

Entropy minimization for automatic correction of intensity nonuniformity

Jean-François Mangin
Service Hospitalier Frédéric Joliot, CEA
4 place du Gal Leclerc, 91401 Orsay, France
E-mail: mangin@shfj.cea.fr

Abstract

This paper outlines a fully automatic method for the correction of intensity nonuniformity in MR images. This method does not require any a priori model of the tissue classes. The basic idea is that entropy is a good measure of the image quality which can be minimized in order to overcome the bias problem. Therefore, the optimal correcting field is defined by the minimum of a functional combining the restored image entropy and a measure of the field smoothness. This measure stems from the usual physical analogy with membranes. A third term added to the functional prevents the optimal field from being uniformly null. The functional is minimized using a fast annealing schedule. The performance of the method is evaluated using both real and simulated data.

1. Introduction

While an impressive amount of effort has been dedicated to the classification of tissues in various magnetic resonance (MR) based images, the difficulties induced by the intensity spatial inhomogeneity usually observed in the MR signal is only a recent subject of discussion. However, the level of intensity variations often reaches 20% in standard images and even more for the high field scanners used by the brain mapping community. Although these variations have little impact on visual diagnosis, they may result in a lot of difficulties for algorithms assuming intra-class homogeneity or endowing isosurfaces with some anatomical meaning.

Sources of intensity inhomogeneity in MR images are numerous [11]. A lot of them can be overcome by regular calibration. However, the difficulties due to induced currents and the spatial inhomogeneity of the excitation field depend on the geometry and electrical properties of the subject [12]. While some solution could stem from the direct *in vivo* measure of the excitation field using spe-

cific MR sequences, such an approach seems rather cumbersome for routine acquisitions. As such, a purely post-processing oriented approach for intensity inhomogeneity correction is required; a subject of increasing interest during the last years in the image analysis community [3, 8, 13, 14, 5, 4, 2, 12, 10, 1].

Most of the methods proposed to date require expert supervision which in turn has prevented their widespread use. The expert has to specify some parameters related to the intensity models of the different tissue classes. These parameters are usually inferred from regions of interest. Among these methods, the more attractive one, which was proposed by Wells et al., consists in using an expectation maximization (EM) strategy to compute iteratively the optimal smooth bias field corrupting the data from a classification in several tissue classes [14]. Usually, the intensity model assumed for each tissue class is a small variance Gaussian distribution. Hence, the EM approach raises the problem of selecting the number of classes which have to be explicitly modeled. Interesting solutions stemming from information theory could allow this difficult choice to be done automatically [6, 5, 9]. Another troublesome point, however, is the model of the *other* class that includes the tissues not modeled by simple Gaussians. The choice of the uniform probability density is advocated in [5] as yielding significantly better results than a large variance Gaussian distribution. Finally the EM algorithm is unfortunately bound to converge to a local minimum for some bias configurations, especially when more than two tissue classes are modeled [5]. An alternative to this probability-based approach consists in using the fuzzy C-means algorithm to perform the classification [10, 1]. This approach leads to a robust unsupervised method but requires a priori knowledge of the class number.

A different approach recently proposed by Sled et al. [12] overcomes most of the difficulties embedded in the EM-like iterative methods that include a classification step. This approach does not require any explicit model of the different classes present in the MR image. Moreover, this nonparametric nonuniform intensity normalization method

(N3) is insensitive to pathological data that might otherwise violate model assumptions. This approach consists in searching for the smooth multiplicative field that maximizes the frequency content of the image intensity distribution. This signal processing oriented approach takes advantage of the usually simple form of the distribution of the logarithm of the bias field. Hence, the method iteratively deconvolves narrow Gaussian distributions from the image intensity distribution. Various experiments show the robust behaviour of this iterative method which does not seem prone to local minima.

The rapid overview of the literature presented above leads to the observation that a bias correction method mainly requires an image quality criterion. Then an optimal correcting field incorporating smoothness properties can be computed. In this paper, we propose to define the quality criterion from an information theory perspective. This idea was proposed first by P. Viola in one prospective chapter of his PhD thesis [13] where preliminary 2D results can be found. Minimum entropy has often been proposed as a good criterion for the design of restoration or classification methods [13, 5, 6, 9]. Indeed, assuming a thin unimodal intensity distribution for each tissue class (ideally a spike), the entropy of a perfect MR image should be very low. The bias field results in a spreading of each class distribution leading to a more uniform image intensity distribution endowed with higher entropy. Therefore, we propose to define a functional U of the correcting field \mathcal{F}_c combining the restored image entropy and a measure of the field smoothness. Hence, the optimal correcting field \mathcal{F}_c^{opt} , which minimizes $U(\mathcal{F}_c)$, provides a trade-off between field smoothness and image quality. In fact, a third term has to be added in $U(\mathcal{F}_c)$ to prevent the optimal field from being uniformly zero, which will be explained further. The functional is minimized by a fast annealing process overcoming the difficulties induced by potential local minima. The efficiency of the method is evaluated using simulated and real data.

