

INHOMOGENEITY CORRECTION OF MAGNETIC RESONANCE IMAGES BY MINIMIZATION OF INTENSITY OVERLAPPING



J.D. Gispert, S. Reig, J. Pascau, R. Martínez Lázaro, J.J. Vaquero and M. Desco

Medicina y Cirugía Experimental
Hospital General Universitario “Gregorio Marañón”
E- 28007 Madrid. Spain.

ABSTRACT

This work presents a new algorithm (NIC; *Non uniform Intensity Correction*) for the correction of intensity inhomogeneities in magnetic resonance images. The algorithm has been validated by means of realistic phantom images and a set of 24 real images. Evaluation using previously proposed phantom images for inhomogeneity correction algorithms allowed us to obtain results fully comparable to the previous literature on the topic. This new algorithm was also compared, using a real image dataset, to other widely used methods which are freely available in the Internet (N3, SPM'99 and SPM2). Standard quality criteria have been used for determining the goodness of the different methods. The new algorithm showed better results removing the intensity inhomogeneities and did not produce degradation when used on images free from this artifact.

1. INTRODUCTION

Segmentation of magnetic resonance (MR) images is a fundamental procedure for the quantitative study of different brain pathologies such as multiple sclerosis, Alzheimer's disease or schizophrenia. Automatic segmentation of MR scans would be very useful, although difficult to achieve, since different artifacts may be present in the images as a result of the MR acquisition technology.

One of those artifacts is the lack of homogeneity in the radiofrequency (RF) or B1 field, also known as 'illumination artifact'. It consists in a smooth multiplicative variation of the intensity levels across the MR image. In a standard 1.5 Tesla MR scanner for clinical purposes, the magnitude of this intensity variation may even exceed 20% of the signal value. Despite this artifact does not impair the visual inspection of the images, it can degrade noticeably the volumetric quantification of cerebral tissues, particularly when using automatic segmentation algorithms relying on intensity levels. The magnitude of this effect is stronger as the RF

frequency increases, as it is the case in those MR scanners with higher static fields.

This artifact is caused by several reasons, such as the lack of uniformity in the sensitivity of the RF emitting and receiving coils, standing wave effects and the attenuation of the signal inside the object [1]. Provided that attenuation is not equal for different samples under study, this latter factor precludes the calibration of the RF inhomogeneity field to compensate its effect.

This problem is usually addressed by employing offline mathematical algorithms for removing the illumination artifact. Usually, those algorithms correct the MR scans by minimizing (or maximizing) a certain quality criterion related to the goodness of the correction. The bias field can be modeled either by means of a family of basis functions, as in the parametric methods [2-5], or directly as in the non-parametric methods [6-8]. Other possibilities include parametric entropy minimization [9] and homomorphic filtering [10].

Great efforts have been devoted to develop a large number of different methods, thus highlighting the need for inhomogeneity correction algorithms. Unfortunately, the evaluation and comparison of those algorithms has not received the same attention. Arnold et al [11] compared quantitatively six of those algorithms by means of realistic MR phantoms, and qualitatively by means of real MR scans. Their results show that non-parametric algorithms commonly provide a more accurate correction than parametric ones.

Our work describes a new algorithm for the correction of the bias field artifact, based on the minimization of the intensity overlap between different cerebral tissues. This algorithm has been validated by means of a realistic phantom freely available in the Internet. This phantom had been previously used in a comparison of six different correction algorithms [11], thus permitting the comparison between our results and those reported in that work. Furthermore, the algorithm has been evaluated on 24 real images and compared to three of the most widely used correction methods (N3 [8], SPM'99 [2] and SPM2 [5]).

2. MATERIAL AND METHODS

2.1. Description of the NIC algorithm

The rationale of our algorithm (NIC; Non-uniform Intensity Correction) is to use the estimated error rate of a tissue segmentation as a goodness parameter for the correction of the bias field. Modeling tissue intensities by means of a mixture of basis functions allows the calculation of the Intensity Overlapping (IO) as the overlap between those basis functions.

The algorithm requires the presence of only gray and white matter and CSF tissues, thus being necessary to exclude all extra-cranial voxels. A trained operator manually edited an extra-cranial mask obtained by thresholding the original image. The histogram of the masked image is modeled with five basis functions: three gaussians corresponding to pure cerebral tissues (white and gray matter and CSF) and two additional probability densities for partial volume voxels containing only two pure tissues (gray - white matter and CSF - gray matter). The parameters of the basis functions are estimated iteratively applying the EM (Expectation-Maximization) algorithm [12]. At each EM iteration, a Bayesian approach is used to calculate two minimum-error-rate classification thresholds from the three basis functions corresponding to pure tissues. The IO associated to this classification is estimated as the error rate between pure tissue distributions. The EM algorithm stops when the segmentation thresholds do not change in ten consecutive iterations. The result of this step is an estimation of the basis function parameters modeling the image histogram.

