

# A Unified Variational Approach to Bias Correction and Denoising in MR

Ayres Fan<sup>1</sup>, William M. Wells III<sup>2,3</sup>, John W. Fisher III<sup>1,2</sup>, Mujdat Cetin<sup>1</sup>, Steven Haker<sup>3</sup>, Robert Mulkern<sup>3</sup>, Clare Tempany<sup>3</sup>, Alan Willsky<sup>1</sup>

<sup>1</sup>Laboratory for Information and Decision Systems  
Massachusetts Institute of Technology  
Cambridge, MA, USA

<sup>2</sup>Artificial Intelligence Laboratory  
Massachusetts Institute of Technology  
Cambridge, MA, USA

<sup>3</sup>Brigham and Women's Hospital  
Harvard Medical School  
Boston, MA, USA

## Abstract

We propose a novel bias correction method for magnetic resonance (MR) imaging that uses complementary body coil and surface coil images. The former are spatially homogeneous but have low signal intensity; the latter provide excellent signal response but have large bias fields. We present a variational framework where we optimize an energy functional to estimate the bias field and the underlying image using both observed images. The energy functional contains smoothness-enforcing regularization for both the image and the bias field. We present extensions of our basic framework to a variety of imaging protocols. We solve the optimization problem using a computationally efficient numerical algorithm based on coordinate descent, preconditioned conjugate gradient, half-quadratic regularization, and multigrid techniques. We show qualitative and quantitative results demonstrating the effectiveness of the proposed method in producing debiased and denoised MR images.

## Problem Statement

- The bias field is a systematic intensity inhomogeneity that corrupts magnetic resonance (MR) images.
  - Fundamental trade-off between SNR and spatially-homogeneous signal response.
  - Uncorrupted image  $\varphi(x)$  would depend solely on the underlying tissue (e.g.,  $T_1$ ,  $T_2$ ,  $\rho$ ) and the imaging parameters (e.g.,  $T_E$ ,  $T_R$ ).
  - The resulting MR signal at the receiver is then the true MR signal  $\varphi(x)$  multiplied by the reception profile of the receiving coil  $\beta(x)$  plus additive thermal noise:
 
$$I(x) = \beta(x)\varphi(x) + n(x)$$
  - Receiving with a body coil results in low signal-to-noise ratio (SNR) but good spatial homogeneity.
  - Surface coils have a signal response that rapidly diminishes with distance. By placing the receiving coil close to the object being imaged, this variable response allows better visualization of the region of interest (ROI) but also causes the bias field.
- Removing the bias field makes both human and computer analysis easier.
- Separating the bias field from the true underlying image is an underconstrained and ill-posed problem—there are half the number of observations as there are free variables.

## Previous Work

- Most previous methods make the following general assumptions:
  - The bias field is slowly varying in space.
  - The bias field is tissue independent.
  - Tissue intensities are piecewise constant.
- Earliest work in mid-80's: Haselgrove and Prammer (1986), Lufkin et al (1986) with homomorphic unsharp filtering, Axel et al (1987) with phantom correction.
- Window/level adequate for human vision tasks, but not for computer-based processing.
- More modern approaches include Dawant et al (1993) with spline fitting, Singh and Nesaiver (1993) with embedded coil markers, Likar et al (2000) with entropy minimization.
- Various simultaneous bias correction and segmentation approaches such as Meyer et al (1995) using adaptive clustering and Wells et al (1996) using Expectation-Maximization.

## Imaging Model

- Brey and Narayana (1988) proposed capturing images from both the body coil and the surface coil. The measurement model is then:
 
$$I_B(x) = \varphi(x) + n_B(x)$$

$$I_S(x) = \beta(x)\varphi(x) + n_S(x)$$
- $I_B$  is homogeneous but noisy.  $I_S$  has high SNR in the region of interest, but a potentially severe bias artifact.
  - The noise in the acquisition process is thermal which results in complex independent and identically-distributed (IID) Gaussian noise.
  - Working with the absolute value images results in Rician noise.
  - The gain in SNR from using a surface coil does not come from reduction of noise, but from increased signal gain from the bias field. Thus the more severe the bias field, the better the local signal strength.
- Other similar approaches:
  - Brey-Narayana filter  $I_B$  and  $I_S$  and divide the results to estimate the bias field.
  - Lai and Fang (1998) fit splines to the bias field.
  - Prussmann et al (2001) fit local polynomials to the bias field.