2. Method

2.1. General strategy

A convenient way of designing a bias correction method consists of modeling the intensity nonuniformity as resulting from a smooth multiplicative field. While this multiplicative model is consistent with some characteristics of the underlying acquisition principle, it should be noted that other nonuniformity sources can not be fully account for by this model. For instance, since intensity non uniformity due to excitation inhomogeneity is related to the tissue longitudinal relaxation time, this artefact is bound to present discontinuities along tissue boundaries. Nevertheless, if we admit that the correction method is intended to overcome

segmentation problems rather than to infer the exact bias field, the smooth multiplicative field is largely sufficient for our purpose.

Let us consider the usual following model of image formation [5, 12]:

$$\mathcal{O}(x) = \mathcal{I}(x)\mathcal{F}(x) + \mathcal{N}(x) , \quad (1)$$

where at location x , $\mathcal{O}(x)$ is the observed signal, $\mathcal{I}(x)$ is the signal emitted by the tissue including partial volume effect, $\mathcal{F}(x)$ is an unknown smoothly varying bias field and $\mathcal{N}(x)$ is white noise. Assuming that without bias, each tissue should lead to a clustered set of values, we propose to use entropy as a measure of the quality of the restored image $\mathcal{F}_c\mathcal{O}$, where \mathcal{F}_c is a smooth multiplicative field. Indeed, corruption from the bias blurs together the clusters which increases entropy of the gray level distribution. The optimal correcting field \mathcal{F}_c^{opt} is then defined as the global minimum of the functional:

$$U(\mathcal{F}_c) = K_S S(\mathcal{F}_c\mathcal{O}) + K_R R(\mathcal{F}_c) + K_M M(\mathcal{F}_c\mathcal{O}) , \quad (2)$$

where $S(\mathcal{F}_c\mathcal{O})$ is the entropy of the restored image, $R(\mathcal{F}_c)$ is a regularizing function which measures the field smoothness, and $M(\mathcal{F}_c\mathcal{O})$ is a quadratic measure of the discrepancy between the observed image mean and the restored image mean. K_S , K_R and K_m are positive weighting constants.

For the sake of efficiency during U 's minimization, the first term $S(\mathcal{F}_c\mathcal{O})$ is simply computed from the histogram of the restored image $\mathcal{H}_{\mathcal{F}_c\mathcal{O}}$. During the computation of this histogram, intensity at location x , given by $\mathcal{F}_c(x)\mathcal{O}(x)$, is rounded off to the nearest integer. Then:

$$S(\mathcal{F}_c\mathcal{O}) = - \sum_i \frac{\mathcal{H}_{\mathcal{F}_c\mathcal{O}}(i)}{n_V} \log\left(\frac{\mathcal{H}_{\mathcal{F}_c\mathcal{O}}(i)}{n_V}\right) , \quad (3)$$

where n_V denotes the total number of locations (the number of voxels) and $\mathcal{H}_{\mathcal{F}_c\mathcal{O}}(i)$ denotes the number of locations with intensity i .

Various parameterized smooth field models can be designed. Among them, spline based approaches are the more attractive ones because they lead to local dependencies on the field parameters. Hence, updating spline's control points during the optimization can be done at a low cost. During the minimization, again for the sake of efficiency, the correcting field is simply modelled by a piecewise constant image (a zero order spline), which turns out to be sufficient for our purpose. For each location (X, Y, Z) , the field value is given by $\mathcal{F}_c(x) = f_c(X/s_{\mathcal{F}}, Y/s_{\mathcal{F}}, Z/s_{\mathcal{F}})$, where $s_{\mathcal{F}}$ denotes the sampling of the field and f_c a little volume made up by the field parameters. Hence each parameter is the correcting value for a volume of $s_{\mathcal{F}}^3$ voxels of the raw image. Nevertheless, the final optimal field \mathcal{F}_c^{opt} is resampled from f_c^{opt} using cubic splines before performing the

final restoration, thus avoiding potential discontinuity artefacts. The logarithm of the piecewise constant field f_c is regularized using the usual membrane analogy. The regularizing energy is then minimal for constant fields and is invariant relative to a multiplicative factor:

$$\begin{aligned} R(\mathcal{F}_c) &= \sum_{x_f} \sum_{x'_f \in N_6(x_f)} (\ln(f_c(x_f)) - \ln(f_c(x'_f)))^2 \\ &= \sum_{x_f} \sum_{x'_f \in N_6(x_f)} \ln^2\left(\frac{f_c(x_f)}{f_c(x'_f)}\right), \end{aligned} \quad (4)$$

where $N_6(x_f)$ denotes the 6-neighbourhood of the location x_f in the field f_c .

