2005

Many Heads are Better than One: Jointly Removing Bias from Multiple MRIs using Nonparametric Maximum Likelihood

Erik G Learned-Miller, *University of Massachusetts - Amherst*
Vidit Jain, *University of Massachusetts - Amherst*

Many Heads are Better than One: Jointly Removing Bias from Multiple MRIs using Nonparametric Maximum Likelihood

Erik G. Learned-Miller and Vidit Jain

Department of Computer Science
University of Massachusetts, Amherst
Amherst, MA 01003, USA,
elm@cs.umass.edu,
WWW home page: http://www.cs.umass.edu/~elm

Abstract. The correction of multiplicative bias in magnetic resonance images is an important problem in medical image processing, especially as a preprocessing step for quantitative measurements and other numerical procedures. Most previous approaches have used a maximum likelihood method to increase the probability of the pixels in a single image by adaptively estimating a correction to the unknown image bias field. The pixel probabilities are defined either in terms of a pre-existing tissue model, or nonparametrically in terms of the image's own pixel values. In both cases, the specific location of a pixel in the image does not influence the probability calculation. Our approach, similar to methods of joint registration, simultaneously eliminates the bias from a set of images of the same anatomy, but from different patients. We use the statistics from the same location across different patients' images, rather than within an image, to eliminate bias fields from all of the images simultaneously. Evaluating the likelihood of a particular voxel in one patient's scan with respect to voxels in the same location in a set of other patients' scans disambiguates effects that might be due to either bias fields or anatomy. We present a variety of "two-dimensional" experimental results (working with one image from each patient) showing how our method overcomes serious problems experienced by other methods. We also present preliminary results on full three-dimensional volume correction across patients.

1 Introduction

The problem of bias fields in magnetic resonance (MR) images is an important problem in medical imaging. We illustrate the problem in Figure 1 using a synthetic image from BrainWeb [10] and an artificial bias field. When a patient is imaged in the MR scanner, the goal is to obtain an image which is a function solely of the underlying tissue (left of Figure 1). However, typically the desired anatomical image is corrupted by a multiplicative bias field (second image) that is caused by engineering issues such as imperfections in the radio frequency coils used to record the MR signal. The result is a corrupted image (third image). (See [1] for background information on bias fields.) The goal of bias correction is to estimate the uncorrupted image from the corrupted image.

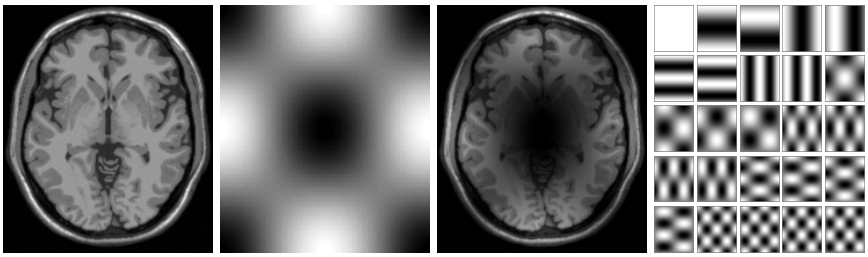

Fig. 1: On the left is an idealized mid-axial MR image of the human brain with little or no bias field. The second image is a simulated low-frequency bias field. It has been exaggerated for ease of viewing. The third image is the result of pixelwise multiplication of the image by the bias field. On the right is the set of basis images used to parameterize smooth bias fields for the slice-based algorithm

Radiologists appear to be remarkably immune to the effects of bias fields under many circumstances.[1] This is probably because radiologists seem to make mostly *relative intensity* judgments based upon local image information. They use so-called window-level adjustments to optimize local contrast for discriminating various properties of the tissues in a specific region. Bias fields, however, are a major problem for automated computer applications like registration, segmentation or pre-screening which depend upon similar tissues having consistent values across a scan. In these applications, the actual numeric brightness value assigned to a tissue is critical and directly affects whether such algorithms will work.

A variety of statistical methods have been proposed to address this problem. Wells et al. [9] developed a statistical model using a fixed number of tissues, with the brightness distribution for each tissue type (in a bias-free image) represented by a one-dimensional Gaussian distribution or by a nonparametric distribution. An expectation-maximization (EM) procedure was then used to simultaneously estimate the bias field, the tissue type, and the residual noise. While this method works well in many cases, it has several drawbacks: (1) Models must be developed *a priori* for each type of acquisition (for each different setting of the MR scanner), for each new area of the body, and for different patient populations (like infants and adults). (2) Models must be developed from "bias-free" images, which may be difficult or impossible to obtain in many cases. (3) The model assumes a fixed number of tissues, which may be inaccurate. For example, during development of the human brain, there is continuous variability between gray matter and white matter. In addition, a discrete tissue model does not handle so-called partial volume effects in which a pixel represents a combination of several tissue types. This occurs frequently since many pixels occur at tissue boundaries.

