line histogram evaluation and a subsequent correction. Another attempt [2] modeled the inhomogeneity as a linear variation in the transverse anatomic plane and computed an approximation to the inhomogeneity as a linear ramp from the neck to the crown. Only vertical (or horizontal) distortions were taken into account, while the actual distortions due to RF field inhomogeneities were often more accurately described by polynomial functions of higher degree. Merickel [3] encountered the problem of the sensitivity profile of the surface coil which falls off with distance from the center of the coil towards the slice borders and along the direction of increasing depth. The intensity gradients were corrected by employing an unsharp masking technique, reducing low frequency surface coil gradients while enhancing the contrast of edges [12]. Lim et al. [4] proposed a smoothing technique to correct the inhomogeneity problem: after extraction of the head contour, the intensity values were extended radially towards the image boundaries and smoothed with a Gaussian filter of a large kernel size. They assumed that the resulting blurred image represents one homogeneous region that is only distorted by the scanner inhomogeneities. The images were corrected with this approximation of the inhomogeneity characteristics. Kohn et al. [9] discuss and illustrate the inhomogeneity problem. They observed that inhomogeneity elongates clusters in feature space in the direction of the origin, but that due to the relative positions of the clusters representing brain and cerebro-spinal fluid, the two classes were still separated. The segmentation of gray and white matter, however, was severely impeded because these two clusters merged.

Homomorphic filtering assumes a separation of the low-frequency bias field from the higher frequencies of the image structures. The assumption is often valid in microscope images of small particles, but can often fail for the structures imaged by MR. A scene, such as a head structure, contains a considerable amount of low-frequency components. Dawant et al. [13] propose a bias correction method relying on user interaction. A user selects typical sample points of a tissue class as input to the estimation of a parametric bias field. Tincher et al. [14] and Meyer et al. [15] present automatic techniques that fit polynomial functions to pre-segmented regions. The individual fits are combined to find an estimate for a global inhomogeneity field. The procedure relies on a preliminary segmentation into region patches.

Wells et al. [16], [17], [18] propose an expectation-maximization (EM) algorithm to achieve an interleaved

Project funding is provided by the Swiss Federal Office for Education and Science (BBW Nr 95.0340).

Martin Styner and Guido Gerig are members of the Department of Computer Science at the University of North Carolina in Chapel Hill, NC 27514, USA. email:martin_styner@ieee.org, gerig@cs.unc.edu

Christian Brechbühler and Gábor Székely are at the Institute of Communication Technology, Image Science Group, ETH Zürich, Switzerland. email:{brech,szekely}@vision.ee.ethz.ch

bias correction/statistical segmentation. In the case of scalar data, the bias estimate $\hat{b}$ is calculated as $\hat{b} = H[Y - WU]$, where H is a low-pass filter, Y the original data and WU a prediction of the signal, which is the sum of the class means weighted by the a posteriori probabilities, $\sum P_c \mu_c$. Wells' formulation includes the bias distortion in the statistical model of the pixel distribution, i.e. the bias field influences the distribution by locally shifting its mean value. The algorithm iterates between two steps, the E-step for calculating the posterior tissue probabilities, and the M-step for estimating the bias field. Wells presented excellent results on double-echo spin-echo data as well as on surface-coil MR images. Our own experiments showed that appropriate initialization is critical. The procedure either has to be initialized with a close estimate of the bias field or with a coarse initial segmentation. Due to low-pass filtering of the residuals, the method performs best in cases where tissue categories represent a spatial frequency pattern much smaller than the frequency of the bias field.

Régis Guillemaud and Michael Brady [19] introduced a modified EM algorithm that replaces the distribution of the class *other*, which includes all tissue not explicitly modeled, by a uniform probability density function. The correction was claimed to be more robust and to overcome some limitations of Wells' original method. They also introduced an automatic estimation of the initial parameters based on a constrained and exhaustive search guided by minimum entropy. Nevertheless, the initialization of the parameters remains critical, as in the original algorithm, and the method is still sensitive to the spatial configuration of image structures.

A correction algorithm fully based on the histogram was proposed by Sled [20], [21]. It requires a parametric model for the bias field but not a decomposition of the intensities into tissue classes. The algorithm assumes that the histogram of an image represents the probability distribution of the given signal. The histogram of the undistorted image is estimated iteratively from the current histogram by estimating a parametric bias field with a histogram sharpening goal function.

We have developed a new approach, based on our own tests and our experience with several inhomogeneity correction schemes. Wells' EM model is based on a probabilistic model in which a bias corruption produces pixels with mixture densities between classes. In contrast, our model is based on the observation that a pixel value of the corrupted image still belongs to a single class, but its value is moved away from the class mean. Therefore, we try to correct for the inhomogeneity by "pushing" each pixel to a value which is near to one of the predefined class means. In contrast to Wells, we do not alternate between image classification and bias estimation, and the estimation of the parameters of a parametric bias field model does not require a preliminary segmentation. The new method PABIC (PArametric Bias field Correction) is driven by the assumption that the idealized scene can be modeled as being composed of pixels that belong to a small number of classes with known statistics. This assumption includes an image model with piecewise constant regions, but it is not restricted to a preferred spatial configuration of image regions with limited spatial frequencies. Our assumptions are closely intertwined with the main goal of bias correction for improved intensity-based segmentation of images into a small number of categories,

Section II provides a discussion of our model of the image formation, the parametric model of the bias field, and the fit of the bias field estimate to the original image according to our assumptions of the imaging process. Section III describes the appropriate settings of the parameters of PABIC and presents tests, validation and applications to 2D and 3D simulated scenes and MRI images from clinical studies.

II. Methods

A. Modeling the imaging process

A typical image data set contains several kinds of tissues or other substances (e.g. cerebral gray and white matter, cerebro-spinal fluid, bone, muscle, fat, air). The ultimate goal of our efforts is the determination of the correct class k for each pixel in the data set, i.e. the *segmentation* of the image data.

The mathematical model of signal formation assumes that each tissue class k has a specific value ρ_k of the property being measured (e.g. proton density). The idealized signal $o(\underline{x})$, therefore, consists of piecewise constant regions, each having one of the values ρ_k. Biological tissues usually have interior structures, which we model with an additive noise term n_{bio}. Characteristics of the measuring device and discrete sampling lead to partial voluming in the blurred border region between tissue classes. This effect can be modeled as a convolution with a small kernel $h(\underline{x})$. The measuring device further corrupts the signal with statistical noise n_{MR} and a systematic bias $b(\underline{x})$. The bias field $b(\underline{x})$ is induced partly by imperfections of the RF field and partly by the individual patient. The bias field is typically very smooth across the whole data set, i.e. it varies slowly and has only very low frequency components. A measured signal $s(\underline{x})$ is formed as follows:

$$s(\underline{x}) \quad = \quad (o(\underline{x}) + n_{bio}(\underline{x})) \star h(\underline{x}) + b(\underline{x}) + n_{MR}(\underline{x}) \;, \quad (1)$$

where $o(\underline{x}) = \mu_{k(\underline{x})}$ and $k(\underline{x})$ is the tissue class found at location $\underline{x}$. Our model of signal formation is visualized in Figure 1.

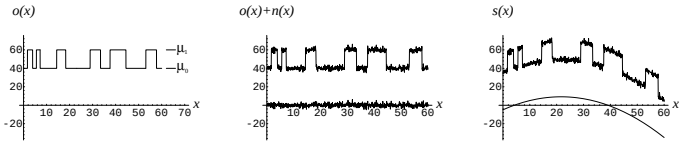

Fig. 1. Signal formation in 1D. The idealized signal $o(x)$ (*left*) is superimposed by noise $n(x)$ (*middle*) and a smooth bias field $b(x)$(*right*, lower curve), yielding the signal $s(x)$ (*right*, upper curve).

For further simplification, $h(\underline{x})$ is neglected, and n_{bio} and n_{MR} are joined into one term, leading to

$$\tilde{s}(\underline{x}) \quad = \quad o(\underline{x}) + b(\underline{x}) + n(\underline{x}) \qquad (2)$$

$$\tilde{s}(\underline{x}) - b(\underline{x}) \quad = \quad o(\underline{x}) + n(\underline{x}) \; . \tag{3}$$

In image acquisition techniques such as MRI, the bias distortion is a multiplicative effect. In such situations, the correction technique described in this paper is to the logarithmic transformed image:

$$\tilde{s}(\underline{x}) \quad = \quad o(\underline{x}) \cdot b(\underline{x}) + n(\underline{x}) \tag{4}$$

$$\log \tilde{s}(\underline{x}) - \log b(\underline{x}) \quad = \quad \log(o(\underline{x}) + \frac{n(\underline{x})}{b(\underline{x})}) \tag{5}$$

The goal of PABIC is a recovery of the idealized image $o(\underline{x})$ in order to obtain an intensity-based segmentation of the scene. Bias correcting an image $\tilde{s}$ means calculating an estimate $\hat{b}(\underline{x})$ of the effective bias field $b(\underline{x})$ and removing the calculated bias field $\hat{b}(\underline{x})$ from the measured image $\tilde{s}$.

We can see from Equation (4) that, due to multiplication and assuming a constant category, the signal-to-noise ratio $o(\underline{x}) \cdot b(\underline{x}) / \sigma^2_{n(\underline{x})}$ becomes smaller in regions with a low bias field than in those with a large bias field. Bias correction by division (5) achieves a common brightness level for these regions, but it results in noise enhancement in dark areas in comparison to bright areas.