The third term $M(\mathcal{F}_c\mathcal{O})$ is used to prevent the optimal field from being uniformly zero. Let us study the behaviour of the functional U relative to constant fields \mathcal{F}_c for which the regularizing energy is minimal. Since the computation of the restored image from the field leads only to integer intensity values (in our implementation), applying a constant field lower than 1 leads to a ‘‘compression with loss’’ of the information embedded in the intensity distribution. Since entropy can be related to the minimum number of bits required to code the intensity distribution, this value is globally decreasing with the constant field level, even if this relationship is complex because of discrete phenomena. Therefore, without the third term of the functional U , the global minimum would turn out to be the null field, which is far from being the anticipated result. Instead information reduction should follow from the degrees of freedom of the field allowing a sharpening of the histogram modes rather than from a global intensity compression. A full untangling of both phenomena seems problematic, and, anyway unnecessary for our purpose. Therefore, a simple solution consists of accommodating the functional with an additional term to prevent the compression from being too high:

$$M(\mathcal{F}_c\mathcal{O}) = (\text{mean}(\mathcal{F}_c\mathcal{O}) - \text{mean}(\mathcal{O}))^2. \quad (5)$$

Another interesting solution would consist of imposing a low threshold on the possible correcting field values [5].

2.2. Implementation details

In order to overcome the potential difficulties induced by the non-convexity of the functional U , the minimization is performed using a stochastic algorithm. This algorithm relies on a fast annealing schedule implemented using a classical update scheme. The temperature decreasing schedule follows a geometric law (multiplicative factor: 0.95). Each annealing iteration is a loop on f_c points with a constant temperature T . For each point x_f , a multiplicative increment i is chosen randomly in the range [0.5,2]. Then the functional variation ΔU related to the transition from

$f_c(x_f)$ to $i * f_c(x_f)$ is computed. The transition is accepted if $\frac{1}{1+e^{\frac{\Delta U}{T}}}$ is greater than a value randomly chosen in [0,1] according to the uniform law. The field is initialized with random values taken in the range [0.5,2].

The whole minimization process, which converges in less than 200 iterations, takes a few minutes on a conventional workstation. Experiments applying several minimizations with different field initializations led to very similar results. Moreover, other experiments using an ICM-like deterministic algorithm also led to interesting solutions, even if annealing improves the result. This behaviour suggests that the landscape of the functional in the domain where the field is travelling is a large valley with some shallow local minima. Provided that this assumption is a good metaphor of the multidimensional nature of this landscape, the robustness of the fast annealing implementation can be understood.

The balancing of the three terms of the functional U is a difficult question which will require further work in the future. For the experiments described in this paper, the weights and the field sampling have been fixed as follows: $K_S = n_f$, $K_R = 10$, $K_M = 0.005 * n_f$, $s_{\mathcal{F}} = 10$, where n_f denotes the number of parameters of the field (the number of voxels in f_c). Various experiments have shown that slight modifications of these values have a very small impact on the final result. The low value chosen for K_M aims at minimizing as far as possible the influence of $M(\mathcal{F}_c\mathcal{O})$ on the shape of U landscape. Indeed, several experiments have shown that a high value of K_M could lead to minimization difficulties requiring a slower temperature decreasing schedule. Intuitively, this behaviour stems from an important reduction of the number of paths in the field space endowed everywhere with a relatively low value of the functional U . Hence, the depth of local minima is increased which reduces the efficiency of the annealing process.

The range of intensities provided by different scanners or different MR sequences presents great variations. According to the acquisition, intensity can be coded in 7 or 16 bits. In order to get a robust estimation of the intensity distribution from the volume histogram, we apply a preprocessing step to the data which consists of compressing the intensities by the largest multiplicative factor $\frac{1}{2^n}$ putting 95% of the histogram within the range [0,64]. This compression, which is only a scaling, amounts to neglecting some of the bits. Thanks to this compression, the entropy and mean variations observed during U 's minimization are comparable between images, which explains why the same set of weighting constants can be used in all cases.

