



# Validating MRI Field Homogeneity Correction Using Image Information Measures

N. A. Thacker, A. J. Lacey and P. A. Bromiley.  
Imaging Science and Biomedical Engineering,  
Stopford Building, University of Manchester,  
Oxford Road, Manchester, M13 9PT  
[a.lacey], [n.thacker]@man.ac.uk  
<http://www.tina-vision.net/>

## Abstract

For image analysis techniques to be of utility in medical diagnosis systems it is necessary to be able to perform quality control over the results they produce. Input data must conform to the assumptions within the algorithm if useful results are to be achieved. Automation of this process is essential if vision algorithms are to form components in analysis systems.

In this paper we present a technique to validate the correction of field inhomogeneity in MR images. The initial intention was to use information measures to check the improvement due to correction. However, it will be shown that the standard log entropy calculation for information measurement does not have the required properties, specifically grey-level scale invariance. We present an alternative, scale-invariant information measure derived using conventional likelihood approaches, that can be applied as an absolute measure of information content. We show this technique in use for the validation of our existing coil correction method [9].

## 1 Introduction and Background

The goal of field inhomogeneity correction can be stated as seeking a multiplicative correction,  $\alpha(x, y)$  for an image,  $g(x, y)$  such that homogeneous regions of equivalent tissue have the same mean grey level value in the image  $\alpha(x, y)g(x, y)$ . Several algorithms that attempt to gain correct volume data have been presented in the literature. Those based on fluid phantom models are able to correct for RF inhomogeneities, but do not take into account patient anatomy and orientation [11]. Methods requiring prior tissue classification or modelling are dependent upon expert supervision, in practise this includes the majority of those classed as *automatic* [1, 2].

The N3 algorithm presented in [5], uses an interactive optimisation strategy to estimate the field correction under the assumption of Gaussian intensity distortions. Whilst it provides a genuinely automatic algorithm, it has several limitations. First, it requires the segmentation of the foreground from the background. Although in a typical image this is trivial to achieve for the majority of pixels, it becomes difficult to correct intensity in those regions where the signal drops close to the background level. The iterative nature of the N3 method coupled with the non-uniform noise levels on the intermediate



data introduces a bias. The iterative process also necessitates sub-sampling in order to reduce the computational load to acceptable levels on typically sized datasets. Finally, the approximately Gaussian nature of the field has only been estimated in head coil data and may not be accurate for different anatomical regions or levels of distortion.

Our own work has produced a statistically based general purpose method for the estimation of field homogeneity which makes minimal assumptions regarding the content of the image [9]. Specifically, it weights the evidence supporting local estimates of image slope using error propagation whilst robustly suppressing derivatives at edge boundaries. The resulting image slopes in the x and y direction are then re-integrated to estimate the bias field. The algorithm assumes only locally homogenous regions separated by edge boundaries and this minimal basis supports the analysis of a wider variety of image formation processes than those which assume particular tissue density models or bias field distributions. However, the technique will fail if presented with data which does not conform to even these minimal assumptions. Such behaviour is **an inevitable consequence of any algorithm which makes assumptions of data** (to a greater or lesser extent this must include all algorithms). If the techniques of image analysis are to develop into usable components in medical diagnosis systems the issue of quality control needs to be addressed. Specifically there must be some way of confirming that the assumptions underpinning an algorithm design have been upheld.

This paper presents a technique to monitor the success of field inhomogeneity correction on MR data sets so that such techniques can be used reliably in an automated fashion. Entropy measures have previously been suggested as the basis for the design of algorithms to correct image inhomogeneity [3]. The intention here was to use an entropy measure as a complementary statistic to monitor the success of the correction procedure.

## 2 Information Measures

Shannon defined a measure of information content analogous to the *entropy* measure used in statistical mechanics. Given a set of  $A$  grey level measurements  $\{g_a\}$  we can write the probability of finding a measurement within interval  $i$  of width  $\Delta_g$  as  $P_i(\{g_a\})$ . The entropy of this set of measurements is then expressed as

$$H = - \sum_i^N P_i(\{g_a\}) \log P_i(\{g_a\}) \quad (1)$$

The value of  $H$  is non-negative and takes on its maximum when the distribution of  $P_i(\{g_a\})$  is flat and is often regarded as a measure of the “peakiness” of the image histogram. When used for inhomogeneity correction, the entropy measure is maximised (or its negative minimised) on the assumption that field non-uniformity introduces information (in the Shannon sense) into the histogram.