The 'bias-free' image is created using the parameters of the basis functions provided by the tissue classification step. All intra-cranial voxels are classified as 'pure tissue' or 'partial volume', being the latter ones those whose intensity level falls within the 95% confidence intervals of any of the two partial volume basis functions. The remaining intra-cranial voxels are classified into gray matter, white matter and CSF by using the minimum-error-rate classification thresholds obtained in the tissue classification step. The value of these pure tissue voxels is set to the mean intensity of their corresponding tissue class (Figure 2b). In both the original and 'bias free' images, the value of all extra-cranial and partial volume voxels is set to the mean intra-cranial intensity to reduce filtering artifacts in the posterior smoothing operation. Then, both images are filtered with the same gaussian kernel whose Full Width at Half Maximum (FWHM) is reduced one third at each iteration (from 200 mm to about 60 mm, depending on the number of iterations).

At this point, the bias field is estimated by the quotient between the smoothed original and 'bias-free' images. However, the estimation of the magnitude of the field can be optimized so as to minimize the IO. To this

end, the original image is corrected with different magnitudes of the estimated bias field, increasing from 0% to 40% in steps of 1%, selecting the bias field magnitude yielding the lowest IO. This goodness index is obtained by using the tissue classification method described above. The final result is a bias corrected MR image calculated as the quotient between the original image and the estimated bias field. This corrected MR image is entered as the original image for the tissue classification step, repeating the whole procedure until no improvement in the classification error is obtained in three consecutive iterations. Note that the parameter minimized by the NIC algorithm is actually the IO, despite the use of a maximum-likelihood EM algorithm in the tissue classification step.

2.2. Validation

2.2.1 Phantom

For the sake of the comparability of our results to previous literature, our evaluation of the NIC algorithm with phantoms followed the methodology described in the work of Arnold et al. [11], that compared six inhomogeneity correction methods. Two known bias fields were added to a realistic MR brain phantom [13] with a 3% of gaussian noise: a paraboloidal field of a maximum magnitude of $\pm 8\%$, and a sinusoidal field of the same magnitude. The algorithm was also applied to the original phantom image in order to quantify possible distortions when no bias field is present. All these images are available in the Internet¹. For both the paraboloidal and sinusoidal fields, the quality of the correction was assessed by means of a scatterplot of the applied vs. the extracted bias fields and its correlation coefficient (r). Ideally, the applied and extracted bias would have exactly the same shape and magnitude. In this case, the regression plot would show a perfect straight line of slope equal to one and a correlation coefficient of one. Similarly, in the case of the plain bias field, the intensities of the original and the corrected images were compared using regression.

2.2.2 Real Images

A total of 24 MR images of healthy volunteers were acquired in two clinical Philips Gyroscan NT MR scanners: 12 in a scanner with a static field of 0.5 Tesla and 12 in a scanner of 1.5 Tesla. MR scans were acquired by using a T1-weighted 3D gradient echo sequence (Flip angle = 30°, Repetition Time = 15.4 ms., Echo Time = 4.6 ms.), matrix size of 256 x 256 x 110, and voxel size of 0.98 x 0.98 x 1.10 mm.