Owing to the nature of the model of image formation proposed in Eq. 1, it seems rather impossible to retrieve the bias field in the image areas where the observed signal is at the noise level. Such areas are, for instance, the out-

side of the head, the skull or the cerebro-spinal fluid (CSF) for inversion-recovery sequences. At first glance, this impossibility is not a problem because we are not really interested in the field in these areas. However, because of the smoothness property imposed on the field, this situation gives rise to difficulties during its estimation. These difficulties could be overcome by the use of a more complex regularizing term including, for instance, bounded penalty potentials (Φ -functions) or line processes. In fact a simpler solution consists of restricting the field effect on the range of intensities located above a fixed threshold t_{low} in the observed image. Hence intensities from low signal areas have no influence on the field parameters. For instance, outside the head, the optimal field values stem exclusively from the regularization term. The threshold t_{low} could be fitted for each image, but with the intensity compression mentioned above, fixing the threshold to 10 turned out to be adequate for all the images we tested. Then, during the minimization process, the field value at each location X, Y, Z is given by the rule: if $\mathcal{O}(X, Y, Z) < t_{low}$, $\mathcal{F}_c(x) = 1$; otherwise $\mathcal{F}_c(x) = f_c(X/s_{\mathcal{F}}, Y/s_{\mathcal{F}}, Z/s_{\mathcal{F}})$. In return, the simple rule without threshold is used during the final restoration to prevent artefacts.

3. Results and Discussion

We first present results with a high quality T1-weighted image of the head. Acquisition was performed on 1.5 T system (Signa, GE, 256x256x124 voxels, 0.86x0.86x1.2mm resolution) using an inversion-recovery sequence which yields a high contrast between gray and white matter, and no signal in CSF. Thanks to this high contrast between tissue classes, most of the usual image processing techniques give good results even though they are corrupted by a bias field. Nevertheless, our experiments show that a bias correction provides some improvements. Fig. 1.A displays two orthogonal slices of this volume with a rainbow colormap (unfortunately converted in grey values because of publication constraints). The volume histogram illuminates the high contrast between gray and white matter. Color variations inside white matter reveal a bias. It should be noted that a part of these variations can be related to normal intra-tissue variations (highlighting of the pyramidal tract in the coronal slice) or partial volume effect (thinnest parts of the gyri). The result of a standard brain segmentation using mathematical morphology [7] is corrupted by the bias: some thin gyri of the parietal lobe are truncated while some sulci of temporal and occipital lobes are closed. These problems are overcome in the equivalent segmentation stemming from the restored image (see Fig. 1.B). Moreover, the color variations of white matter in the restored image suggest a more uniform intensity distribution. The histogram of the restored image shows that the spreading of the white matter

distribution has been largely reduced. In order to evaluate this effect we have performed a simple threshold based classification of the restored brain in gray and white matter. The accuracy of this classification is of course debatable, and is given here only for demonstrative purposes. Similar to previous works [5, 12], we compared the coefficient of variations in these two volumes of interest before and after correction. We observed an important uniformity improvement for white matter from 10.1% to 6.9%, while the improvement is not significant for gray matter (see Fig. 1.C). In fact, the largest part of grey matter intensity variation seems related to partial volume effect which masks a potential nonuniformity correction. This bias correction experiment has been reproduced successfully with 40 3D images obtained from three different 1.5 T scanners using various T1-weighted sequences.

The second experiment consists of using the corrected image yielded by the previous experiment in order to simulate large magnitude bias fields. This idea, which appears as an alternative to the use of purely simulated volumes [12], may be discussed. However, it allows us to use real images to study the correction method's behaviour. Fig.2 presents the results obtained for a simulated field corresponding to $(1 + 6 * e^{-\frac{\|x-C_1\|^2}{60^2}})(1 + 4 * e^{-\frac{\|x-C_2\|^2}{70^2}})$, where $C_1 = (50, 50, 50)$ and $C_2 = (150, 170, 170)$. The simulated field histogram illustrates that this field magnitude is largely beyond the usual ones. While the histogram of the corrupted image does not present any clue to the two underlying tissue classes, various results presented in the figure argue that the correction performed well. For instance, the coefficient of variation in white matter decreases from 30.1% with the bias to 8.6% after correction (versus 6.9% in the initial volume). It should be noted that the field estimation is not possible outside the head, which explains the differences with the simulated field. Similar experiments with other field parameters led to equivalent results.

The third experiment has been performed with a T1-weighted volume obtained from a 3T Brucker (128x256x256 voxels, 1.5x1x1 mm resolution). The bias field observed for such high field acquisitions is usually endowed with a high magnitude which leads to a flat histogram. The different results presented in Fig.3 suggest once again the good behaviour of the correction methods. For instance, the two intensity modes related to gray and white matter are clearly visible in the corrected image histogram. Histogram analysis using a scale-space based method described elsewhere [7] provides a clear detection of the two modes. This bias correction experiment has been reproduced successfully with 10 3D images obtained from the 3T scanner.