B. Bias field estimation by a smooth parametric bias field

We propose to estimate the bias field $b(\underline{x})$ by a parametric model. Because the bias field $b(\underline{x})$ is usually very smooth across the whole data set, we assume that $b(\underline{x})$ belongs to a family of smooth functions. We have chosen $b(\underline{x})$ to be a linear combination of m smooth basis functions $f_i(\underline{x})$. The parameter vector $\underline{p}$ of length m specifies one such linear combination, an *estimated* bias field $\hat{b}$ (Equation (7)). In 2D, we selected products of Legendre polynomials P in x and y as the basis functions f_{ij} for the estimated bias field. In 3D, Legendre polynomials in x, y and z are used. Given detailed information regarding the bias inhomogeneities of a specific imaging system, the Legendre polynomials could be replaced by an alternative parametric model. The image coordinates x_j are scaled to the range $[-1, 1]$. For Legendre polynomials up to the degree l, the size m of the parameter vector $\underline{p}$ is given for the 2D case by $m = (l+1)\frac{(l+2)}{2}$ and for the 3D case by $m = (l+1)\frac{(l+2)}{2}\frac{(l+3)}{3}$. For instance, Legendre polynomials up to the third degree in 3D would therefore require 20 coefficients for $\underline{p}$. The choice of the maximal degree of Legendre polynomials largely depends on prior knowledge of the coil and the expected type and smoothness of the bias field. The bias field estimate $\hat{b}$ is determined as follows:

$$\text{in 2D} : \hat{b}(\underline{x}, \underline{p}) = \sum_{i=0}^{m-1} p_i f_i(\underline{x}) = \sum_{i=0}^{l} \sum_{j=0}^{l-i} p_{ij} P_i(x) P_j(y) \tag{6}$$

$$\text{in 3D} : \hat{b}(\underline{x}, \underline{p}) = \sum_{i=0}^{l} \sum_{j=0}^{(l-i)} \sum_{k=0}^{(l-i-j)} p_{ijk} P_i(x) P_j(y) P_k(z) \tag{7}$$

with $P_i(.)$ denoting a Legendre polynomial of degree i (cf. Appendix V-A).

If $\hat{b}(\underline{x}, \underline{p})$ is a sufficient estimate of the real bias field $b(\underline{x})$, we subtract it from the measured signal and, using (2), define the *bias corrected* image $\hat{o}(\underline{x})$ as

$$\begin{aligned} \hat{o}(\underline{x}) \quad &= \quad \tilde{s}(\underline{x}) - \hat{b}(\underline{x}, \underline{p}) \\ &= \quad o(\underline{x}) + b(\underline{x}) - \hat{b}(\underline{x}, \underline{p}) + n(\underline{x}) \qquad (8) \\ &\approx \quad o(\underline{x}) + n(\underline{x}) \; . \qquad (9) \end{aligned}$$

Following the assumption of images being composed of pixels from a small number of categories, we expect the values of the corrected image to be close to the mean of one of the classes. For the case of a multiplicative bias corruption we divide the measured signal by the estimated bias field, i.e. a logarithmic transformed signal is subtracted (see Equation (12)). This results in a bias corrected signal $\log(\hat{o})$, which is only degraded by noise. Because of the division, the statistics of the noise varies across the spatial domain.

$$\begin{aligned} \tilde{s}(\underline{x}) / \hat{b}(\underline{x}, \underline{p}) \quad &= \quad (o(\underline{x}) \cdot b(\underline{x}) + n(\underline{x})) \, / \hat{b}(\underline{x}, \underline{p}) \\ &\approx \quad o(\underline{x}) + n(\underline{x}) / \hat{b}(\underline{x}, \underline{p}) \qquad (10) \\ \log(\tilde{s}(\underline{x})) - \log(\hat{b}(\underline{x}, \underline{p})) \quad &= \quad \log(o(\underline{x}) + n(\underline{x}) / \hat{b}(\underline{x}, \underline{p})) \qquad (11) \\ &= \quad \log(\hat{o}(\underline{x})) \; . \qquad (12) \end{aligned}$$

The appropriateness of choosing a smooth, parametric bias field model for MR applications is discussed in Appendix V-C. In this section, the model is validated by estimating the bias field of a MRI phantom image with known ground truth.

C. Estimating parameters of a single class fitting

Our model of the imaging process assumes the ideal case of piecewise constant regions or pixels assigned to one of a few categories, which is violated by effects like partial voluming. Therefore, the estimation of classes needs to be based on robust estimators, and minor violations of the model should not result in incorrect solutions.

Starting from the well-known *least squares* technique, we derive a measure for fitting a parameterized single-class model $v_i(\underline{p})$ to data u_i.

$$e(\underline{p}) \quad = \quad \sum_i (u_i - v_i(\underline{p}))^2 \; = \; \sum_i square(d_i) \tag{13}$$

To get a more robust estimation, i.d. an estimation that is less sensitive to outliers, we considered various functions besides "square" (see Figure 2). The "valley" function described below belongs to the family of M-estimators, which are robust estimators ([22], [23]).

$$valley(d) \quad = \quad \frac{d^2}{d^2 + 3\sigma^2} \tag{14}$$

This function shows inflection points at $d \pm \sigma$. Outliers have less influence on the position of the minimum if "valley" is used instead of "square". Supplying $(d - \mu)$ as an argument will shift the position of the minimum from 0 to μ. An energy function of the form $e(\underline{p}) = \sum_i valley\left(u_i - v_i\left(\underline{p}\right)\right)$ will yield a small value if the model v is well fitted to the data samples u.

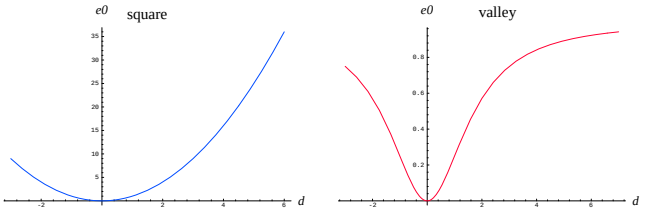

Fig. 2. Left: square function. Right: valley-shaped function.

D. Estimating parameters of a multi-class fitting

PABIC estimates the bias field directly from image data. No image contrast is required to estimate this field. A contrast between different image regions, however, should not disturb the bias estimation, i.e. the different brightness levels of regions should be ruled out by the procedure.

We assume that the images are composed of regions with piecewise constant intensities with mean values μ_k. Each pixel of the idealized signal corrupted by noise $o(\underline{x}) + n(\underline{x})$ must take values close to one of these class means. Therefore a multi-class energy function $e_0(d)$ must have a minimum at each class mean μ_k.

We define for each class k a "valley"-function with the minimum located at the class mean μ_k. The "valley" functions of all classes are multiplied to form the energy function $e_0(\hat{d}) = \prod_k valley(\hat{d} - \mu_k)$ (Figure 3). Bias corrected pixels which are close to a μ_k have a value close to zero for the product over all classes k and therefore give a small contribution to the total energy $e_{tot\_image}(\underline{p})$.

$$e_{tot} = \sum_i e_0(\hat{d}_i) = \sum_i \prod_k valley(\hat{d}_i - \mu_k) \quad (15)$$

$$e_{tot\_img}(\underline{p}) = \sum_{\underline{x} \in img} \prod_k valley\left(\tilde{s}(\underline{x}) - \hat{b}(\underline{x}, \underline{p}) - \mu_k\right) \quad (16)$$

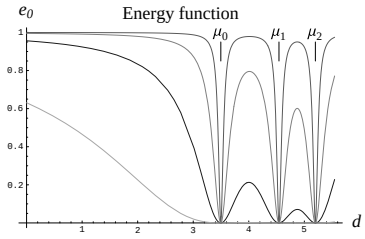

Fig. 3. Multi-valued energy function. The class values are chosen as $\mu_0 = \log 33 = 3.49$, $\mu_1 = \log 94 = 4.54$, and $\mu_2 = \log 181 = 5.20$ with $\sigma_1 = \sigma_2 = \sigma_3 = 0.03, 0.1, 0.3$, and 1.0, from top to bottom respectively.

E. Multi-class model and partial volume effect

As stated before, our model of the imaging process assumes the ideal case that every pixel can be assigned to one of a few categories. But in medical imaging, the partial volume effect violates this assumption, since partial volume voxels have an intensity that is composed of multiple class intensities. Low resolution datasets can have as much as 30% of partial volume voxels. If such a high amount of partial volume voxels is present, the calculation of the bias field can be impeded and the calculated bias field can be partially incorrect. In high resolution images, the robust estimator used for our class model can cope with outliers like partial volume voxels.

In order to robustize the bias field correction to account for the large partial volume effect, we apply a preprocessing step to partially mask out these voxels. The masking method assumes that partial volume voxels mainly occur at the borders of class regions. The gradient magnitude of a slightly blurred image is calculated, and voxels with large gradient magnitude are considered as partial voluming voxels and masked out.

F. Assumptions of PABIC

To summarize, PABIC uses the following assumptions:

- A *smooth* bias field $b(\underline{x})$ is partly caused by imperfections of the RF field and partly induced by the patient's influence on the magnetic and electric fields.
- The bias field $b(\underline{x})$ can be approximated by a parametric model $\hat{b}(\underline{x}, \underline{p})$, which is chosen as the superposition of Legendre polynomials in the actual implementation .
- The appearance of a class k in the image can be modeled by a mean intensity μ_k and a variance σ_k^2.
- The hypothetical, idealized signal $o(\underline{x})$ consists of pixels $\underline{x}$, each of which can be assigned to a unique tissue class k with mean value μ_k, which includes images composed of piecewise constant regions with values μ_k.
- Noise $n(\underline{x})$ is caused by the granularity and interior structures of biological tissues and the measuring device.