Unfortunately, measures of entropy have several undesirable characteristics. First, it is possible to get interference between distinct regions of data which happen to have similar but different mean grey level values. An optimisation algorithm will achieve an *improvement* by combining such data into a single peak. Of more concern however, is the lack of invariance to scale changes in the ordinal values (grey levels) of observed data. In other words,



$$\sum_i^N P_i(\{g_a\}) \log P_i(\{g_a\}) \neq \sum_i^N P_i(\{kg_a\}) \log(P_i(\{kg_a\})) \quad (2)$$

This poses a particular problem when using the measure to assess arbitrary scaling in grey levels, as in the case of the coil correction problem, because the maximum is always obtained with the null correction field ( $0 \times g_a$ ). This inevitably biases the result towards the trivial solution of  $k = 0$ . Therefore,  $H$  cannot be used as an absolute measure of information content. It is not possible to correct this using scaling processes based upon data dependent parameters such as maximum and minimum values. Only features such as local maxima and minima can reliably be used for algorithmic purposes.

A more direct method of solving the bias problem involves working with the distribution of the logarithm of grey levels  $P(\{\log(g_a)\})$  instead. Scale changes in this distribution result in simple shifts in the histogram but no systematic changes in the information measure. However, for images with uniform random noise (the most common form of noise model), the logarithm transform also produces histograms with a sensitivity to intrinsic measurement accuracy, which varies across the dynamic range of the variable. This is an undesirable property if we wish to approximate the information measure using histograms rather than continuous analytic approaches. However, this approach does point the way to a more statistically justifiable measure of information content.

## 2.1 Scale invariant information measure

We can consider an information measure as a way of assessing the difference between the observed histogram and an uninformative (flat) distribution. With this in mind, we can define the statistical information content of an image based on the likelihood of the histogram of the logarithm of image data compared to a flat, uninformative distribution. Although this could be defined as a standard chi-squared statistic it is generally not well known that it is more accurately measured by the Matusita measure for the difference between two normalised distributions  $P$  and  $Q$ .

$$M = \sum_i^N (\sqrt{P_i} - \sqrt{Q_i})^2 \quad (3)$$

The square-root transforms Poisson distributed variables into Gaussian ones so that a squared difference is appropriate across the entire range of probability values. As a difference measure, this has the property of being well behaved for large statistical differences [6]. This property is not important when the data and assumed distribution are in close agreement (as is generally the case in log-likelihood approaches). However, in this work we are interested in achieving the largest possible difference between the data and the assumed distribution and thus are interested in a measure which is valid over all separations.

For a logarithmic image histogram and constant  $Q$  (i.e.  $Q_i = Q$ ) this can be written (up to a scale factor) in the equivalent Bhattacharyya form as

$$B = \sqrt{Q} \sum_i^N \sqrt{P_i(\{\log(g_a)\})} \quad (4)$$



Although we are seeking the worst possible agreement between  $P$  and  $Q$ ,  $Q$  is currently an unknown and the value required here is that which gives the best possible match to  $P$ . Differentiating  $M$  with respect to  $Q$  and setting to zero we find that the optimal value for  $Q$  is

$$\sqrt{Q} = \frac{1}{N} \sum_i^N \sqrt{P_i(\{\log(g_a)\})} \quad (5)$$

Therefore

$$B = \frac{1}{N} \left( \sum_i^N \sqrt{P_i(\{\log(g_a)\})} \right)^2 \quad (6)$$

Therefore we will get the same characteristic behaviour of the measure regardless of whether we use optimal scaling of  $Q$  or a fixed arbitrary constant. However, this scaling needs to be made explicit if we wish to compute covariances. We will continue to work with an explicit  $Q$  term for simplicity.