Tissue-free modeling approaches have also been suggested, as for example by Viola [11]. In that work, a nonparametric model of brightness values was developed from a single image. Using the observation that the entropy of the pixel brightness distribution

[1] Anecdotally, moderate bias fields do not seem to significantly effect radiologists' ability to make diagnoses.

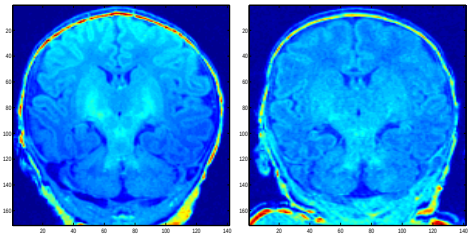

Fig. 2: The infant brain image on the left shows a coronal MR image with a strong bias field. The image is too bright at the top and too dark at the bottom. This is easy to see and can be corrected successfully by a variety of bias correction techniques. The right image, however, is a more difficult case. In particular, the subtle increase in intensity in the middle of the image is, from an algorithmic point of view, difficult to categorize. Is it a subtle increase in intensity due to a low frequency bias field, or is it a slight increase in intensity due to say, partial myelination of white matter in a developing infant? Due to the location of the increased intensity, a radiologist would usually guess that this is developing white matter in an infant brain, but algorithms that do not take into account spatial location and the appearance of other similar scans cannot make such an assessment. It is exactly this sort of information which is leveraged by our algorithm.

for a *single image* is likely to increase when a bias field is added, Viola's method postulates a bias-correction field by minimizing the entropy of the resulting pixel brightness distribution. This approach addresses several of the problems of fixed-tissue models, but has its own drawbacks: (1) The statistical model may be weak, since it is based on data from only a single image. (2) There is no mechanism for distinguishing between certain low-frequency image components and a bias field. That is, the method may mistake signal for noise in certain cases when removal of the true signal reduces the entropy of the brightness distriibution. We illustrate this problem in Figure 2.

The present method, first presented in [5] overcomes or improves upon problems associated with both of these methods and their many variations (see, e.g., [1] for recent techniques). It models tissue brightness nonparametrically, but uses data from multiple images to provide improved distribution estimates and alleviate the need for bias-free images for making a model. Most importantly, it conditions the distributions on spatial location, taking advantage of a rich information source ignored in other methods. Experimental results demonstrate the effectiveness of our method.

2 The Image Model and Problem Formulation

We assume we are given a set $\mathbf{I}$ of observed images I_i with $1 \leq i \leq N$, as shown on the left side of Figure 3. Each of these images is assumed to be the product of some bias-free image L_i and a smooth bias field $B_i \in \mathcal{B}$. We shall refer to the bias-free images as *latent images* (also called *intrinsic images* by some authors). The set of all latent images shall be denoted $\mathbf{L}$ and the set of unknown bias fields $\mathbf{B}$. Then each observed image can be written as the product $I_i(x,y) = L_i(x,y) * B_i(x,y)$, where (x,y) gives the pixel coordinates of each point, with P pixels per image.