G. Analysis of the high-dimensional parameter space

The components of the parameter vector $\underline{p}$ stretch a multi-dimensional feature space $e(\underline{p})$. The dimensionality of this feature space is $m = (l+1)\frac{(l+2)}{2}$ for 2D and $m = (l+1)\frac{(l+2)}{2}\frac{(l+3)}{3}$ for 3D applications, with l as the maximum degree of Legendre polynomials used to approximate the bias field. Visualization of the energy landscape $e(\underline{p})$ is very difficult because of the high dimension m. In order to evaluate the topography of the high-dimensional energy function at a specific location, we calculate the eigenvectors of the local Hessian matrix and plot the energy values along each eigenvector. Both, local and global minima are minimal in each eigenvector direction.

We performed a detailed analysis of a 2D step-edge image representing a simple two-class image. This image structure was chosen to demonstrate that our algorithm is independent of the relationship between spatial frequencies of the bias field and the image structures. Correction algorithms based on homomorphic filtering would fail in such cases. The image was distorted by various types of bias fields and different noise levels (see Figure 4). Even for a simply shaped bias field and no noise, we observed many non-optimal minima. Therefore, finding the global minimum requires an optimization technique offering a mechanism to step out of non-optimal minima reliably.

To our surprise the energy function is always smooth,

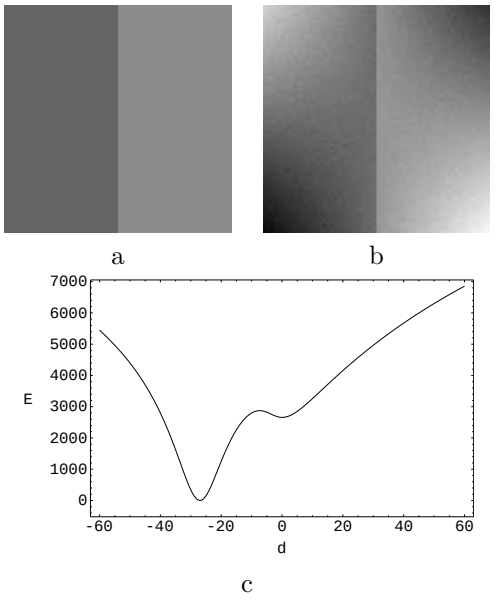

Fig. 4. Images of a step-edge function along the x-axis. a: No bias field, no noise. b: Bias field $x \cdot y$, Gaussian noise ($\sigma = contrast/4$). d: bias field as in c, Plot of the energy function along the line connecting the global and the smallest non-optimal minimum.

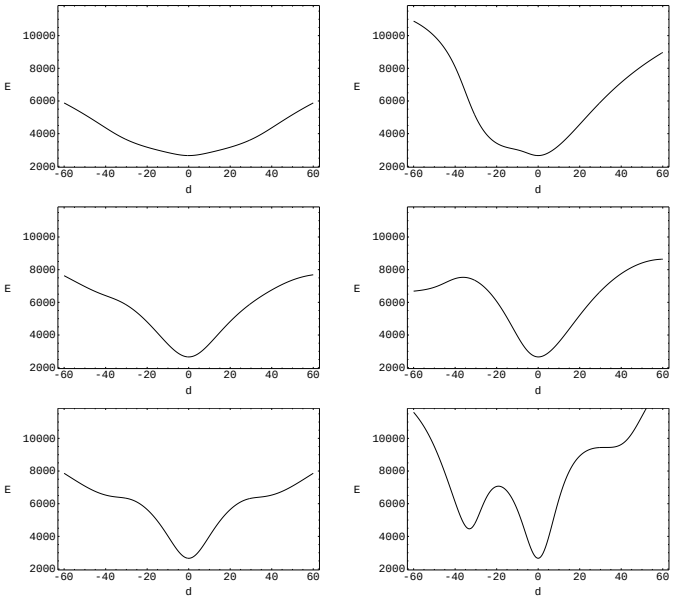

Fig. 5. Plots of the energy function of a bias distorted step-edge image with Gaussian noise ($\sigma = contrast/4$). Plots were obtained along the eigenvectors of the Hessian matrix at a non-optimal minimum in parameter space and sorted by decreasing eigenvalues. From the plots alone it cannot be concluded whether this minimum is global or not. Further, the energy function is smooth at any location.

no matter how high the chosen noise level (see Figure 5). This leads to the hypothesis that, within certain limits, the noise level does not severely affect the result of the bias correction. Our tests supported this hypothesis, as long as the noise level was lower than the contrast between neighboring classes. Images with higher noise levels require a preceding noise reduction step, e.g. one with nonlinear filtering.

H. Bias field parameter estimation by nonlinear optimization

Finding the parameter vector $\underline{p}$ with minimum energy $e(\underline{p})$ is a nonlinear optimization problem, independent of the type of bias field and energy function. In principle, any nonlinear optimization method could be applied.

An early version of PABIC used a discrete taboo search (TS) technique ([24],[25]). The major advantage of TS is its ability to step out of non-optimal minima. However, TS does not have an automatic step size adaptation and restricts the search directions only to directions parallel to the parameter axes, resulting in a poor convergence rate.

To improve optimization we developed a new method that locally adjusts search direction and step size, and provides a mechanism to step out of non-optimal minima. Furthermore, the method is fast enough to cope with our large data sets and overcomes the problem of parameters with different scaling. The developed algorithm uses a so called $(1+1)$-Evolution Strategy (ES), which belongs to the family of Evolutionary Algorithms (for an introduction see [26] and [27]).

In an Evolution Strategy, a parameter vector $\underline{p}$ represents an individual. The fitness of an individual is determined by its energy value $e(\underline{p})$. In our case, small energy values yield a high fitness and vice versa. The existing individuals form the population at the beginning of each optimization step. As a result of an optimization step, new individuals ("children") are generated by some sort of mutation of the population("parents"). The children are added to the population, which is then reduced back to the former size as the fittest individuals survive. In the $(1 + 1)$-ES strategy both the population size and the number of children generated are equal to 1. A mutation is a random vector of the multi-dimensional normal distribution with mean $\mu = \underline{p}_{parent}$ and covariance matrix Σ^2. The covariance matrix Σ^2 gets adapted at each optimization step. It is increased by a factor c_{grow} if the new population consists of fitter individuals, and it is reduced by a factor c_{shrink} if otherwise. Details of the formalism of the $(1 + 1)$-ES algorithm are described in the Appendix in Section V-B.

III. Applications

A. Parameter setting

The parameters of the bias correction can be divided into three groups: Parameters for the class model, the bias field model and for the optimization. This paper gives a short description of the settings of the parameters. For a detailed evaluation, please refer to [28].

Class parameters

- Number of classes: The number of classes is given by the number of dominant categories in the image scene; well-separated by intensity differences. Categories of low occurrence only give small contributions to the bias field and

can be neglected. If a dominant category is omitted from the model, the optimization will result in a non-optimal solution.

- Mean and Variances of a class: An inappropriate choice of the mean values and variances can lead to non-optimal solutions. Therefore, PABIC is sensitive regarding the choice of class statistics, but experiences with hundreds of datasets has shown that PABIC runs stable regarding minor changes of the class statistics.
- Class-mask: For each image a *coarse* class-mask must be defined. The class-mask separates the regions that are well-modeled by the class probability density functions from the rest of the image. This masks is usually generated by a coarse segmentation clustering all modeled classes into one main category. PABIC can handle class-masks that are very coarse and contain several misclassified regions. Experiments indicate that PABIC is not sensitive to segmentation errors smaller than 20%.

Bias estimation parameters

- Type of bias field: Images obtained with a MRI scanner usually have a multiplicative bias field, but for other applications the bias field may be additive.
- Order of Legendre polynomials: The maximum degree of Legendre polynomials determines the accuracy and stability of the calculated bias field. Choosing a too-large degree results in inefficient computation time, instable coefficients, and the calculated bias field is likely to adapt itself to anatomical structures. When choosing a too-low degree, PABIC will only be able to remove partially the bias field. Our experiments showed that, for MRI with head-coils, Legendre polynomials up to the third degree sufficiently model the distortion. For 3D MR mammography we need higher order polynomials, which encourages the future use of a more appropriate parametric model than Legendre polynomials.

Optimization parameters

- Mutation factors c_{grow} and c_{shrink}: The choice of the parameters c_{grow} and c_{shrink} determines the efficiency and the correctness of the optimization. Generally, the lower the choice of c_{grow} the more probable the optimization finds the optimal solution; the higher the c_{grow}, the higher the convergence rate. A good choice for c_{shrink} is $c_{grow}^{-1/4}$ (see [26]) and for c_{grow} a value between 1.01 and 1.1.
- Maximum number of iterations: The major criterion is a good approximation of the global bias function and not a determination of a set of parameters to a very high precision. This allows us to limit the number of iterations to a reasonable value.
- Initial value of coefficients $\underline{p}$: This parameter is not critical, but a good initial choice will speed up the calculation.

B. Validation of the calculated bias field

The validation of the resulting bias field can be done with several methods, using either a visual assessment of the quality of the correction or a quantitative analysis by comparison to a gold standard.