The formulations above consider the logarithm of the image data,  $\log(g_a)$ , but we wish to work with the original image data if we are to base the estimation process on the construction of quantised histograms. Importantly, because  $B$  simply relates relative frequencies observed in two distributions, and the quantisation (binning) of these distributions is arbitrary, equivalent forms of this statistic can be generated by any monotonic co-ordinate mapping<sup>1</sup>. As a consequence, we can now determine the equivalent form of this measure for the original image histogram by applying an appropriate transformation to the comparison distribution. In fact a histogram bin at location  $g_i$  in the original histogram should correspond to a quantity proportional to  $(\log(g_{i+1}) - \log(g_i))$  in the logarithmic distribution,

$$B = \sqrt{Q} \sum_i^N \sqrt{P_i(\{g_a\})} \sqrt{(\log(g_{i+1}) - \log(g_i))} \quad (7)$$

writing this as;

$$B = \sqrt{Q} \sum_i^N \sqrt{\frac{P_i(\{g_a\})}{\Delta_g}} \sqrt{\frac{(\log(g_i + \Delta_g) - \log(g_i))}{\Delta_g}} \Delta_g \quad (8)$$

and letting  $\Delta_g \rightarrow dg$  and  $P(\{g_a\}) \rightarrow p(g) dg$  we can generate the continuous form of the measure;

$$B \rightarrow \sqrt{q} \int \sqrt{p(g)/g} dg \quad (9)$$

where  $q$  is the continuous form of  $Q$ . Thus a uniform distribution  $q$  in the logarithm domain corresponds to a  $q/g$  distribution for the original grey levels. Again the scaling of  $q$  is arbitrary and can therefore be taken as unity for practical purposes.

This measure is invariant to scaled co-ordinate changes,  $g \rightarrow \alpha g$ . It also has the property that making a change in the measurement domain selected from a set of scaled

<sup>1</sup>The continuous form of the measure has complete invariance to redefinition of the ordinal space.



measure	strong	weak	noise model ( $\sigma_g$ )
$\int p(g) \log(p(g)) dg$	$g + \gamma$	$\alpha g + \gamma$	uniform additive
$\int p(\log(g)) \log(p(\log(g))) dg$	$\alpha g$	$\alpha g^\beta$	uniform multiplicative
$\int \sqrt{p(\log(g))} dg$	$\alpha g$	$\alpha g^\beta$	uniform multiplicative
$\int \sqrt{p(g)/g} dg$	$\alpha g$	$\alpha g^\beta$	uniform additive
$\int \sqrt{p(\sqrt{g})/\sqrt{g}} d\sqrt{g}$	$\alpha g$	$\alpha g^\beta$	Poisson

Table 1: Invariance Characteristics

polynomial mappings, changes the measure by only a normalisation factor. This means that the location of the minima  $g_{max}$  of this function (defining the maximum difference) is unaffected. We can call this **weak** invariance, as it is not a true invariant of the measure though it will result in equivalent results for optimisation algorithms,

$$B_w = \int \sqrt{p(w)/w} dw \propto B \quad (10)$$

We can similarly refer to complete invariance of the statistic to a transformation as **strong** invariance.

$$B_s = \int \sqrt{p(s(g))/g} dg = B \quad (11)$$

We can prove polynomial weak invariance for the new measure by observing that

$$B_w = \int \sqrt{p(w)/w} dw = \int \sqrt{\frac{p(g)}{\partial w(g)/\partial g} \frac{1}{w} \frac{\partial w(g)}{\partial g}} dg \quad (12)$$

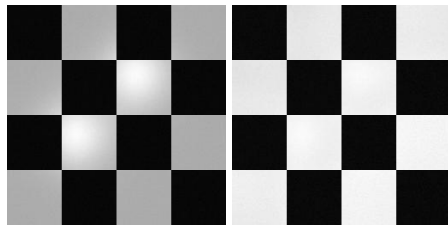
which is simply a scaled version of the original statistic ( $\beta^{0.5}B$ ) provided that  $w(g)$  is a solution of

$$w(g) = \beta^{-1} g \frac{\partial w(g)}{\partial g} \quad (13)$$

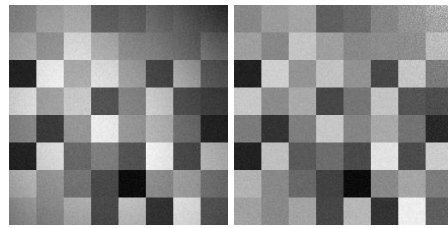
The general solution for  $w(g)$  is  $w(g, \alpha, \beta) = \alpha g^\beta$ . Such invariance is perhaps to be expected for a sensible (continuous) information measure but is clearly absent for the standard entropy measure unless computed using the logarithm intensity histogram (Table 1). These invariants carry over into the discrete forms of these measures (constructed from sampled data) as statistical equivalences. It is also useful for practical calculation of the measure in discrete form, as it can be used to match histogram binning to the noise model for the measurement process ( e.g.  $\beta = 0.5$  for Poisson processes). This is summarised for three common noise models in Table 1, but more generally, if the mapping required to represent the data with uniform errors is known then a suitable (scale invariant) form for the Bhattacharyya measure can always be constructed by the process presented here.