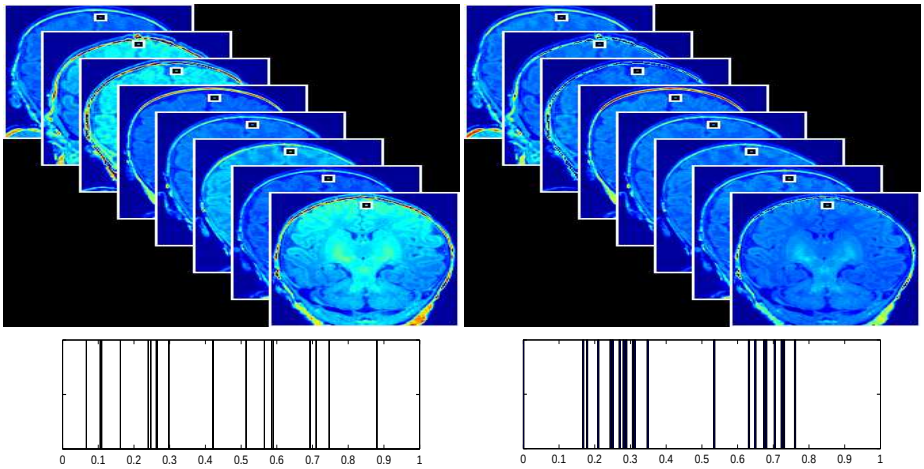

Fig. 3: **Top left.** A set of mid-coronal brain images from eight different infants, showing clear signs of bias fields. A *pixel-stack*, a collection of pixels at the same point in each image, is represented by the small square near the top of each image. The plot beneath the images shows the values of the pixels in the pixel stack (plus points from an additional 13 images). Note the wide distribution (high entropy) of brightness values in the stack. The estimated entropy of this distribution was -0.4980. **Top right.** The same mid-coronal images after bias correction. Note the uniformity of the images and the higher concentration (lower entropy) of brightness values from the pixel stack. The estimated entropy for these samples was -0.8389.

Consider again Figure 3. A *pixel-stack* through each image set is shown as the set of pixels corresponding to a particular location in each image (not necessarily the same tissue type). Our method relies on the principle that the pixel-stack values are likely, on average, to have lower empirical entropy when the bias fields have been removed. We now explain what exactly this means and why it should be true.

2.1 Entropy, Nonparametric Distributions, and Maximum Likelihood

Consider some infinite set of images taken from a fixed population, such as mid-coronal images of infants between zero and two years of age. Now pick a particular location in each image, such as the middle pixel. The distribution over *tissue values* at this location, across the images, is a random variable (call it T). We might expect white matter, cerebrospinal fluid, vasculature, or a handful of other tissue at this location, each with some relative frequency. The *entropy* (defined formally below) of this random variable gives us a measure of the variability of tissues at this location.

In MR images with no bias fields, each tissue is mapped to a fairly consistent brightness value, another random variable (call it L, for latent image brightness). Thus, the entropy of the tissue types at a particular spatial location is closely related to the entropy of brightness values in bias-free MRs at that location. An empirical sample of true brightness values from such a set of images is in the lower right of Figure 3.

Now consider what happens when random bias fields are *introduced* into each image (going from right to left in Figure 3). If we consider the random variable B to be the contribution of a random bias field to each image, then we will be perturbing the original distribution of brightness values L to values $L \times B$. This tends to spread out the brightness values in the pixel stack, increasing their empirical entropy, as shown by the set of samples on the lower left of Figure 3. In fact, in dealing with an infinite sample, it can be proven [3] that the entropy of a random variable (brightness) will always increase (or remain the same) when an independent random variable is added to it.[2]

The idea that entropy increases when random variables are added together has another interpretation in terms of probability theory. In particular, the average log probability density (which is just the negative entropy) of points in a distribution of one random variable is guaranteed to be *higher* than the average log probability density of another random variable which is the original random variable plus an independent source of randomness. In other words, the probability density of our data under a bias-free distribution should be higher than the probability of our data under distribution that include bias. This is only guaranteed when we have an infinite amount of data, but is usually true even for the case of finite data. This is true irrespective of the form of the distributions. That is, these ideas make no assumptions about the *parametric form* of the distributions, and are thus completely nonparametric. It is these ideas upon which our method is based. We now describe the specifics of our model and method.

2.2 The Model

The latent image generation model assumes that each pixel is drawn from a fixed distribution $p_{x,y}(\cdot)$ which gives the probability of each gray value at the the location (x,y) in the image. Furthermore, we assume that all pixels in the latent image are independent, given the distributions from which they are drawn. It is also assumed that the bias fields for each image are chosen independently from some fixed distribution over bias fields. Unlike most models for this problem which rely on statistical regularities within an image, we take a completely orthogonal approach by assuming that pixel values are independent given their image locations, but that pixel-stacks in general have low entropy when bias fields are removed.