- The energy image of the bias field represents the value of the energy function at each pixel of the bias corrected image (see Figure 6). The energy image should show a low energy distribution in the whole image, except in those parts of the image that are not modeled by the class statistics (e.g. the background). High energy values may also exist at the boundaries between regions due to partial voluming effects.
- A coarse segmentation of the corrected image can be performed by simple thresholding; for example, by selecting only one single class k with mean value μ_k. Poor homogeneity of the segmented region would indicate that the bias field was calculated incorrectly.

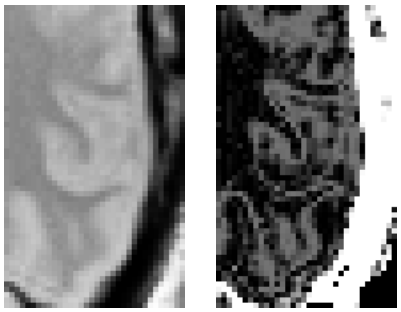

Fig. 6. Zoomed detail (left) of the bias corrected image and its corresponding energy image (right) of a human head surface coil image. The energy is low (dark) in homogeneous regions and higher (bright) at the borders between regions included in the tissue model. The skin and csf were not included in the class-model but in the class-mask. They are represented by regions of maximal energy.

C. Synthetic test images in 2D and 3D

Four simulated image restoration problems were examined using different bias fields up to the fourth degree and various noise levels. The first problem was the correction of a bias distorted step-edge image (see also Section II-G). In the second problem the step image was replaced by a checkerboard (Figure 7). The step-edge test was performed only in 2D, and the checkerboard tests in 2D and in 3D. Additionally, a synthetic 3D onion-like object was constructed, which consisted of multiple layers of different intensities. The Mean Squared Differences (MSD) of the synthetic to the estimated bias fields were calculated directly from the coefficient vectors (see Appendix V-A). At least 5 different experiments were carried out for each noise level and each synthetic test image, resulting in a maximum and average deviation, which are illustrated in Table I. In all synthetic tests, the calculated bias field matched the modeled field to a very high accuracy, i.e. the deviation $\sqrt{MSD}$ was very low. The major difference between the 3D tests and the 2D tests was increased computation time and a larger number of parameters.

D. Simulated MRI images

The simulated brain MRI image database (SBD) of Montréal Neurological Institute (MNI) at the McGill University was used to validate and to test the sensitivity of PABIC [29], [30], [31]. In this database there is one

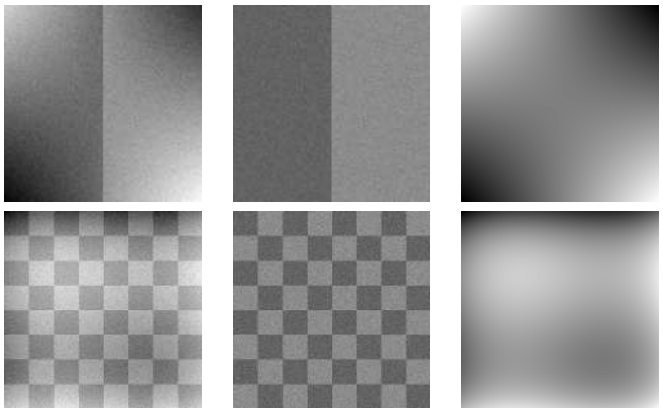

Fig. 7. Bias field correction of a step-edge (top row) and a checkerboard (bottom row) (Gaussian noise $\sigma = 40\%contrast$). Left: Noisy, distorted image. Middle: Corrected image. Right: Calculated bias field.

TABLE I

Several multiplicative bias fields and Gaussian noise (amplitude in percentage of the contrast) were applied to a set of synthetic two-class test images. The table shows the maximum and average $\sqrt{MSD}$ in percentage of the contrast.

Image	Noise $n_1 = 0\%$		Noise $n_2 = 40\%$	
	d_{avg}	d_{max}	d_{avg}	d_{max}
Step edge	0.0000311	0.0000472	0.148947	0.253286
Checkerboard	0.0005471	0.001972	0.051534	0.067458
3d onion	0.0000615	0.0001090	0.619582	1.68283
	Noise $n_3 = 80\%$		Noise $n_4 = 120\%$	
Step edge	0.65116	0.86723	3.98527	6.82645
Checkerboard	0.162641	0.25067	0.873322	1.55208
3d onion	1.6433	4.04466	-	-

isotropic atlas dataset of a human head with corresponding ground truth for the segmentation of various tissues, including white and gray matter. Based on this atlas, MRI images of the three protocols PD, T2 and T1 are simulated. Additionally, noise of different levels can be added, and there is a choice between three levels of bias distortions. We applied the same sequence of image processing steps as in the case of processing real MR images, which includes a preprocessing to reduce noise using a locally adaptive anisotropic filtering method.

The first validation test series was created by bias-distortion of noisy images (3 MR protocols, 5 different noise levels) with 20 different parameterized bias fields simulated by randomized parameter vectors. The correction was performed with a two-class model using a white and a gray matter class. The class parameters were computed by fitting two Gaussians to the histogram of the masked dataset. The mask was determined with a brain-masking method using thresholding, mathematical morphology operations and removal of partial volume voxels (see II-E). After applying PABIC, the Mean Squared Differences between the synthetic and the estimated bias fields were calculated directly from the parameter vectors (see Appendix V-A). The average and maximal $\sqrt{MSD}$'s, normalized to the contrast between white and gray matter, are shown in Figure II. For all test cases the deviation $\sqrt{MSD}$ turned out to be very small.

TABLE II

5 different noise levels (amplitude as percentage of the intensity range) and for each noise level 20 random bias fields were applied to simulated MRI datasets of the MNI-SBD-database. The table shows the maximum and average $\sqrt{MSD}$ in percentage of the contrast between white matter and gray matter.

Protocol	Noise $n_1 = 0\%$		Noise $n_2 = 3\%$	
	d_{avg}	d_{max}	d_{avg}	d_{max}
PD	0.05236	0.17399	0.44090	0.59771
T2	0.05902	0.10983	0.70192	0.74620
T1	0.03007	0.08011	0.44802	0.58714
	Noise $n_3 = 5\%$		Noise $n_4 = 7\%$	
PD	0.72606	0.84473	0.93939	1.21379
T2	0.99343	1.08403	2.32724	2.37639
T1	0.71379	0.83062	0.436402	0.50541
	Noise $n_5 = 9\%$			
PD	0.91545	0.93876		
T2	2.88804	3.22135		
T1	0.56793	0.60492		

In the second validation test series, simulated SBD datasets including unknown bias fields distortion at 3 different levels were corrected (see Figure 8). The class parameters and masks were determined as in the first validation series. The estimated bias fields were composed of Legendre polynomials up to the third degree. All the corrections were additionally performed using polynomials up to the second degree. The resulting bias fields and segmentation errors varied less than 1% between Legendre polynomials up to the third and second degree. All tests used a subsampling factor of 2 for each dimension to speed up calculations.

Visualization of intensity-isolines in the uncorrected and corrected images (see Figure 9) was created to demonstrate the effect of the bias field on an intensity-based segmentation. The intensity-isolines were chosen at an intermediate level between the intensities of gray and white matter, thus the isolines represent the dividing line between gray matter and white matter of a segmentation based on thresholding at this level. The bias-distorted dataset clearly shows an incorrect segmentation of the gray-white matter boundary.

The validation of the correction was mainly done by calculating the error of the white and gray matter segmentations using a simple thresholding technique. For all white matter (WM) and gray matter (GM) segmentations, the Alpha (False-Positive) and Beta (True-Negative) errors were computed in relation to the amount of corresponding true class voxels, e.g. the Beta error of white matter was calculated as $\beta_{WM} = \frac{\sum \text{voxels wrongly classified as WM}}{\sum \text{WM voxels in ground truth}}$. The overall segmentation error was computed as the maximal value of the 4 segmentation errors $\{\alpha_{WM}, \beta_{WM}, \alpha_{GM}, \beta_{GM}\}$. The minimal overall segmentation error for white and gray matter

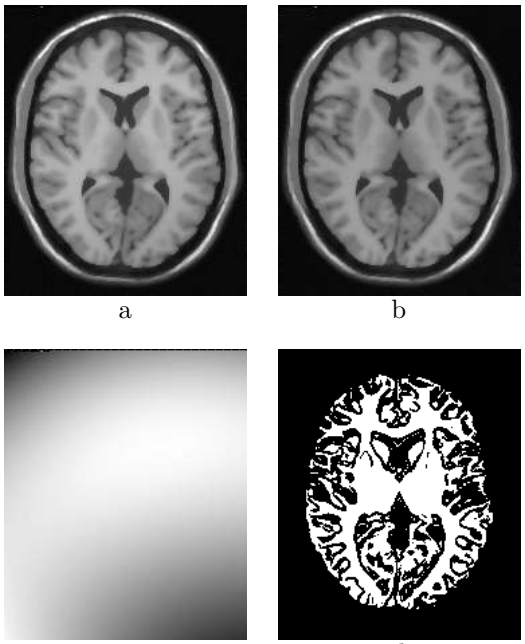

Fig. 8. Bias field correction of a simulated 3D MRI head image of the MNI-SBD-database (classes white and gray matter; Legendre polynomials up to third degree). a: Slice in the original image. b: Corrected image. c: Computed bias field. d: Class-mask for brain tissue and partial voluming voxels,