### 3 Experiments

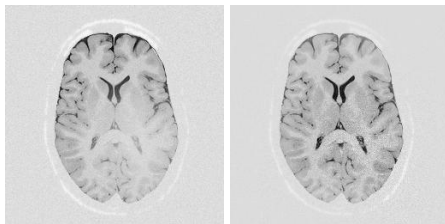
We have taken a set of eight test images, three simulated with known ground truth and five real MR images with known uniform additive noise and manually verified coil correction.



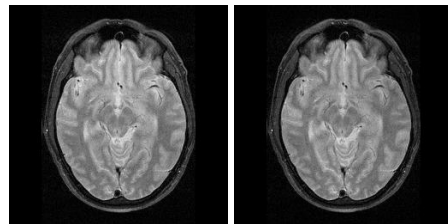
(a) checker



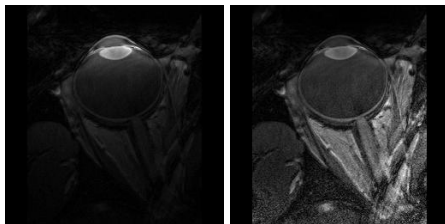
(b) drift



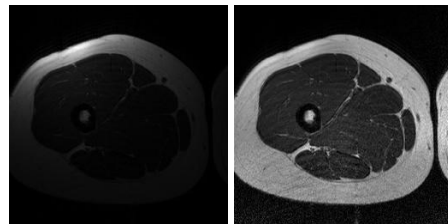
(c) phantom



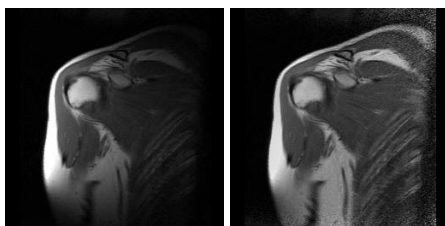
(d) brain



(e) eye



(f) leg



(g) shoulder



(h) spine

Figure 1: Original (left) and corrected (right) images



The three simulated images comprise; one generally used to demonstrate field homogeneity correction algorithms in the literature; one specifically designed to test the edge boundary capabilities of the slope estimation; and another constructed from a modified image of a real brain. The remaining images comprise a set of real images from a variety of clinical MR scenarios with various levels of gain inhomogeneity and spatially varying signal to noise. First we compute the summed robust error between the observed  $x$  and  $y$  derivatives in the estimated coil correction surface and the original data ( $L_{rg}$ ). This measure is implicitly optimised by our derivative based correction technique. Second we compute the standard log entropy measure,

$$L_{le} = \int p(g) \log(p(g)) dg \quad (14)$$

with the histogram bin size fixed a priori (as suggested by [3]). Third we compute the likelihood based scale invariant statistic,

$$L_{si} = \int \sqrt{p(g)/g} dg \quad (15)$$

which estimates the difference between the histogram and the information free distribution (uniform in the log domain).

In order to test the above measures and our theoretical analysis we have computed these three estimates of correction performance as a function of a scaled modification of the known correction,  $\alpha(x, y)^\zeta$ .  $\zeta$  greater than 1 implies more correction, less than 1 implies less correction,  $\zeta = 0$  implies no correction. We would expect useful measures to demonstrate appropriate local minima or maxima around  $\zeta = 1$  as this has already been visually validated as an appropriate correction. If any measure does not demonstrate such an optima then it could not be used as the basis for a correction procedure. In addition, if the local optima is not better than the original uncorrected image then we know that it cannot be used as an absolute measure of information content or as the basis for monitoring correction performance.