The last experiment has been performed with a T1-weighted image of a baboon head acquired without a dedicated coil, leading to a high magnitude bias field. Our cor-

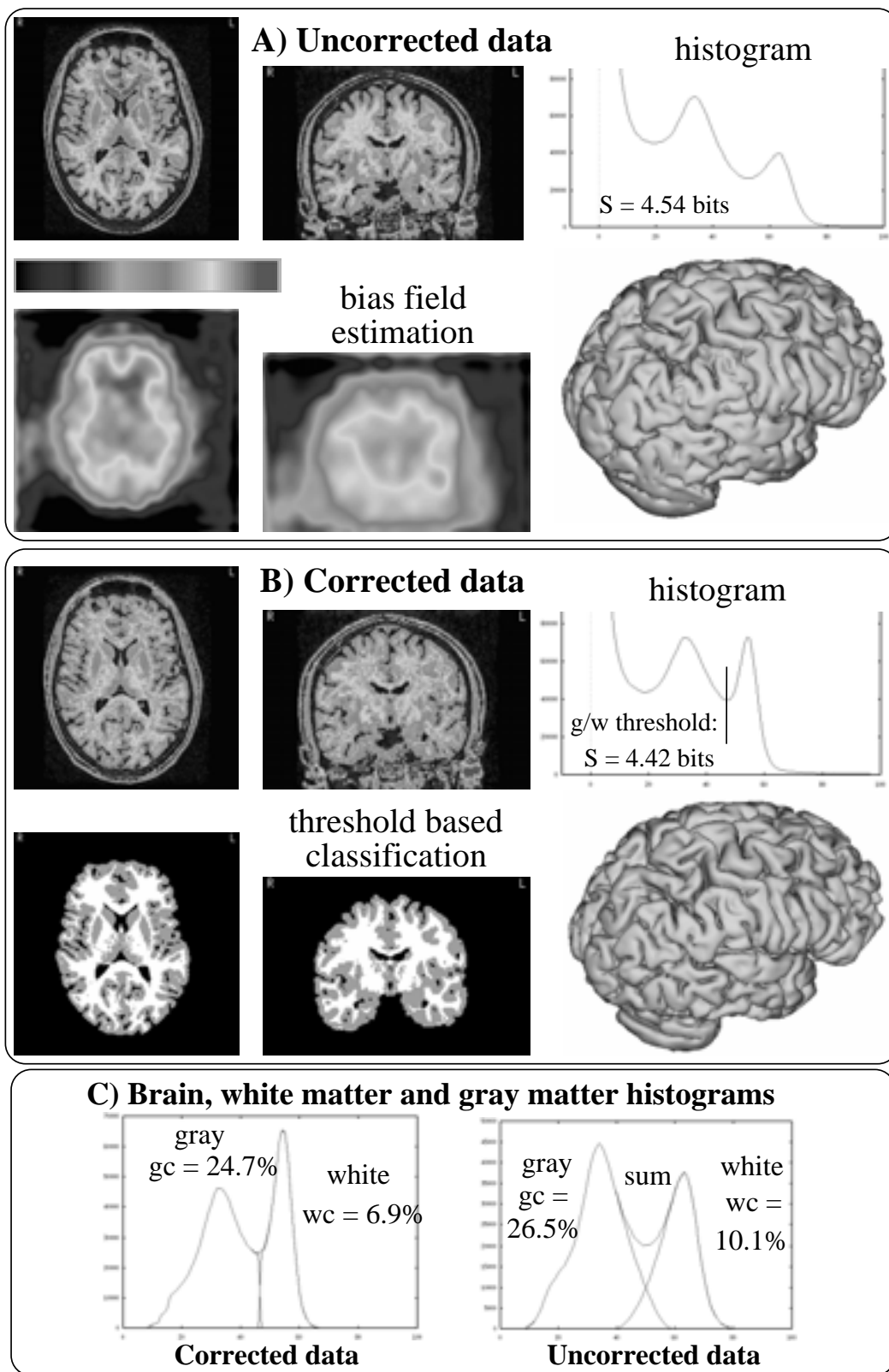


Figure 1. Bias correction in high contrast inversion-recovery images (gc and wc denote the coefficients of variation in gray and white matters)