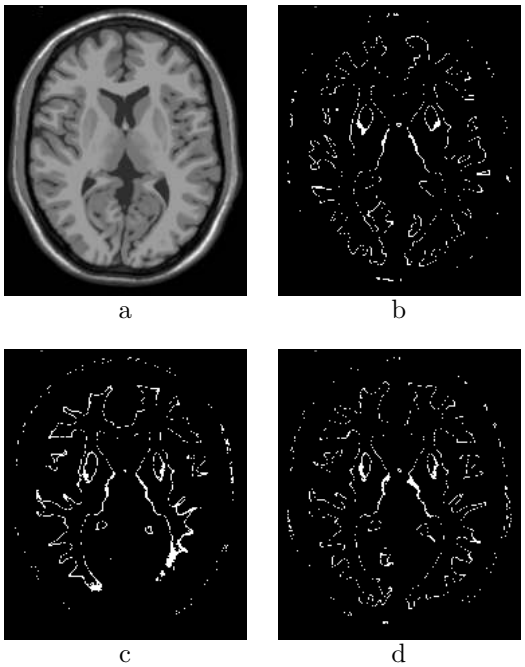

Fig. 9. Bias field correction of a simulated 3D MRI head image of the MNI-SBD-database. Visualization of intensity-isolines in a slice (a) of the undistorted base image (b), the distorted image (c) and the bias corrected image (d). The incorrect course of the gray-white matter boundary in the distorted image is clearly visible.

segmentation on the noiseless, undistorted images is not zero as illustrated in Figure 10. The reason is overlapping intensity distribution clusters with the inherent probability of misclassification errors for any thresholding technique. Additionally, the resulting energy image was visually checked for possible spatial accumulations of high energy.

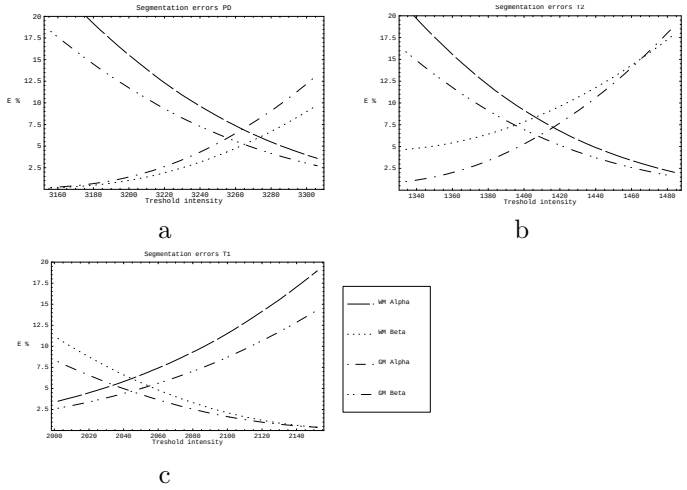

Fig. 10. Alpha/False-Positive and Beta/True-Negative segmentation errors of the white matter and gray matter segmentation of the PD (a), T2 (b) and T1 (c) images of the MNI-SBD-database (no noise, no bias field) depending on the threshold value. Clearly it can be seen that a segmentation error of 0% cannot be achieved for all 4 errors simultaneously.

PABIC was applied to images of all 3 protocols with 3 different levels of distortion and 5 different noise levels, a total of 45 datasets. All corrections were successful, but 2 datasets (PD and T2, both noise level 9%, bias level 40%) needed an additional recalculation. The segmentation errors of all corrected datasets are shown in Figure 11. For all test cases the segmentation errors were small and stable in regard to noise. The increased error in the noiseless T2 image is an effect of the anisotropic diffusion filter that was part of the standardized segmentation protocol.

E. Real 2D and 3D MRI images

All the following correction problems were preprocessed with a small anisotropic diffusion filter to reduce noise.

E.1 2D MRI images

The first test with real MRI was performed on 2D images of a phantom with moderate bias field. It served mainly for model validation and is discussed in detail in the Appendix V-C.

The second test was performed on a 2D sagittal image of a human head scanned with a head surface coil. Since the sensitivity of the head surface coil decreases approximately quadratic with distance from the coil, we modeled the bias field using Legendre polynomials up to the second degree. The correction used the classes white and gray matter; other image data was masked out using a coarse mask generated by manual segmentation. In the original

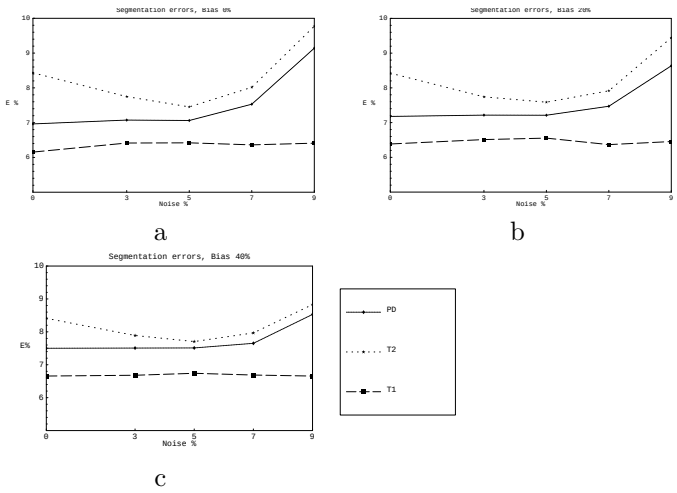

Fig. 11. Maximal segmentation errors of bias corrected images of the MNI-SBD-database for PD, T2 and T1, and bias levels 0% (a), 20% (b), 40% (c).

image (Figure 12) an adequate segmentation is extremely difficult; the pixel values of the different tissues strongly overlap. In the corrected image this is not the case, so it is more suitable for visual analysis and computer-assisted processing.

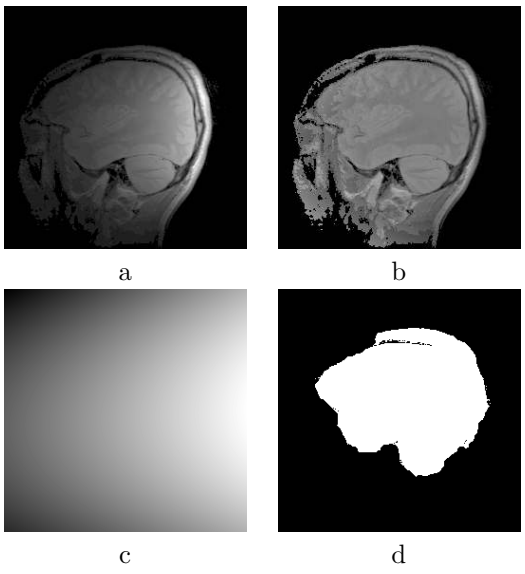

Fig. 12. Bias field correction of a sagittal head MRI image (classes white and gray matter; Legendre polynomials up to second degree): a: Original image. b: Corrected image. c: Computed bias field. d: Class-mask

Another test was performed on a breast image obtained by MRI mammography. MRI mammography uses C-coils. The uncorrected mammography images are very difficult to interpret (Figure 13). The correction was done with Legendre polynomials up to the third degree with one class for the fat tissue of the breast and another class for the mammary ducts glands and connective tissue. A mask generated by manual segmentation masks out other image data. Compared to the original image, the bias corrected image provides details throughout the whole breast and the pixel values of different tissues no longer overlap.

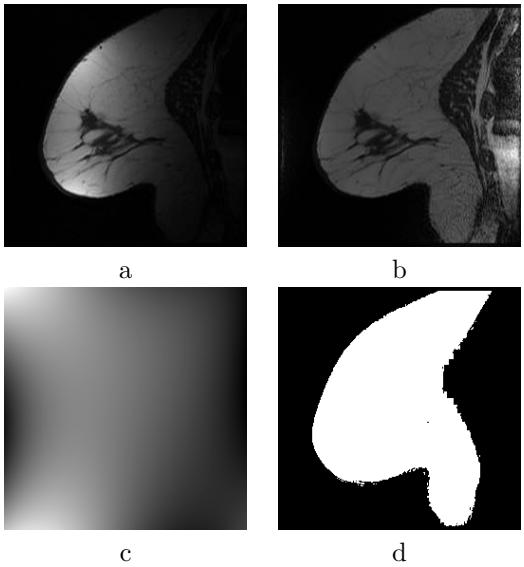

Fig. 13. Bias field correction of a breast MRI image (classes fat tissue and mammary ducts glands and connective tissue; Legendre polynomials up to third degree). a : Original image. b: Corrected image. c: Computed bias field. d: Class-mask

E.2 3D MRI images

3D bias correction has been applied to 3D head coil MR images from several clinical studies. Here we present corrections of serial image data in a Multiple Sclerosis study. Two time series (proton density and T2-weighted) were acquired at 12 different time points for each patient. These images were all obtained using an interleaved 2D acquisition protocol with high inter-slice intensity variations. These inter-slice intensity changes were removed in a preprocessing step by applying a 2D 0^{th}-order bias correction/histogram adjustment (see III-E.3) to each slice.

Two classes white and the gray matter were chosen for modeling brain tissue. Other tissues were masked out with brain masks generated by a coarse statistical classification based on a training set from one slice for each volume. The geometry of the head coil suggests the use of Legendre polynomials up to the second degree.

Visual inspection reveals no noticeable difference between the uncorrected and the corrected volumes (Figure 14), because the detected distortions are small. The importance of bias correction for a subsequent tissue segmentation is demonstrated in Figure 15 by comparing the values of the voxels along a line through white matter in an uncorrected and a corrected volume. The direction of the line is approximately perpendicular to the isosurfaces of the local bias field.