We formulate the problem as a maximum a posteriori (MAP) problem, searching for the most probable bias fields given the set of observed images. Letting $\mathcal{B}$ represent the 25-dimensional product space of smooth bias fields (corresponding to the 25 basis images of Figure 1), we wish to find

$$\arg\max_{\mathbf{B}\in\mathcal{B}} P(\mathbf{B}|\mathbf{I}) \stackrel{(a)}{=} \arg\max_{\mathbf{B}\in\mathcal{B}} P(\mathbf{I}|\mathbf{B})P(\mathbf{B}) \tag{1}$$

$$\stackrel{(b)}{=} \arg\max_{\mathbf{B}\in\mathcal{B}} P(\mathbf{I}|\mathbf{B}) \tag{2}$$

$$\stackrel{(c)}{=} \arg\max_{\mathbf{B}\in\mathcal{B}} P(\mathbf{L}(\mathbf{I},\mathbf{B})) \tag{3}$$

[2] Here we are *multiplying* random variables rather than adding them, so this result does not strictly apply. However, when one of the random variables is near 1 (as is the bias random variable) and we force its mean to be 1, this result will usually hold even for multiplication.

$$= \arg\max_{\mathbf{B}\in\mathcal{B}} \prod_{x,y}\prod_{i=1}^{N} p_{x,y}(L_i(x,y)) \tag{4}$$

$$= \arg\max_{\mathbf{B}\in\mathcal{B}} \sum_{x,y}\sum_{i=1}^{N} \log p_{x,y}(L_i(x,y)) \tag{5}$$

$$\overset{(d)}{\approx} \arg\min_{\mathbf{B}\in\mathcal{B}} \sum_{x,y} H(p_{x,y}) \tag{6}$$

$$\overset{(e)}{\approx} \arg\min_{\mathbf{B}\in\mathcal{B}} \sum_{x,y} \hat{H}_{\text{Vasicek}}(L_1(x,y),...,L_N(x,y)) \tag{7}$$

$$= \arg\min_{\mathbf{B}\in\mathcal{B}} \sum_{x,y} \hat{H}_{\text{Vasicek}}(\frac{I_1(x,y)}{B_1(x,y)},...,\frac{I_N(x,y)}{B_N(x,y)}). \tag{8}$$

Here H is the Shannon entropy ($-E(\log P(x))$) and $\hat{H}_{\text{Vasicek}}$ is a sample-based entropy estimator discussed below. (a) is just an application of Bayes rule. (b) assumes a uniform prior over the allowed bias fields. The method can easily be altered to incorporate a non-uniform prior. (c) expresses the fact that the probability of the observed image given a particular bias field is the same as the probability of the latent image associated with that observed image and bias field. The approximation (d) replaces the empirical mean of the log probability at each pixel with the negative entropy of the underlying distribution at that pixel. This entropy is in turn estimated (e) using the entropy estimator of Vasicek [8] directly from the samples in the pixel-stack, without ever estimating the distributions $p_{x,y}$ explicitly.

The inequality (d) becomes an equality as N grows large by the law of large numbers, while the consistency of Vasicek's entropy estimator [2] implies that (e) also goes to equality with large N. (See [2] for a review of entropy estimators.)

2.3 The Entropy Estimatior

The entropy estimator used is similar to Vasicek's estimator [8], given (up to minor details) by

$$\hat{H}_{\text{Vasicek}}(Z^1,...,Z^N) = \frac{1}{N-m}\sum_{i=1}^{N-m} \log\left(\frac{N}{m}(Z^{(i+m)} - Z^{(i)})\right), \tag{9}$$

where Z^i's represent the values in a pixel-stack, $Z^{(i)}$'s represent those same values in rank order, N is the number of values in the pixel-stack and m is a function of N (like $N^{0.5}$) such that m/N goes to 0 as m and N go to infinity. These entropy estimators are discussed at length elsewhere [4].

To understand the intuition behind this estimator, consider the case when $m = 1$. In this case $Z^{(i+m)} - Z^{(i)}$ just represents the distance between two adjacent samples. The result of Vasicek's estimator is just proportional to the sum of the log of these distances. Thus, if many points are clustered in one area, many of these values will be small resulting in a low entropy. If points are spread out, then many of these values will be large, resulting in a large entropy.

3 The Algorithm

Using these ideas, it is straightforward to construct algorithms for joint bias field removal. As mentioned above, we chose to optimize Equation (8) over the set of band-limited bias fields. To do this, we parameterize the set of bias fields using the sine/cosine basis images shown on the right of Figure 1:

$$B_i = \sum_{j=1}^{25} \alpha_j \phi_j(x,y).$$

We optimize Equation (8) by *simultaneously* updating the bias field estimates (taking a step along the numerical gradient) for each image to reduce the overall entropy. That is, at time step t, the coefficients α_j for each bias field are updated using the latent image estimates and entropy estimates from time step $t-1$. After all α's have been updated, a new set of latent images and pixel-stack entropies are calculated, and another gradient step is taken. Though it is possible to do a full gradient descent to convergence by optimizing one image at a time, the optimization landscape tends to have more local minima for the last few images in the process. The appeal of our joint gradient descent method, on the other hand, is that the ensemble of images provides a natural smoothing of the optimization landscape in the joint process. It is in this sense that our method is "multi-resolution", proceeding from a smooth optimization in the beginning to a sharper one near the end of the process.