¹ http://pet.med.va.gov:8080/distrib/nu_compare.html

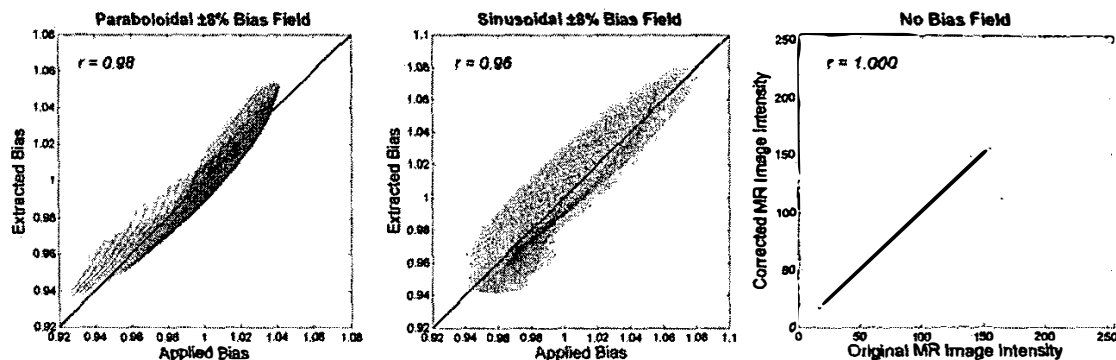


Figure 1. Scatterplots of the applied vs. extracted bias fields for the $\pm 8\%$ paraboloidal (left) and sinusoidal (middle) bias fields, and of the original vs. corrected MR scan intensities (right). Correlation coefficients (r) indicate accuracy in the corrections in the two first cases and lack of distortion in the last one.

Inhomogeneity artifact was corrected in all the images by means of our NIC algorithm and three other methods: N3 [8], SPM'99 [2] y SPM2 [5]. These algorithms were chosen on the basis of its public availability in the Internet² and their wide use by the neuroscience community. The SPM'99 software includes two bias correction techniques, one intended for high magnitude bias fields (SPM-S) and one for moderated bias fields (SPM-L). Input parameters in all algorithms were left at their default values.

In order to evaluate the quality of the corrections, four different goodness criteria were selected: the coefficient of variation of the white and gray matters, their coefficient of joint variation and an estimation of the segmentation error calculated as the overlap between the gaussian basis functions that model image intensities. The coefficient of joint variation is calculated as the sum of the standard deviations of the voxel intensities in the two tissue classes; normalized by the difference of their means. The coefficient of variation of the white and gray matters is a widely employed goodness parameter [8, 9]. The coefficient of joint variation was first proposed in [9] as an approximation to segmentation error. We propose the use of the intensity overlap as a direct measurement of the classification quality.

3. RESULTS

3.1. Phantom

Scatterplots of the applied vs. extracted bias fields are shown in Figure 1 for the $\pm 8\%$ paraboloidal field and the $\pm 8\%$ sinusoidal field, as well as the correlation between the original vs. the corrected intensities when no bias

field was present. Coefficients of correlation were $r = 0.98$ for the paraboloidal bias, $r = 0.96$ for the sinusoidal bias and $r = 1.000$ in the case in which no bias was applied.

3.1. Real Images

Mean estimated segmentation error of the twelve original 0.5 Tesla images was of 10.29% and of 9.08% in the twelve 1.5 Tesla images. All four goodness criteria were calculated in the original and corrected images and then subtracted in order to determine improvement in the image quality.

Table 1 shows the mean values of the differences between goodness criteria in the original and corrected images for the twelve 0.5 Tesla and the 12 1.5 Tesla scans. Lower values indicate higher improvement in the image quality.

	CV-W	CV-G	CJV-WG	IO
0.5 Tesla (N = 12)				
SPM99-L	3.75 %	-0.53 %	20.40 %	0.85 %
SPM99-S	4.74 %	-1.58 %	23.28 %	0.54 %
SPM2	-0.10 %	4.07 %	13.42 %	0.24 %
N3	-0.24 %	-0.83 %	4.90 %	0.22 %
NIC	-0.37 %	-0.64 %	-2.94 %	-0.50 %
1.5 Tesla (N = 12)				
SPM99-L	-0.26 %	0.59 %	22.05 %	3.82 %
SPM99-S	-0.99 %	-0.44 %	9.48 %	1.53 %
SPM2	0.98 %	-0.78 %	4.77 %	0.69 %
N3	-0.64 %	-0.28 %	0.03 %	-0.24 %
NIC	-0.68 %	-1.05 %	-3.15 %	-1.04 %

Table 1. Results of four goodness parameters for the comparison of bias field correction algorithms with real images: coefficient of variation of the white (CV-W) and gray (CV-G) matters, coefficient of joint variation of the white and gray matters (CJV-WG), and intensity overlap (IO). Figures indicate mean diff between the original and corrected scans. Lower values denote greater improvement.