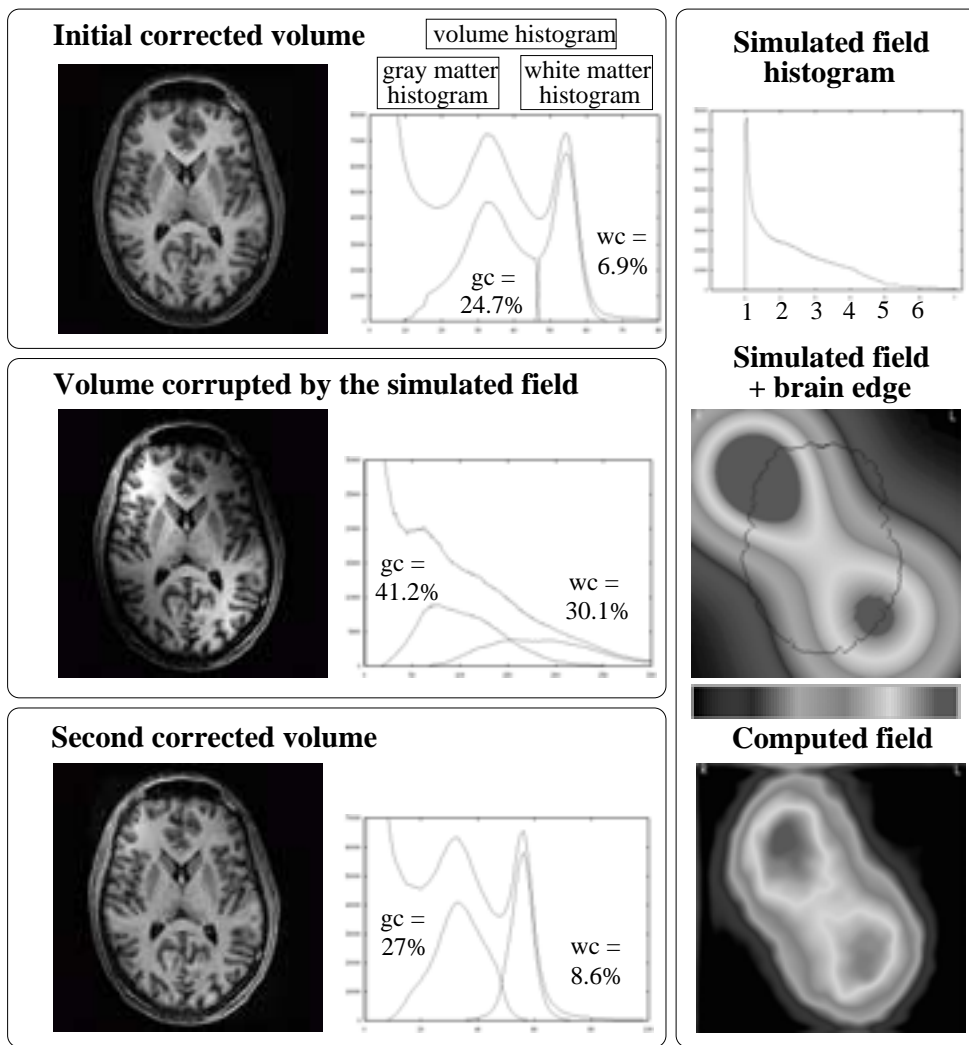


Figure 2. Correcting a high magnitude simulated field

rection method led to interesting results which turned out to be sufficient for performing automatic brain segmentation. Nevertheless, this experiment highlights some weaknesses of the entropy based criterion when one specific tissue class outmatches the others in size and is located in one specific image area (this problem occurs with monkeys because of their large muscles around a small brain relative to the large human brain surrounded by small muscles). In such cases, the final histogram includes only one tissue mode.

One important point to be discussed is the fact that the estimated bias fields contain some anatomical information. For instance, the field in Fig. 1.A looks like a blurred version of a gray/white classification. It should first be noted that this experiment was performed with an inversion-recovery MR sequence which gives no signal in CSF. Hence CSF has no influence on the bias estimation. Secondly, further experiments were performed with the same image

but with a set of increasing regularization weights. While the estimated field's low spatial frequencies were similar across the experiments, a more or less blurred version of the gray/white matter interface was visible in all these fields. In our opinion, this observation leads us to question the hypothesis of the field smoothness. This point is difficult to discuss exhaustively because of the large variety of T1-weighted MR sequences. Nevertheless, a lot of them lead to similar dependencies between the measured signal and the flip angle θ which is corrupted by inhomogeneities. For one of the simplest sequences, this dependency during the steady state is the following:

$$M = M_0 \sin \theta \frac{1 - \exp^{-\frac{T_R}{T_1}}}{1 - \cos \theta \exp^{-\frac{T_R}{T_1}}}, \quad (6)$$

where M denotes the measured signal, T_1 the tissue longitudinal relaxation time and T_R the radiofrequency pulse