We now summarize the algorithm:

1. Initialize the bias field coefficients for each image to 0, with the exception of the coefficient for the DC-offset (the constant bias field component), which is initialized to 1. Initialize the gradient descent step size δ to some value.
2. Compute the summed pixelwise entropies for the set of images with initial "neutral" bias field corrections. (See below for method of computation.)
3. Iterate the following loop until no further changes occur in the images.
   (a) For each image:
      i. Calculate the numerical gradient $\nabla_\alpha H_{\text{Vasicek}}$ of (8) with respect to the bias field coefficients (α_j's) for the current image.
      ii. Set $\alpha = \alpha + \delta \nabla_\alpha \hat{H}_{\text{Vasicek}}$.
   (b) Update δ (reduce its value according to some schedule).

Upon convergence, it is assumed that the entropy has been reduced as much as possible by changing the bias fields, unless one or more of the gradient descents is stuck in a local minimum. Empirically, the likelihood of sticking in local minima is reduced by increasing the number of images (N) in the optimization. In our experiments described below with only 21 real infant brains, the algorithm appears to have found a global minimum of all bias fields to the extent that this can be discerned visually.

Note that for a set of *identical* images, the pixel-stack entropies are not increased by multiplying each image by the same bias field (since all images will still be the same). More generally, when images are approximately equivalent, their pixel-stack entropies are not signficantly affected by a "common" bias field, i.e. one that occurs in all of the

images.[3] This means that the algorithm cannot, in general, eliminate all bias fields from a set of images, but can only *set all of the bias fields to be equivalent.* We refer to any constant bias field remaining in all images after convergence as the *residual bias field*.

Fortunately, there is an effect that tends to minimize the impact of the residual bias field in many test cases. The residual bias field tends to consist of components for each α_j that approximate the mean of that component across images. For example, if half of the observed images have a positive value for a particular component's coefficient, and half have a negative coefficient for that component, the residual bias field will tend to have a coefficient near zero for that component. Hence, the algorithm naturally eliminates bias field effects that are non-systematic, i.e. that are not shared across images.

If the same type of bias field component occurs in a majority of the images, then the algorithm will not remove it, as the component is indistinguishable, under our model, from the underlying anatomy. In such a case, one could resort to within-image methods to further reduce the entropy. However, there is a risk that such methods will remove components that actually represent smooth gradations in the anatomy. This can be seen in the bottom third of Figure 5, and will be discussed in more detail below.

4 Slice-Based Experiments

To test our algorithm, we ran two sets of experiments, the first on images with simulated bias fields, and the second on real brain images. In the first experiment, we started with a single brain image and created a set of "different" brain images by first adding different known bias fields to each image and then randomly translating the images from zero to five pixels in a random direction. The random translation creates an image set in which the pixel stacks have variability similar to a true set of images, but for which the latent images are still known.

If our algorithm works as claimed, then the final recovered images should not necessarily be equal to the original images (since shared bias components cannot be detected) but should recover bias fields that, up to some shared bias field, are equivalent to the originally introduced bias fields. Another way to say this is that the difference $\hat{\alpha} - \alpha$ between the estimated biasfield coefficients $\hat{\alpha}$ and the original bias field coefficients α for each image should be constant across images. If this is true, than the variance of these differences across images should go to zero as the algorithm runs. Figure 4 demonstrates that this is exactly what happens in our experiments. The plot shows that as the algorithm runs, the difference between the estimated bias field coefficients and the true bias field coefficients becomes equal (its variance goes to zero).

More interesting are the results on real images, in which the latent images come from different patients. We obtained 21 pre-registered[4] infant brain images (top of Fig-

[3] Actually, multiplying each image by a bias field of small magnitude can artificially reduce the entropy of a pixel-stack, but this is only the result of the brightness values shrinking towards zero. Such artificial reductions in entropy can be avoided by normalizing a distribution to unit variance between iterations of computing its entropy, as is done in this work.