² N3: <http://www.bic.mni.mcgill.ca/software/N3>

SPM'99: <http://www.fil.ion.ucl.ac.uk/spm>

SPM2: <http://www.fil.ion.ucl.ac.uk/spm/spm2b.html>

4. DISCUSSION

On phantom images, our results on the evaluation of the NIC algorithm are directly comparable to those reported in [11], as long as the same methodology and datasets were used. When correcting the $\pm 8\%$ parabolic bias field, the NIC algorithm yielded the same correlation coefficient ($r = 0.98$) than the best algorithms in the comparison: N3 [8] and BFC [4]. Similarly, the NIC algorithm achieved the same quality in the correction of the $\pm 8\%$ sinusoidal bias field ($r = 0.96$) than the N3 method, that also showed the best behavior in this case. Finally, when correcting phantom images without bias field, the NIC method showed the best result in the comparison. Regarding the rest of the algorithms, the HUM method [10] yielded the best result ($r = 0.999$) and the N3 method achieved a correlation coefficient of $r = 0.997$. Bringing together all the results of the phantom evaluation, it can be stated that the NIC algorithm shows a better performance than those evaluated in [11], as it reached the best results in both correcting the bias fields, and not distorted the MR scans when no bias field was present.

On real images, the magnitude of the correction was always higher in the 1.5 Tesla images than in the 0.5 Tesla ones, denoting a greater intensity inhomogeneities in the former images, as expected. Results in Table I show that the NIC algorithm achieved a higher reduction in the coefficient of joint variation between the white and gray matters, as well as in the intensity overlap. This indicates a higher reduction in the segmentation error and, therefore, lower intensity inhomogeneities. Despite reductions in segmentation error might seem too small to be significant, it must be noted that they represent a mean improvement of 4.86 % for the 0.5 Tesla scanner images and of 11.45 % for the 1.5 Tesla ones, with respect to the original mean segmentation errors of 10.29 % and 9.08 %, respectively. Segmentation error is lower in the 1.5 Tesla scans than in the 0.5 Tesla ones, as expected given the better signal to noise ratio achieved by higher static field MR scanners.

The N3 method obtains a higher reduction in the coefficient of variation of the white matter in the 0.5 Tesla images as well as the SPM-S method in both the 0.5 and 1.5 Tesla images. However, their worse results in the coefficient of variation of the gray matter make them fail to reach the overall performance of the NIC algorithm. Moreover, it must be remarked that the NIC method is the only one that never distorts the original scans, allowing its use for images from both high and low static field MR scanners. This better behavior of our algorithm could be explained by the explicit modeling of partial volume voxels in the segmentation step, and its subsequent rejection when estimating the bias field.

5. REFERENCES

- [1] J. G. Sled and G. B. Pike, "Standing Wave and RF Penetration Artifacts Caused by Elliptic Geometry: An Electrodynamical Analysis of MRI." *IEEE Transactions on Medical Imaging*, vol. 17, pp. 653-662, 1998.
- [2] J. Ashburner and K. J. Friston. "Voxel Based Morphometry: The Methods," *NeuroImage*, vol. 11, pp. 805-821, 2000.
- [3] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. "Automated Model-Based Bias Field Correction of MR Images of the Brain," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 885-896, 1999.
- [4] D. W. Shattuck, S. R. Sandor Leahy, K. A. Schaper, D. A. Rottenberg, and R. M. Leahy, "Magnetic Resonance Image Tissue Classification Using a Partial Volume Model." *NeuroImage*, vol. 13, pp. 856-876, 2001.
- [5] J. Ashburner, "Another MRI bias correction approach," presented at 8th International Conference on Functional Mapping of the Human Brain. Sendai, Japan, 2002.
- [6] 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, pp. 429-442, 1996.
- [7] R. Guillemaud and M. Brady, "Estimating the Bias Field of MR Images," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 238-251, 1997.
- [8] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data." *IEEE Transactions on Medical Imaging*, vol. 17, pp. 87-97, 1998.
- [9] B. Likar, M. A. Viergever, and F. Pernus, "Retrospective Correction of MR Intensity Inhomogeneity by Information Minimization," *IEEE Transactions on Medical Imaging*, vol. 20, pp. 1398-1410, 2001.
- [10] B. H. Brinkmann, A. Manduca, and R. A. Robb, "Optimized Homomorphic Unsharp Masking for MR Grayscale Inhomogeneity Correction," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 161-171, 1998.
- [11] J. B. Arnold, J. S. Liow, K. A. Schaper, J. J. Stern, J. G. Sled, D. W. Shattuck, A. J. Worth, M. S. Cohen, R. M. Leahy, J. C. Mazziotta, and D. A. Rottenberg, "Qualitative and Quantitative Evaluation of Six Algorithms for Correcting Intensity Nonuniformity Effects." *NeuroImage*, vol. 13, pp. 931-943, 2001.
- [12] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm." *Journal of the Royal Statistical Society*, vol. 39, pp. 1-38, 1977.
- [13] 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, pp. 463-468, 1998.