To demonstrate further the effect of the bias correction, we performed simple thresholding to segment white and gray matter. The threshold values were adapted manually for each volume to produce the best result (see Figure 16). The segmentation in the uncorrected volume contained

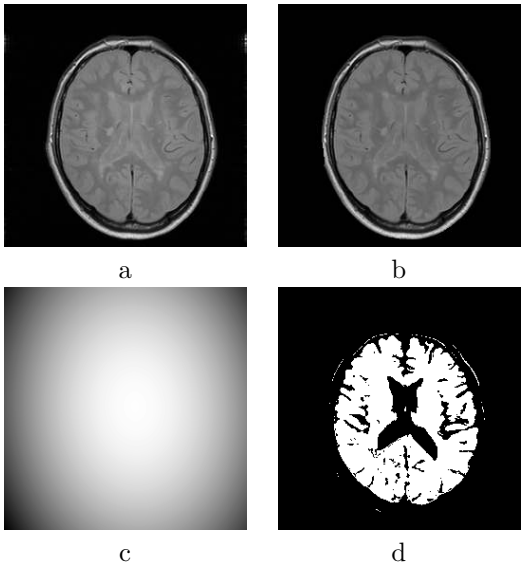

Fig. 14. Correction of a 3D MRI dataset taken from a Multiple Sclerosis study using spin-echo dual-echo imaging. A single slice of the first echo is shown (classes white and gray matter; Legendre polynomials up to the second degree). a: Uncorrected volume, b: Corrected volume, c: Calculated bias field. d: Class-mask. The bias correction is small but significantly improves gray/white matter separation (see Figs 15, 16).

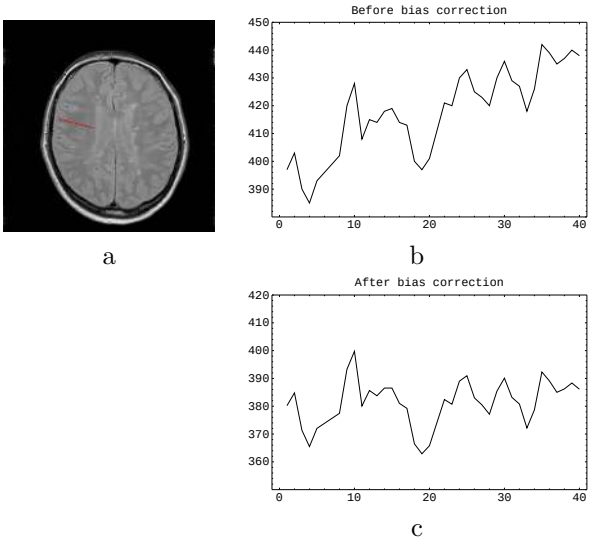

Fig. 15. Correction of a 3D MRI dataset taken from a Multiple Sclerosis study using spin-echo dual-echo imaging: Plots of the voxel values along a line in white matter. a: Slice of the uncorrected volume with plot line, b: Plot of the uncorrected volume, c: Plot of the corrected volume (c).

misclassified regions, and boundaries were poorly delimited. Besides misclassification due to small MS-lesions in the image, the segmentation in the corrected volume was much better, since there were fewer misclassified regions, well-delineated boundaries and a more homogeneous appearance. Consider the slices at medium level shown in Figure 16. Large parts of the ventricles in the uncorrected image are classified as gray matter and the deep gray matter is not classified as gray matter. These misclassified regions were labeled correctly after bias correction.

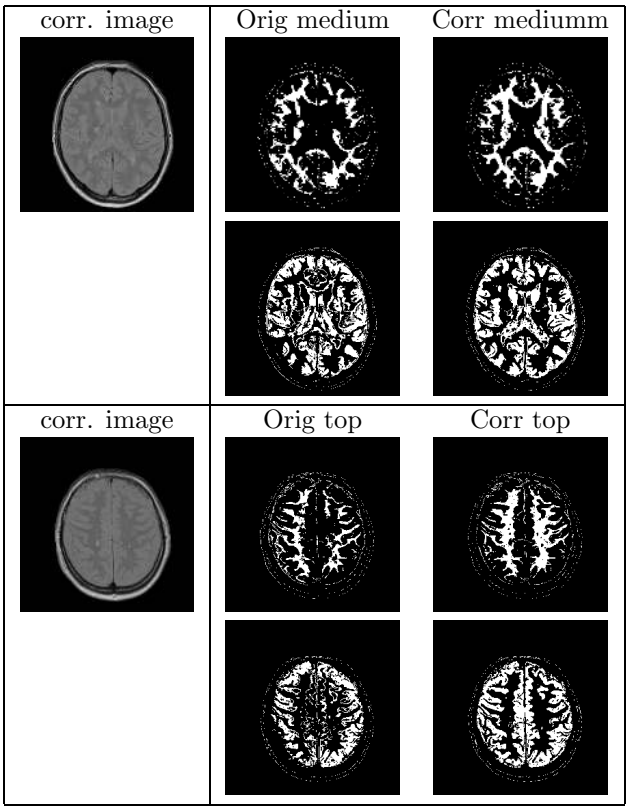

Fig. 16. 3D Segmentation (MS PD time series) by simple thresholding of white matter (WM, first row) and gray matter (GM, second row) shown on axial slices at medium and top level. The corresponding gray value image is displayed on the far left. Orig: Segmentation of the uncorrected volume. Corr: Segmentation of the corrected volume.

After these initial tests, PABIC has been applied routinely in various large clinical studies using MRI imaging. It is a necessary preprocessing step for a white matter/gray matter segmentation method applied to image data in several Schizophrenia studies. One of these studies was directed by the European BIOMORPH consortium, in which 40 Schizophrenia and 40 control datasets (SPGR, high resolution) were successfully corrected using PABIC. Generally, in newer studies that are based on high resolution single contrast acquisition protocols (gradient echo protocols like SPGR, IRprepped, Turboflash, e.g.), the observed bias distortions were small, which is due to improved scanner techniques and built-in correction methods. However, bias correction is still necessary for correcting patient induced inhomogeneities and the natural variability of tissue intensities in the spatial domain.

E.3 4D MRI images - Time series analysis - Histogram adjustment

The study of temporal intensity changes requires bias correction, registration, and normalization of tissue intensities along the time axis. Initially, we performed two time series analysis on the earlier mentioned Multiple Sclerosis datasets. After registration using the MIRIT-software [32], the datasets were transformed into a normalized 4D image data set. Intensity normalization was achieved by applying PABIC with the same tissue model, i.e. the same mean values and variances for the tissue classes, for all datasets in the time series. PABIC has since been applied routinely for bias correction and histogram adjustment in MR time series, for example, it is a required preprocessing step for a method developed at ETH Zürich to segment and characterize Multiple Sclerosis lesions based on dynamic changes over time [33]. The segmentation accuracy was significantly increased and we were able to perform comparisons across time and space. In total, PABIC has been applied to over 150 high- and low-resolution 3D MR datasets stemming from Multiple Sclerosis studies.

IV. Conclusions and Discussion

The bias correction method PABIC presented in this paper estimates the parameters of a 2D/3D inhomogeneity field by nonlinear optimization. PABIC is based on clear assumptions about the imaging process and about the observed scene. The procedure does not rely on a presegmented image and operates independently of the spatial configuration of structures in the image. The pixels can even be scrambled arbitrarily to achieve the same bias estimation. This has been demonstrated with the synthetic images of a checkerboard structure and a single step-edge showing that they are treated exactly the same way. The latter image cannot be corrected by homomorphic filtering nor by methods which need an initial close segmentation. Further, PABIC can handle bias distortions much larger than the range of class means.

The estimation of a parametric bias field offers several advantages. First, the estimate can be constrained by selecting the degree of the polynomial function and the range of the parameter values. Secondly, prior knowledge about the type of inhomogeneity allows choosing the appropriate parametric model. The Legendre polynomials applied in our implementation are just one of various possibilities. In the MR mammography image (Figure 13), for example, one could think of using a mixture of two Gaussian simulating the effect of the C-coil instead of a higher order polynomial function.

As a coarse summary, PABIC can be viewed as an optimization of polynomial parameters with respect to a cost function based on an image classification likelihood.

Extensive applications have been performed on 2D datasets as well as on 3D datasets (more than 250 scans in total). The bias correction procedure runs stably and can be automated to a certain degree. The quality of brain tissue segmentation by intensity thresholding has been significantly improved. The normalization of the intensities was useful in normalizing each volume in time series and removing brightness changes between slices within image volumes, for example slice-by-slice variations in interleaved acquisitions.

The results demonstrate the universality of the approach. PABIC cannot only be applied to MRI images but also to any kind of image data satisfying the initial assumptions. The algorithm was also tested successfully on different kinds of microscopy images and biological scenes measured by a video camera.

A major disadvantage of PABIC might be its computation time. A correction of a volume of dimensions 256x256x128, using two classes and Legendre polynomials up to the second degree, runs about 2 hours on a SUN Sparc-Ultra 10. The computation time decreases linearly when subsampling is used. Several tests on the accuracy of subsampling were performed and the results differed only marginally. In the 3D tests presented in this paper subsampling of 2x2x2 was used without affecting the results of the bias correction. This subsampling sped up the correction by a factor of 8, resulting in a computation time of about 15 minutes.