[4] It is interesting to note that registration is not strictly necessary for this algorithm to work. The proposed MAP method works under very broad conditions, the main condition being that the bias fields do not span the same space as parts of the actual medical images. It is true,

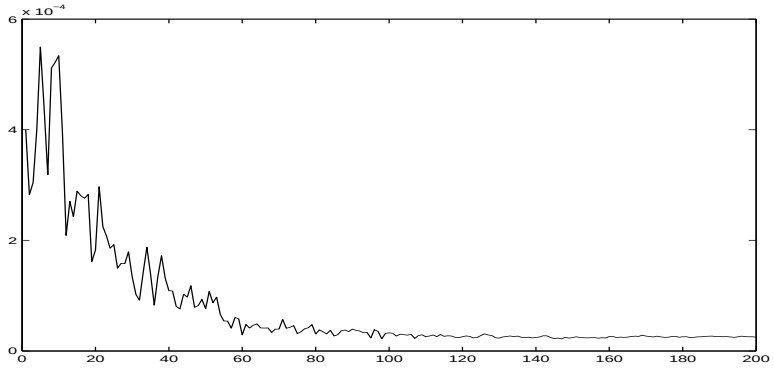

Fig. 4: Typical convergence of the variance of the difference between bias field coefficient estimates and their true values, across images. This convergence implies that the true bias field is recovered up to some "shared" component.

ure 5) from Brigham and Women's Hospital in Boston. Large bias fields can be seen in many of the images. Probably the most striking is a "ramp-like" bias field in the sixth image of the second row. (The top of the brain is too bright, while the bottom is too dark.) Because the brain's white matter is not fully developed in these infant scans, it is difficult to categorize tissues into a fixed number of classes as is typically done for adult brain images; hence, these images are not amenable to methods based on specific tissue models developed for adults (e.g. [9]).

The middle third of Figure 5 shows the results of our algorithm on the infant brain images. (These results must be viewed in color on a good monitor to fully appreciate the results.) While a trained technician can see small imperfections in these images, the results are remarkably good. All major bias artifacts have been removed.

It is interesting to compare these results to a method that reduces the entropy of each image individually, without using constraints between images. Using the results of our algorithm as a starting point, we continued to reduce the entropy of the pixels *within* each image (using a method akin to Viola's [11]), rather than across images. These results are shown in the bottom third of Figure 5. Carefully comparing the central brain regions in the middle section of the figure and the bottom section of the figure, one can see that the butterfly shaped region in the middle of the brain, which represents developing white matter, has been suppressed in the lower images. This is most likely because the entropy of the pixels *within a particular image* can be reduced by increasing the bias field "correction" in the central part of the image. In other words, the algorithm strives to make the image more uniform by removing the bright part in the middle of the image. However, our algorithm, which compares pixels across images, does not suppress these real structures, since they occur across images. Hence coupling across images can produce superior results.

---

however, that as the latent images become less registered or differ in other ways, that a much larger number of images is needed to get good estimates of the pixel-stack distributions.