## Variational Formulation

- We model the noise as Gaussian and IID. This is a good approximation in medium-to-high SNR regions.
- We stack the 2D or 3D images into vectors. We now want to estimate the vector quantities  $\mathbf{f}$  and  $\mathbf{b}$ . The discrete measurement model is:
 
$$\mathbf{y}_B = \mathbf{f} + \mathbf{n}_B$$

$$\mathbf{y}_S = \mathbf{b} * \mathbf{f} + \mathbf{n}_S$$
- We can then write the log likelihood as (ignoring constant terms):
 
$$\ell(\mathbf{y}_B, \mathbf{y}_S; \mathbf{b}, \mathbf{f}) = -\frac{1}{2\sigma_B^2} \|\mathbf{y}_B - \mathbf{f}\|^2 - \frac{1}{2\sigma_S^2} \|\mathbf{y}_S - \mathbf{b} \circ \mathbf{f}\|^2$$
- Finding the ML estimate results in trivial estimates, so we construct an augmented energy function that encourages smoothness in  $\mathbf{b}$  and piecewise smoothness in  $\mathbf{f}$ :
 
$$E(\mathbf{b}, \mathbf{f}) = \lambda_B \|\mathbf{y}_B - \mathbf{f}\|^2 + \lambda_S \|\mathbf{y}_S - \mathbf{b} \circ \mathbf{f}\|^2 + \alpha \|D\mathbf{b}\|^2 + \gamma \|\mathbf{L}\mathbf{f}\|_p^p$$
  - $D$  and  $L$  are matrices chosen to implement differential operators. Generally choose  $p < 1$  to help preserve edges.
  - $\lambda_B$ ,  $\lambda_S$ ,  $\alpha$ , and  $\gamma$  are all positive constants, and the  $\lambda$ 's are related to the inverse noise variances of the observed images.
- Easy to generalize formulation for multiple surface coils/multiple imaging parameters.
  - One error term for each observed image, one smoothing term for each coil, and one regularizing term for each true image.

## Minimizing the Energy Function

- The overall problem is non-convex, so we use coordinate descent to alternately optimize  $\mathbf{b}$  and  $\mathbf{f}$ .
  - Minimizing the energy simultaneously with respect to  $\mathbf{b}$  and  $\mathbf{f}$  is difficult. But given  $\mathbf{b}$ ,  $\mathbf{f}$  is relatively easy to obtain, and vice versa.
  - A stationary point found using coordinate descent is also a stationary point of the overall energy functional.

### b-step

- With  $\mathbf{f}$  fixed, the energy is quadratic and strictly convex in terms of  $\mathbf{b}$ .
- Setting the gradient to zero leads to a linear equation for the solution:

$$(\lambda_S \mathbf{F}^2 + \alpha D^T D) \mathbf{b} = \mathbf{F} \mathbf{y}_S$$

- Direct matrix inversion possible, but slow.
- We use preconditioned conjugate gradient to find a sub-optimal iterative solution.
- Conjugate gradient exhibits a superlinear convergence rate for convex quadratic optimization problems.
- We find that using a tridiagonal preconditioner constructed from the three main diagonals of the matrix increases convergence speed by 2x-3x, while increasing computation time by 10-15% per iteration.