The present implementation runs semi-automated. Prior to the correction, class parameters have to be defined and two coarse masks need to be constructed (Section III-A). The degree of automation can be increased by registration of the image to an atlas prior to these preprocessing steps [34]. The parameters of the noise reduction, the class parameters, and the masks could then be derived from the result of the registration. This requires, however, a statistical atlas for every part of the body to be segmented.

Acknowledgments

The Multiple Sclerosis time series have been provided in connection with the BIOMORPH consortium (EU BIOMED 2). The MRA surface-coil head data have been provided by Prof. F. Jolesz, Dr. R. Kikinis and Dr. S. Wells, Brigham and Women's Hospital, Boston, USA. Prof. M. Brady, Oxford, is acknowledged for providing the MRI mammography image data. The simulated brain MRI images were provided by the Montral Neurological Institute (MNI) at the McGill University (http://www.bic.mni.mcgill.ca/brainweb).

We thank Sandy Wells for the many helpful discussions on the EM algorithm. We would also like to thank the reviewers for their insightful and constructive comments.

V. Appendix

A. Legendre Polynomials

The basis functions for the bias estimation must be smooth. First experiments confirm our presumption that the choice of an orthogonal basis will result in a better-posed condition for changing the parameters p_i. Legendre polynomials are just one choice that satisfies these requirements. A visualization of Legendre polynomials in 2D is displayed in Figure 17. Polynomials in each coordinate x_j are multiplied to define multivariate basis functions,

$P_{\underline{l}}(\underline{x}) = \prod_{j=0}^{n-1} P_{l_j}(x_j)$.

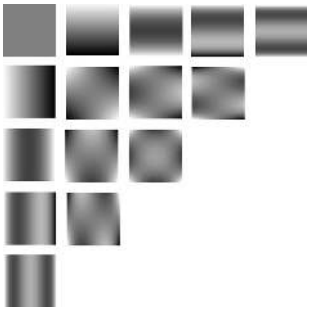

Fig. 17. Visualization of Legendre polynomials in 2D (0^{th} to 4^{th} degree in x and y): increasing x-degree from left to right, y-degree from top to bottom.

The Legendre polynomials are orthogonal, but not normalized to 1.

$$\int_{-1}^{1} P_t^2(x), dx \quad = \quad \frac{2}{2t+1} \tag{17}$$

The orthogonality is passed on to all products $P_{\underline{l}}(\underline{x})$. Using the theorem of Parseval, this property can be used to derive a formula for the computation of the Mean Squared Difference of two fields composed of Legendre polynomials directly from its coefficients. For the 3D case, this results in following formula

$$\begin{aligned} MSD &= \int_{-1}^{1} (\hat{b}_1(\underline{x}) - \hat{b}_2(\underline{x}))^2 d\underline{x} \\ &= \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} \sum_{i=0}^{l}\sum_{j=0}^{(l-i)}\sum_{k=0}^{(l-i-j)} \\ &\quad \left((p_{ijk}^{(\hat{b}_1)} - p_{ijk}^{(\hat{b}_2)}) P_i(x)P_j(y)P_k(z)\right)^2 dxdydz \qquad (18) \\ &= \sum_{i=0}^{l}\sum_{j=0}^{(l-i)}\sum_{k=0}^{(l-i-j)} \frac{(p_{ijk}^{(\hat{b}_1)} - p_{ijk}^{(\hat{b}_2)})^2}{(2i+1)(2j+1)(2k+1)} \qquad (19) \end{aligned}$$

B. Optimization with $(1+1)$-ES algorithm

An Evolution Strategy (ES) for non-linear optimization has been implemented. The $(1+1)$-ES optimization (described in [27]) iterates by randomly selecting a new position in parameter space, which is controlled by a probability function centered at the current location. The original $(1+1)$-ES algorithm uses a common variance for the mutation of all parameters. The isotropic probability function is subject to grow or shrink by an update rule which considers local information. The algorithm can be summarized as follows (terminology from [27]):

$$\begin{aligned} \underline{r}_t &\sim N(\underline{0}, I) \\ \underline{x}_{t+1} &= \underline{x}_t + a_t \underline{r}_t \\ a_{t+1} &= \begin{cases} a_t \cdot c_{grow} & \text{if } f(\underline{x}_{t+1}) < f(\underline{x}_{opt}) \\ a_t \cdot c_{shrink} & \text{otherwise} \end{cases} \\ \underline{x}_{opt} &= \begin{cases} \underline{x}_{t+1} & \text{if } f(\underline{x}_{t+1}) < f(\underline{x}_{opt}) \\ \underline{x}_{opt} & \text{otherwise} \end{cases} \end{aligned} \tag{20}$$

The parameters c_{grow}, c_{shrink}, x_0 and a_0 are given as input and $\underline{r} \sim N(\underline{0}, I)$ denotes a multi-dimensional random vector of an isotropic normal distribution with zero mean and variance one.

As a consequence of using an isotropic probability function, no adjustments to unequally scaled parameters and local properties of the energy landscape are possible, as mutations occur in all directions in parameter space with the same probability. Hence, the $(1+1)$-ES algorithm has been improved by additionally using second order statistics for the update of the multi-dimensional Gaussian probability function. The simple scaling of the normal density function by a scalar a_t is replaced by a multiplication with matrix A, which relates to the covariance matrix as $\Sigma = A \cdot A'$, A' being the transposed matrix A. We use the following property

$$\begin{aligned} \underline{r}_t &\sim N(\underline{0}, I) \\ A\underline{r}_t + \underline{m} &\sim N(\underline{m}, A \cdot A') \end{aligned} \tag{21}$$

Therefore, a scaled and translated vector $\underline{r}$ is an instance of a new normal distribution with mean m and covariance matrix $\Sigma = A \cdot A'$. At each stage of the iterative search, the covariance matrix is updated with new local information.

$$\begin{aligned} \underline{r}_t &\sim N(\underline{0}, I) \\ \underline{x}_{t+1} &= \underline{x}_t + A_t \cdot \underline{r}_t \\ A_{t+1} &= \begin{cases} A_t(I + (c_g - 1)\frac{r_t r_t'}{r_t' r_t})) & \text{if } f(\underline{x}_{t+1}) < f(\underline{x}_{opt}) \\ A_t(I + (c_s - 1)\frac{r_t r_t'}{r_t' r_t})) & \text{otherwise} \end{cases} \\ \underline{x}_{opt} &= \begin{cases} \underline{x}_{t+1} & \text{if } f(\underline{x}_{t+1}) < f(\underline{x}_{opt}) \\ \underline{x}_{opt} & \text{otherwise} \end{cases} \end{aligned} \tag{22}$$

Figure 18 illustrates the optimization of the Rosenbrock function $f(x,y) = 100(x^2 - y)^2 + (1-x)^2$ as an example. The Rosenbrock function is a standard test function for optimization algorithms. It is a badly conditioned problem, with a narrow winding valley from a start point to the global minimum. Figure 18b nicely illustrates the change of the second-order statistics in our method. Our optimization scheme is more flexible than the originally proposed $(1+1)$-ES and results in an improved convergence rate.

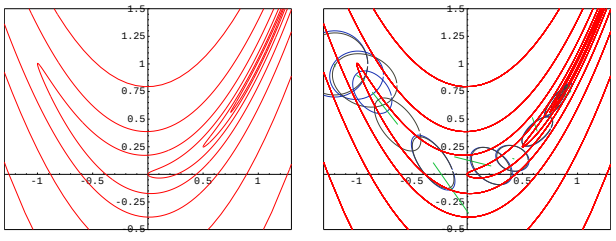

Fig. 18. Convergence of our $(1+1)$-ES algorithm on the Rosenbrock function. Left: Isolines of the Rosenbrock function. Right: Illustration of the iterative search procedure, starting upper left. The shapes of the ellipses represent isolevels of the probability distributions for finding a new child.

C. Model validation by MRI phantom images

The model's appropriateness to real MRI data was validated using a brain phantom with known substance classes and its MRI image. One echo from a dual echo MRI image of the brain phantom and the segmentation result are shown in Figure 19.

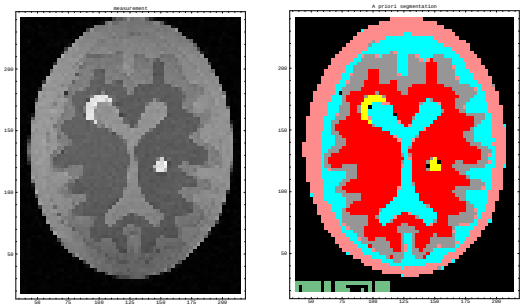

Fig. 19. The input data for the experiment is one MRI echo from a head phantom (*left*, courtesy of Dr. Ron Kikinis). The digitized phantom scene (*right*) is used as ground truth.

The image has been segmented using a statistical classification method applied to both echoes, and this segmentation (Figure 19) is considered to be ground truth $k(\underline{x})$ for the present experiment. We select 2D Legendre polynomials up to the third degree. Defining a characteristic function

$$\chi_{k'}(\underline{x}) = \begin{cases} 1, & \text{if } k(\underline{x}) = k' \\ 0, & \text{otherwise} \end{cases} ,$$

the model (2) can be made linear in the parameters μ and $\underline{p}$

$$\tilde{s}(\underline{x}) \quad = \quad \sum_k \mu_k \chi_k(\underline{x}) \cdot \sum_{i,j} p_{ij} P_i(x) P_j(y) + n(\underline{x}) \,. \quad (23)$$

Fitting a linear model is mathematically simple and yields the function fit shown in Figure 20**a** and **b**.