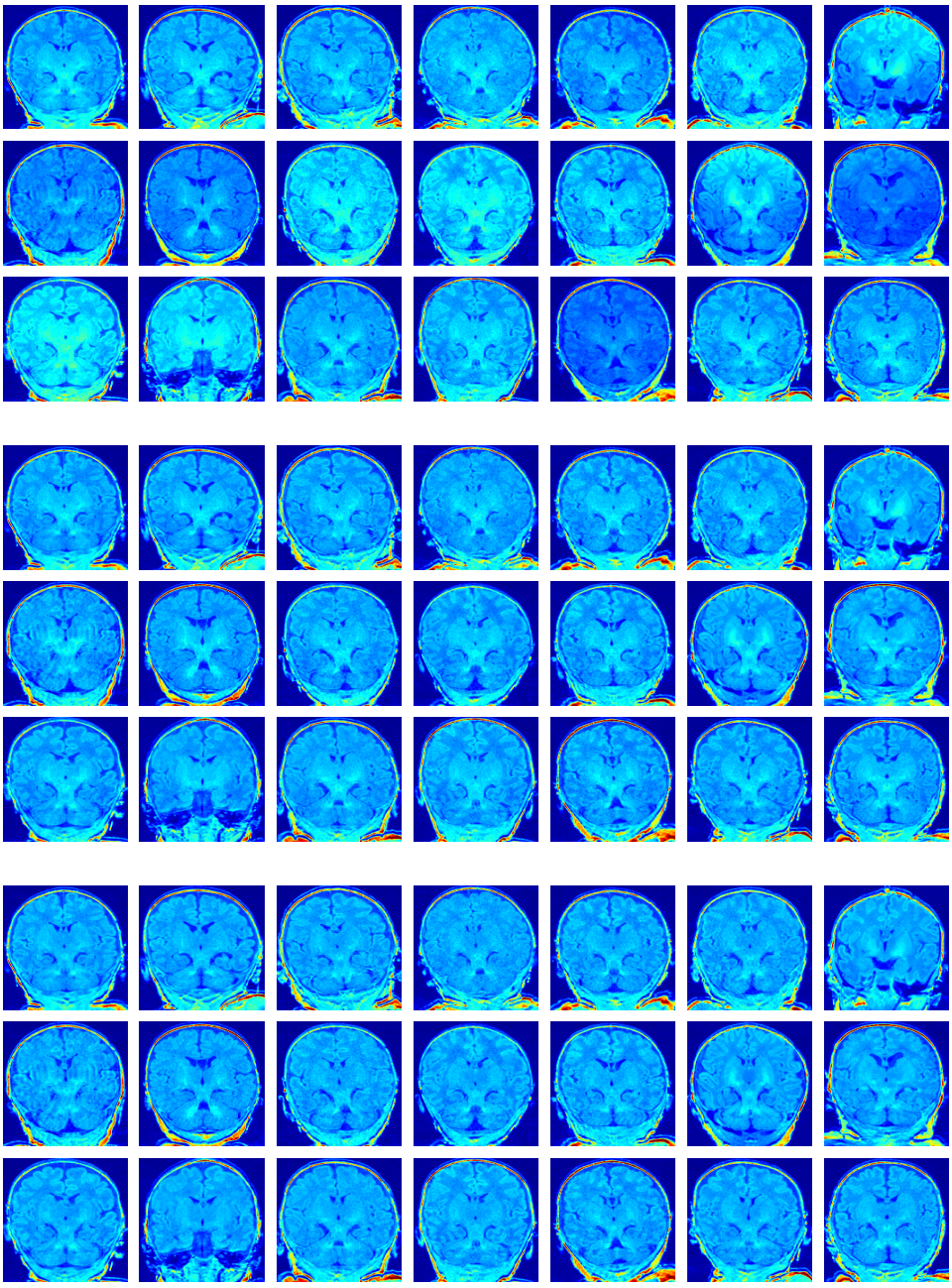

Fig. 5: **NOTE**: This image must be viewed in color (preferably on a bright display) for full effect. **Top.** Original infant brain images. **Middle.** The same images after bias removal with our algorithm. Note that developing white matter (butterfly-like structures in middle brain) is well-preserved. **Bottom.** Bias removal using a single image based algorithm. Notice that white matter structures are repressed.

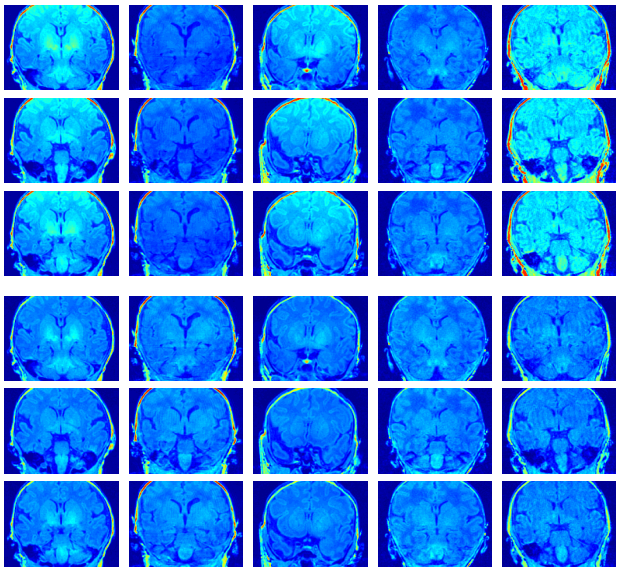

Fig. 6: This figure shows the results of our volumetric joint bias removal algorithm. 15 patient volumes were used, and the bias in each volume was reduced using the 27-component basis volumes for smooth three-dimensional bias fields. The top half of the figure shows 3 images in each column from 5 different patients (rows). The bottom shows the corrected images.