## 4 Results

The resulting curves for each statistic on each of the datasets are presented in Figure 2 and the images (both original and corrected) are presented in Figure 1. The dotted curve in each graph demonstrates that a distinct minima is located for all datasets using the robust gradient measure  $L_{rg}$ , confirming the validity of the approach. The displaced minima of this curve for the spine image (h) is consistent with the problem of over-correction which can occur with this algorithm in very low signal to noise situations. The entropy calculation  $L_{le}$  (solid curve) fares less well with half of the image datasets failing to generate a maxima in the region of the known correction result. Also, this measure often displays the wrong trend altogether, implying that (in direct contradiction to the results presented in [3]) it cannot be used as an absolute measure of image quality. The new method  $L_{si}$  (dashed curve) shows a maxima for all but the shoulder image (g). Even in the case of (g) the value of the measure has increased, implying an improved image. We believe that the lack of a distinct maxima for this data is evidence of an interference process due to the particular alignment of data in the histogram as described in the introduction, which is a potential problem for all histogram based methods.

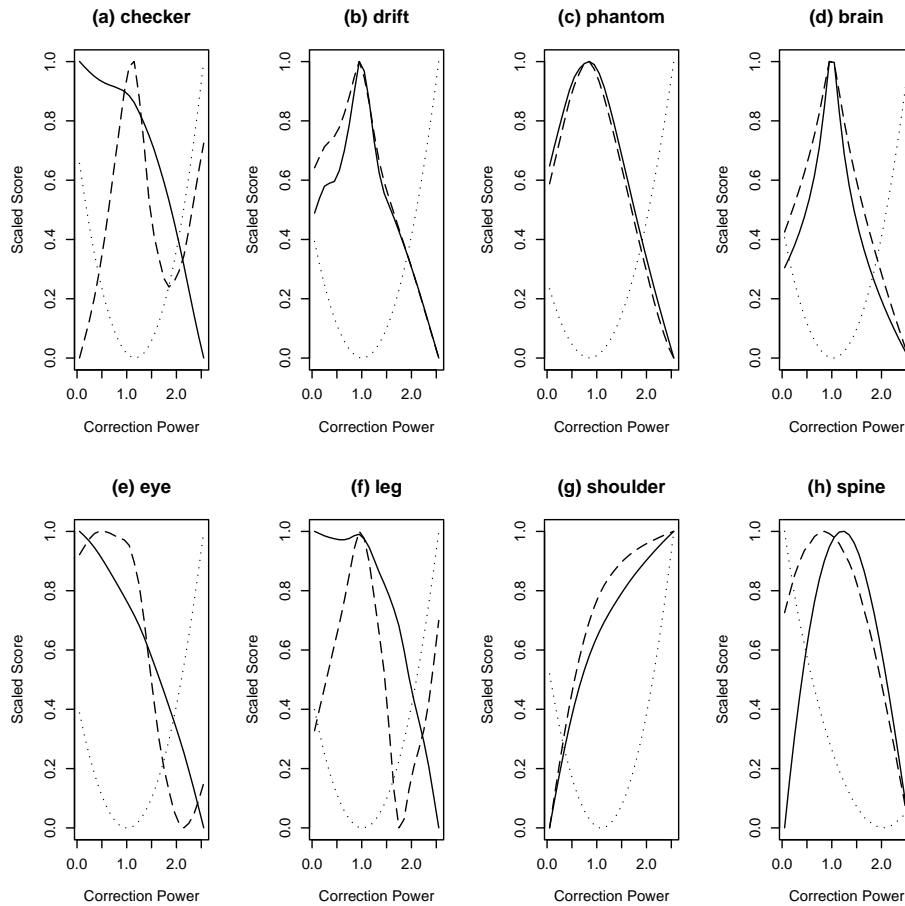


Figure 2: Correction power graphs.  $L_{rg}$  = dotted line,  $L_{le}$  = solid line,  $L_{si}$  = dashed line

## 5 Discussion and Conclusions

The results indicate that our coil correction procedure is consistent with a non-iterative robust fitting of local image derivatives. In addition it demonstrates that a log entropy based scheme cannot be used as the basis for coil correction, without some modification to avoid the bias towards the trivial solution of multiplication by zero. The need for such a correction appears to be fundamental to the use of such a measure if it is ever to be interpreted as an absolute quantity, though is outside the scope of the original theory. This observation casts doubt as to the theoretical validity of using entropy as a measure of image information content. Indeed, it would probably be argued that if we were following the strict Shannon definition of information then the information content of a grey level measurement is the number of bits required to represent the signal when scaled to the noise. As multiplication boosts both of these factors equally the contribution to Shannon



entropy from the image is strictly unchanged. The changing term would instead come from our correction field. This kind of confusion between the use of entropy measures as surrogates for similarity measures is quite common. We agree with other authors, that these approaches be interpreted instead as approximations to bootstrapped likelihoods [4].