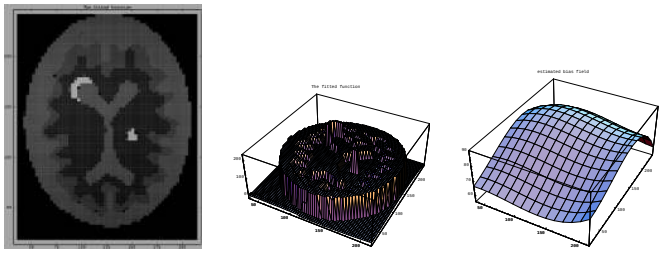

Fig. 20. Result of fitting the piecewise smooth model to the MR phantom. On the left is the resulting 2D image and in the middle a corresponding 3D plot. On the right, a scaled (factor 5) 3D plot of the estimated bias function is shown.

After subtracting the estimate for the idealized signal o and the bias estimate $\hat{b}$ from the signal $\tilde{s}$, we are left with the residual $\tilde{s} - o - \hat{b} = n$ (2) as illustrated in Figure 21. It is obvious that the residual is largest near the boundaries of tissues, where large changes in the image intensity can be observed. The flat regions of the residual indicate that the parametric bias estimate could sufficiently compensate for the inhomogeneity.

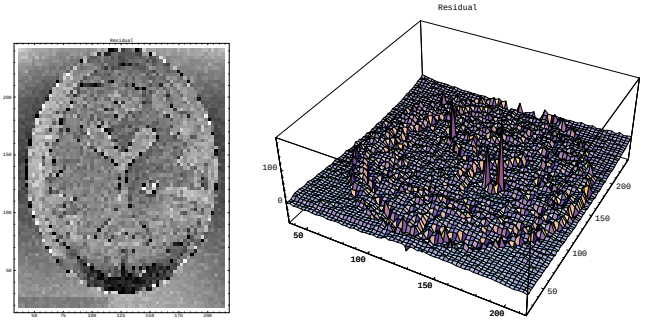

Fig. 21. Residual difference between the signal and the fitted model, shown as a gray level image and as 3D mesh.

References

[1] D.A. Ortendahl and J.W. Carlson, "Segmentation of magnetic resonance images using fuzzy clustering," in *Proc. of Information Processing in Medical Imaging, IPMI'87*, 1988, vol. 39, pp. 91–106.

[2] M.W. Vannier, Speidel Ch.M., D.L. Rickman, L.D. Schertz, et al., "Validation of Magnetic Resonance Imaging (MRI) Multispectral Tissue Classification," in *Proc. of 9th Int. Conf. on Pattern Recognition, ICPR'88*, Nov. 1988, pp. 1182–1186.

[3] M.B. Merickel et al., "Multispectral pattern recognition of MR imagery for the noninvasive analysis of atherosclerosis," in *Proc. of 9th Int. Conf. on Pattern Recognition, ICPR'88*, Nov. 1988, pp. 1192–1197.

[4] K.O. Lim and A.J. Pfefferbaum, "Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter," *Journal of Computer Assisted Tomography*, vol. 13, pp. 588–593, 1989.

[5] H.E. Cline, W.E. Lorensen, St.P. Souza, F.A. Jolesz, R. Kikinis, G. Gerig, and Th.E. Kennedy, "3D surface rendered MR images of the brain and its vasculature," *Journal of Computer Assisted Tomography*, vol. 15, no. 2, pp. 344–351, Mar. 1991.

[6] L.P. Clarke, R.P. Velthuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, and M.L. Silbiger, "MRI segmentation: methods and applications," *Magnetic Resonance Imaging*, vol. 13, pp. 343–368, 1995.

[7] E.T. Bullmore, M.J. Brammer, G. Rouleau, B.S. Everitt, A. Simmons, T. Sharma, R. Frangou, R.M. Murray, and G. Dunn, "Computerised brain tissue classification of magnetic resonance images: A new approach to the problem of partial volume artefact," *Neuroimage*, vol. 2, pp. 133–147, 1995.

[8] A. Simmons, S.R. Arridge, G.K. Barker, and S.C.R. Williams, "Simulation of MRI cluster plots and application to neurological segmentation," *Magnetic Resonance Imaging*, vol. 14, pp. 73–92, 1996.

[9] M.I. Kohn, N.K. Tanna, G.T. Herman, S.M. Resnick, P.D. Mozley, R.E. Gur, A. Alavi, R.A. Zimmerman, and R.C. Gur, "Analysis of brain and cerebrospinal fluid volumes with MR imaging. Part I. Methods, reliability, and validation," *Radiology*, vol. 178, pp. 115–122, Jan. 1991.

[10] M. Jungke, W. von Seelen, G. Bielke, S. Meindl, et al., "A system for the diagnostic use of tissue characterizing parameters in NMR-tomography," in *Proc. of Information Processing in Medical Imaging, IPMI'87*, 1987, vol. 39, pp. 471–481.

[11] M.W. Vannier, Ch.M. Speidel, and D.L. Rickman, "Magnetic resonance imaging multispectral tissue classification," *NIPS*, vol. 3, pp. 148–154, August 1988.

[12] J. Haselgrove and M. Prammer, "An algorithm for compensation of surface-coil images for sensitivity of the surface coil," *Magnetic Resonance Imaging*, vol. 4, pp. 469–472, 1986.

[13] B.M. Dawant, A.P. Zjidenbos, and R.A. Margolin, "Correction of intensity variations in MR images for compuer-aided tissue classification," *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp. 770–781, 1993.

[14] M. Tincher, C.R. Meyer, R. Gupta, and D.M. Williams, "Polynomial modeling and reduction of RF body coil spatial inhomogeneity in MRI," *IEEE Transactions on Medical Imaging*, vol. 12, no. 2, pp. 361–365, 1993.

[15] Charles R. Meyer, Peyton H. B land, and James Pipe, "Retro-

spective correction of Intensity Inhomogeneities in MRI," *IEEE Transactions on Medical Imaging*, vol. 14, no. 1, pp. 36–41, Mar. 1995.

[16] W.M. Wells, W.E.L. Grimson, R. Kikinis, and F.A. Jolesz, "Statistical gain correction and segmentation of MRI data," 1993, internal report BWH Boston.

[17] William Wells, Ron Kikinis, and Ferenc A. Jolesz, "Statistical intensity correction and segmentation of magnetic resonance image data," in *Proceedings of the Third Conference on Visualization in Biomedical Computing VBC'94*, Oct. 1994, vol. 2359, pp. 13–24.

[18] W.M. Wells, W.E.L. Grimson, R. Kikinis, and F.A. Jolesz, "Adaptive Segmentation of MRI Data," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 429–443, Aug. 1996.

[19] Régis Guillemaud and Michael Brady, "Estimating the Bias Field of MR Images," *IEEE Transactions on Medical Imaging*, vol. 16, no. 3, pp. 238–251, June 1997.

[20] John G. Sled, P. Zijdenbos, and Alan C. Evans, "A Comparison of Retrospective Intensity Non-uniformity Correction Methods for MRI," in *Information Processing in Medical Imaging*, 1997, vol. 1230, pp. 459–464, Proceedings 15th Int. Conf. IMPI'97.

[21] John G. Sled, P. Zijdenbos, and Alan C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in mri data," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 87–97, Feb. 1998.

[22] P. Huber, *Robust Statistics*, Wiley, 1981.

[23] S.Z. Li, "Robustizing m-estimation using deterministic annealing," *Pattern Recognition*, vol. 29, no. 1, pp. 159–166, 1996.

[24] Colin R. Reeves, *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell Scientific Publications, 1993.

[25] Ch. Brechbühler, G. Gerig, and G. Székely, "Compensation of spatial inhomogeneity in MRI based on a multi-valued image model and a parametric bias estimate," in *Visualization in Biomedical Computing (VBC) '96*, Sept. 1996, pp. 141 – 146.

[26] Hans-Paul Schwefel, *Evolution and optimum seeking*, Wiley, cop., 1995.

[27] Willfried Wienholt, *Entwurf neuronaler Netze: ein datenbasiertes Verfahren zur Modellierung nichtlinearer Systeme mit Hilfe evolutionärer Algorithmen und unscharfer Logik*, Harri Deutsch, 1996.

[28] Martin Styner and Guido Gerig, "Evaluation of 2d/3d bias correction with 1+1es-optimization," Tech. Rep. 179, Image Science Lab, ETH Zürich, 1997.

[29] C.A. Cocosco, V. Kollokian, R.K.-S. Kwan, and A.C. Evans, "Brainweb: Online interface to a 3d mri simulated brain database," *NeuroImage*, vol. 5, no. 4, pp. 425, 1997.

[30] R.K.-S. Kwan, A.C. Evans, and G.B. Pike, "An extensible mri simulator for post-processing evaluation," in *Visualization in Biomedical Computing, Proc. VBC '96*, Sept. 1996, pp. 135–140.

[31] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 463–468, June 1998.

[32] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, Apr. 1997.

[33] G. Gerig, D. Welti, C. Guttmann, A. Colchester, and G. Székely, "Exploring the discriminating power of the time domain for segmentation and characterization of lesions in serial mr data," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI 98)*, 1998, pp. 469–480.

[34] V.L. Koen, F. Maes, D. Vandermeulen, and P. Suetens, "Automatic Segmentation of Brain Tissues and MR Bias Correction Using a Digital Brain Atlas," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI 98)*, 1998, pp. 1222–1229.