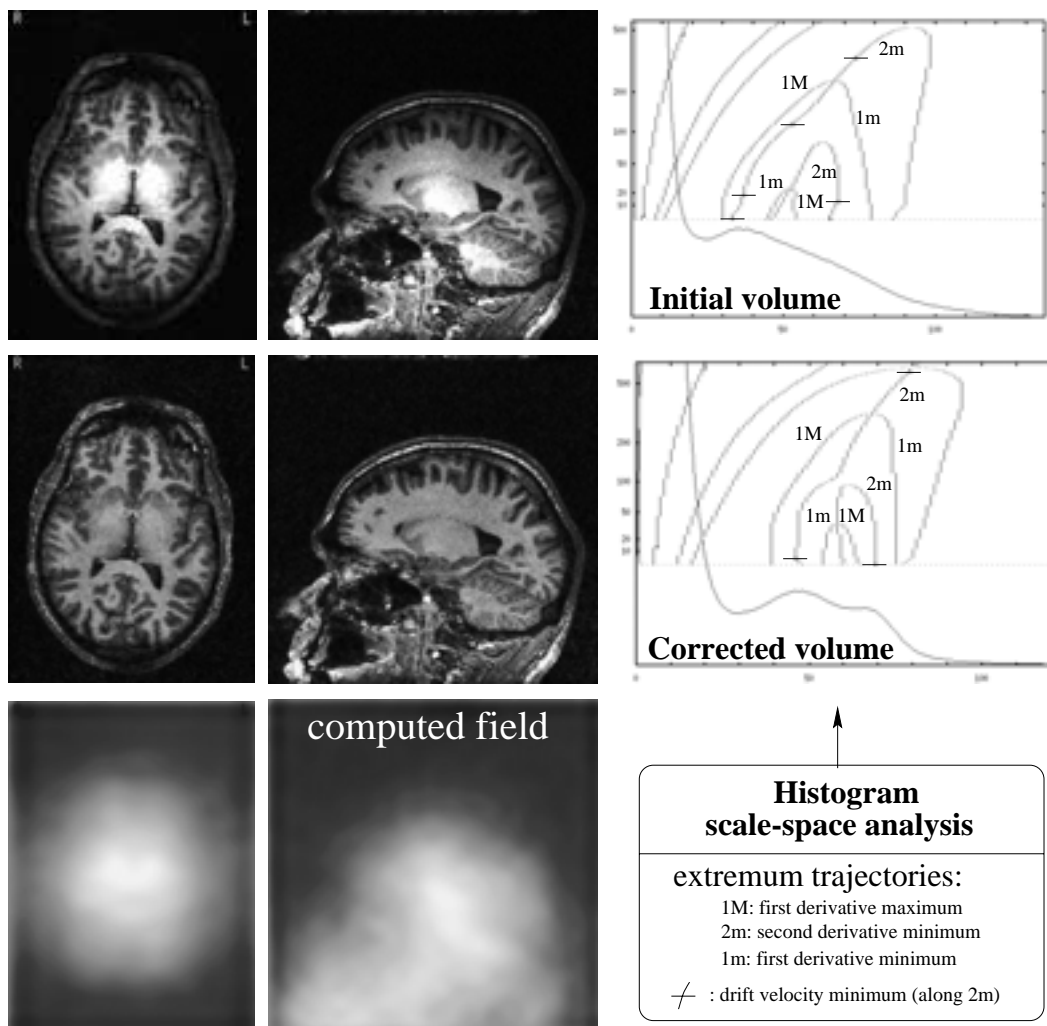


Figure 3. Correcting high magnetic field based images

repetition time. A typical set of values is: $T_R = 10ms$, $T_1(\text{white}) = 800ms$, $T_1(\text{gray}) = 1000ms$, $\theta = 10^\circ$. Inhomogeneities lead to a range of θ values between 7° and 15° . While θ spatial variations may be smooth, the discontinuity in T_1 corresponding to the gray/white matter interface will lead to a discontinuity of the correcting multiplicative field. Although this may call for a more sophisticated regularization model using for instance bounded penalty functions, this improvement is not required when the only purpose is tissue classification. Indeed we do not aim at recovering the exact bias field, but instead at a delineation of the tissue gray level distributions.

4. Conclusion

A fully automatic and fast bias correction method has been presented. The robustness of this method for T1-weighted brain imaging suggested by all the experiments

made in our institution have led us to use it systematically without supervision ahead of any further processing operations. The extension of this method to other applications is currently under study. The difficulties highlighted by the experiments with the baboon head, however, have shown that this entropy based approach may not be universal.

References

- [1] M. N. Ahmed, S. M. Yamany, N. A. Mohamed, and A. A. Farag. A modified fuzzy C-means algorithm for MRI bias field estimation and adaptive segmentation. In *MICCAI'99, Cambridge, UK*, pages 72–81, 1999.
- [2] C. Brechbühler, G. Gerig, and G. Szekely. Compensation of spatial inhomogeneity in MRI based on a parametric bias estimate. In *VBC'96, vol. 1131 of LNCS, Springer-Verlag*, pages 141–146, 1996.