The statistical measure presented in this paper, equation 15, works as expected and an analysis such as that presented above can be used as independent confirmation of the successful application of the coil correction process to unfamiliar data. Problems of interference between similar but distinct tissue regions prevent us from recommending this measure as the sole basis for a correction algorithm. Practical use of this algorithm could introduce systematic bias in subsequent quantitative analysis. We would recommend robust image gradient based methods as the approach which makes a minimum number of assumptions regarding the image data while giving the possibility of an unbiased solution.

Now that the main principle of construction of likelihood measures with required invariance has been established, there is no particular need to restrict ourselves to one dimensional histograms. Information measures for two dimensional histograms based upon mutual information (where information is derived from a log-entropy measure [8]) have been widely used in the literature to define similarity measures for problems such as co-registration of medical data [10]. Despite attempts to relate these approaches to likelihood measures, they can not be used quantitatively (and are therefore not statistically valid) due to the inability to construct an appropriate covariance. This suggests that more work is needed to understand the origins of these measures and more justifiable measures should be sought. We are now in a position to suggest equation 16 as a likelihood based measure with strong scale invariant and weak polynomial invariant characteristics which is appropriate for grey levels  $g_1$  and  $g_2$  acquired with uniform additive noise;

$$B = \int \int \sqrt{\frac{p(g_1, g_2)}{g_1 g_2}} dg_1 dg_2 \quad (16)$$

It should be emphasised that robust calculation of this measure will require statistically valid procedures for stabilisation such as the process of *modal arithmetic* [7]. It is our intention to investigate this measure as a basis for co-registration of medical data volumes in future work.

## 6 Acknowledgements

The authors wish to thank the support of the European Commission (PCCV and OS-MIA projects) and the EPSRC/MRC MIAS IRC project. We also wish to thank personal contributions from Bill Crum, Carole Twining and Tim Cootes at various stages of the development of this work.

## References

- [1] M. Özkan, B. M. Dawant, and R. J. Maciunas. Neural network based segmentation of multimodal medical images: A comparative and prospective study. *IEEE Trans. Med. Imag.*, 12:534–544, September 1993.



- [2] W. M. Wells III, W. E. L. Grimson, R. Kikiris, and F. A. Jolesz. Adaptive segmentation of mri data. *IEEE Trans. Med. Imag.*, 15(4):429–442, 1996.
- [3] Boštjan Likar. *Registration and Restoration of Medical Images*. PhD thesis, University of Utrecht, 2000.
- [4] A. Roche, G. Malandian, N. Ayache, and S. Prima. Towards a better comprehension of similarity measures used in medical image registration. In *Proceedings of MICCAI'99*, pages 555–566, September 1999.
- [5] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. A nonparametric method for automatic correction of intensity non-uniformity in mri data. *IEEE Trans. Med. Imag.*, 17(1):87–97, February 1998.
- [6] N. A. Thacker, F. Aherne, and P. I. Rockett. The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika*, 34(4):363–368, 1997.
- [7] N. A. Thacker and A. J. Reader. Modal division and its application to medical image analysis. In *Proc. MIUA*, pages 7–10, July 2000.
- [8] P. Viola. *Alignment by Maximisation of Mutual Information*. PhD thesis, M.I.T., 1995.
- [9] E. Vokurka, N. A. Thacker, and A. Jackson. A fast model independent method for automatic correction of intensity non-uniformity in mri data. *JMRI*, 10(4):550–562, 1999.
- [10] J. West, J. M. Fitzpatrick, and et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J. Comput. Assist. Tomography*, 21:554–566, 1997.
- [11] D. A. G. Wicks, G. J. Barker, and P. S. Tofts. Correction of intensity non-uniformity in mr images of any orientation. *Mag. Reson. Imag.*, 11:183–196, 1993.