5 Volumetric Bias Removal

Extending this basic method to work with a full series of images from each patient, rather than a single image from each patient, is straightforward and requires only minor modifications to the source code. First, we must parameterize the set of smooth three-dimensional bias fields, which means we need a three-dimensional Fourier basis of volumes. In this work, we used 3-D bases consisting of either 27 or 125 basis volumes, representing bias fields limited, respectively, to either one Hertz or two Hertz in spatial frequency. The 125-volume basis is analogous to the basis shown in Figure 1.

To understand the advantage of correcting bias across volumes rather than across sets of slices one at a time, consider what happens when a set of patient scans are corrected one slice at a time (still grouped across patients of course). In this case, the estimates of bias fields may change sharply from one image to the next within the same patient, ignoring the fact that bias fields tend to be smooth in all three dimensions. This can be avoided by forcing the volumetric bias fields to be parameterized by a smooth three-dimensional basis that enforces smoothness of the bias fields in all directions, and gives us another constraint with which to separate the patients' true anatomical data from smooth bias fields.

In Figure 6, we show the results of our volumetric bias removal algorithm. The bias removal algorithm was done using the 27-volume basis on 15 patients simultaneously.

Results are shown for 3 slices from each of five patients. In future work, we plan to make specific comparisons of volumetric joint bias removal techniques with sequential slice-based joint bias removal to see if the former offers any significant advantage.

The idea of minimizing pixelwise entropies to remove nuisance variables from a set of images is not new. In particular, Miller et al. [6, 7] presented an approach they call *congealing* in which the sum of pixelwise entropies is minimized by *separate affine transforms* applied to each image. Our method can thus be considered an extension of the congealing process to non-spatial transformations. We are currently combining such approaches to do registration and bias removal simultaneously.

This work uses information unused in other methods, i.e. information across images. This suggests an iterative scheme in which both types of information, both within and across images, are used. Local models could be based on weighted neighborhoods of pixels, *pixel cylinders*, rather than single pixel-stacks, in sparse data scenarios. For "easy" bias correction problems, such an approach may be overkill, but for difficult problems in bias correction, where the bias field is difficult to separate from the underlying tissue, as discussed in [1], such an approach could produce critical extra leverage.

We thank Dr. Terrie Inder and Dr. Simon Warfield for graciously providing the infant brain images for this work. The images were obtained under NIH grant P41 RR13218. Also, we thank Neil Weisenfeld and Sandy Wells for helpful discussions. This work was partially supported by Army Research Office grant DAAD 19-02-1-0383.

References

1. Fan, A., Wells, W., Fisher, J., Cetin, M., Haker, S., Mulkern, C., Tempany, C., Willsky, A.: A unified variational approach to denoising and bias correction in MR. IPMI, 2003.
2. Beirlant, J., Dudewicz, E., Gyorfi, L. and van der Meulen, E.: Nonparametric entropy estimation: An overview. *Int. J. of Math. and Stat. Sci., 6*. pp.17-39. 1997.
3. Cover, T. and Thomas J. *Elements of Information Theory*. Wiley, 1991.
4. Learned-Miller, E. G. and Fisher, J.: ICA using spacings estimates of entropy. *Journal of Machine Learning Research*, Volume 4, pp. 1271-1295, 2003.
5. Learned-Miller, E. G. and Ahammad, P.: Joint MRI Bias Removal Using Entropy Minimization Across Images. *Neural Information Processing Systems 17*, pp. 761-768, 2005.
6. Miller, E. G., Matsakis, N., Viola, P. A.: Learning from one example through shared densities on transforms. *IEEE Conference on Computer Vision and Pattern Recognition*. 2000.
7. Miller, E. G.: Learning from one example in machine vision by sharing probability densities. Ph.D. thesis. Massachusetts Institute of Technology. 2002.
8. Vasicek, O.: A test for normality based on sample entropy. *Journal of the Royal Statistical Society Series B, 31*. pp. 632-636, 1976.
9. Wells, W. M., Grimson, W. E. L., Kikinis, R., Jolesz, F.: Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging, 15*. pp. 429-442, 1996.
10. Collins, D.L., Zijdenbos, A.P., Kollokian, J.G., Sled, N.J., Kabani, C.J., Holmes, C.J., Evans, A.C.: Design and Construction of a realistic digital brain phantom. IEEE Transactions on Medical Imaging, 17. pp. 463-468, 1998.
11. Viola, P.A. : Alignment by maximization of mutual information. Ph.D. Thesis. Massachusetts Institute of Technology. 1995.