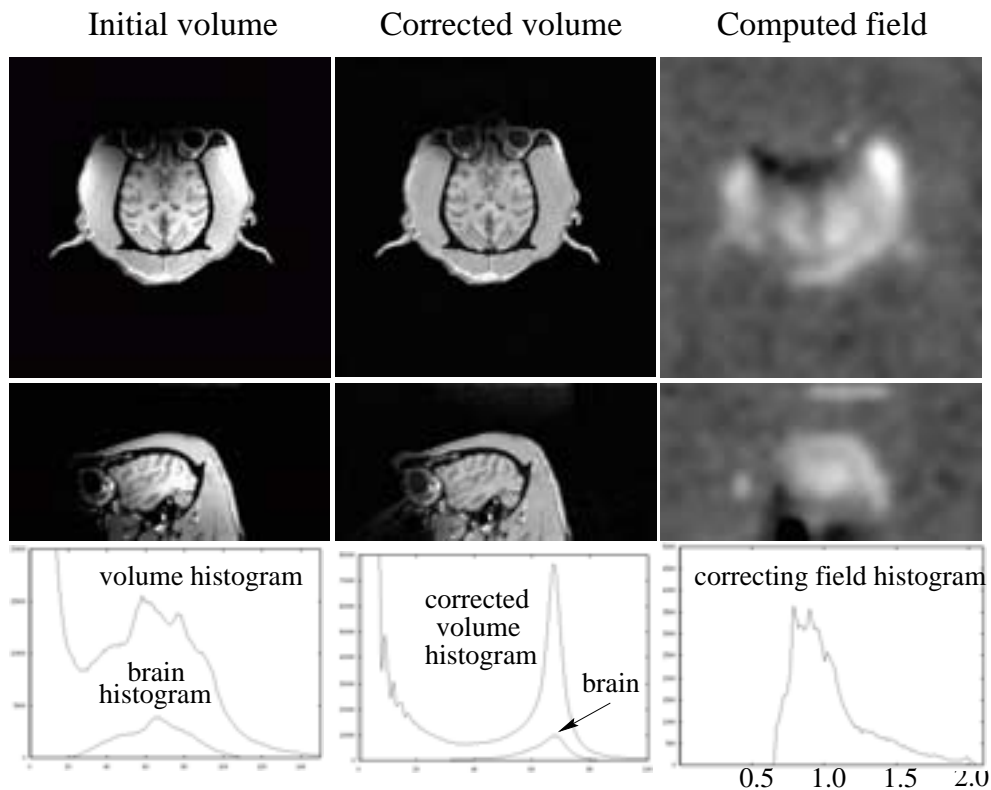


Figure 4. Correcting a baboon image acquired without dedicated coil

- [3] B. M. Dawant, A. P. Zijdenbos, and R. A. Margolin. Correction of intensity variation in MR images for computer-aided tissue classification. *IEEE Trans. on Medical Imaging*, 12:770–781, 1993.
- [4] S. Gilles, M. Brady, J.-P. Thirion, and N. Ayache. Bias field correction and segmentation of MR images. In *VBC'96, vol. 1131 of LNCS, Springer-Verlag*, pages 153–158, 1996.
- [5] R. Guillemaud and M. Brady. Estimating the bias field of MR images. *IEEE Trans. on Medical Imaging*, 16(3):238–251, 1997.
- [6] Y. Leclerc. Constructing simple stable descriptions for image partitioning. *Int. J. Comput. Vision*, 3, 1989.
- [7] J.-F. Mangin, O. Coulon, and V. Frouin. Robust brain segmentation using histogram scale-space analysis and mathematical morphology. In *MICCAI'98, MIT, LNCS 1496, Springer-Verlag*, pages 1230–1241, 1998.
- [8] C. R. Meyer, P. H. Bland, and J. Pipe. Retrospective correction of intensity inhomogeneities in MRI. *IEEE Trans. on Medical Imaging*, 14:36–41, 1995.
- [9] G. Palubinskas, X. Descombes, and F. Kruggel. An unsupervised clustering method using the entropy minimization. In *ICPR'98, Brisbane, Australia*, 1998.
- [10] D. L. Pham and J. L. Prince. An adaptive fuzzy C-means algorithm for image segmentation in the presence of image inhomogeneities. *Pattern Recognit. Lett.*, 20(1):57–68, 1999.
- [11] A. Simmons, P. S. Tofs, G. J. Barker, and S. R. Arridge. Sources of intensity nonuniformity in spin echo images at 1.5T. *Magn. Reson. Med.*, 32:121–128, 1994.
- [12] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A non-parametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. on Medical Imaging*, 17(1):87–97, 1998.
- [13] P. A. Viola. *Alignment by maximization of mutual information*. A.I.-1548, chap. 6, MIT, AI Lab., 1995.
- [14] W. M. Wells III, W. E. L. Grimson, R. Kikinis, and F. A. Jolesz. Adaptive segmentation of MRI data. *IEEE Trans. on Medical Imaging*, 15(4):429–442, 1996.