### f-step

- With no regularization on  $\mathbf{f}$ , we can minimize the function pointwise:  $\hat{f}^{[n]} = \frac{\lambda_B \mathbf{y}_B[n] + \lambda_S \mathbf{b}^{[n]} \mathbf{y}_S[n]}{\lambda_B + \lambda_S \mathbf{b}^{[n]}}$ 
  - This equation can be interpreted as a noise-weighted convex combination between  $\mathbf{y}_B$  and  $\mathbf{y}_S/\mathbf{b}$ .
  - When  $\mathbf{b}^{[n]}$  is large, we mainly use the surface coil for the reconstruction. When  $\mathbf{b}^{[n]}$  is small, we mainly use the body coil.
  - In regions far from the surface coil, there will be a significant advantage to incorporating the body coil measurements.
  - This is essentially the same fusion equation found by Roemer et al (1990).
- For the general minimization problem with  $l_p$  regularization, we use half-quadratic optimization (Geman-Reynolds 1992).
  - Fixed-point iterative method where we form a succession of quadratic approximations to the  $l_p$  norm at  $\mathbf{f}^{(0)}$ ,  $\mathbf{f}^{(1)}$ , ...:
 
$$\|\mathbf{L}\mathbf{f}\|_p^p \approx \mathbf{f}^T \mathbf{L}^T \mathbf{W}^{(i)} \mathbf{L} \mathbf{f}$$
  - Choose the diagonal matrix  $\mathbf{W}^{(0)}$  so that the relationship holds for  $\mathbf{f}^{(0)}$ :
 
$$\mathbf{W}^{(i)}[n, n] = \frac{p}{2} \left( ((D\mathbf{f}^{(i-1)})[n])^2 + \xi \right)^{p/2-1}$$
  - This makes each f-step quadratic, so we just need to solve this linear equation at each half-quadratic iteration:
 
$$(\lambda_B \mathbf{I} + \lambda_S \mathbf{B}^2 + \gamma D^T \mathbf{W}^{(i)} D) \hat{\mathbf{f}}^{(i)} = \lambda_B \mathbf{y}_B + \lambda_S \mathbf{B} \mathbf{y}_S$$
  - This is again a positive definite system, so we can use PCG to solve. This results in three sets of nested iterations.
- We use multigrid, a multiresolution technique that can help avoid local minima and increase convergence speed.
  - Simple coarse-to-fine implementation finds the low frequency components on coarser grids and propagates the results to the original scale.
- For the generalized case (multiple surface coils/multiple pulse sequences), we can use coordinate descent again, but on each b-step we need to compute an estimate for every  $\mathbf{b}$ , and similarly on each f-step.
- 3D implementation is a straightforward application of the exact same energy functional (structure of  $D$  and  $L$  changes).
  - Slower than independent processing of slices, but enables better coupling across slices than independent processing allows.
  - Can also interpolate one bias field estimate to correct volumes obtained with other slice orientations (e.g., axial to correct sagittal).

### Overview of algorithm

- Downsample to finest desired grid and find estimates at that scale.
- Solve problem at scale  $s$  (using input from scale  $s+1$ )
  - b-step: minimize  $\lambda_S \|\mathbf{y}_S - \mathbf{F}\mathbf{b}\|^2 + \alpha \|D\mathbf{b}\|^2$  ( $\mathbf{F}$  is a diagonal matrix with  $\mathbf{f}$  along the diagonal)
    - PCG iterations
  - f-step: minimize  $\lambda_B \|\mathbf{y}_B - \mathbf{f}\|^2 + \lambda_S \|\mathbf{y}_S - \mathbf{B}\mathbf{f}\|^2 + \gamma \|\mathbf{L}\mathbf{f}\|_p^p$  ( $\mathbf{B}$  is a diagonal matrix with  $\mathbf{b}$  along the diagonal)
    - Half-quadratic iterations
      - PCG iterations
    - Repeat A and B until convergence.
- Repeat II until we reach scale 0 (the original scale).

## Results

### MNI Brain Phantom

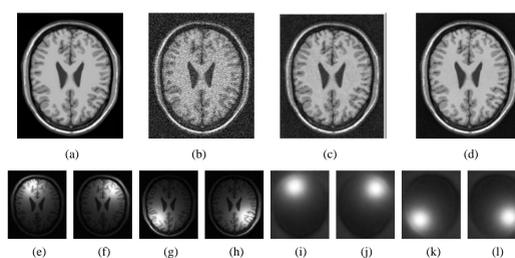


Fig. 1: (a) Ground truth. (b) Body coil image. (c) Surface coil images. (d) Estimated true image ( $\gamma=0$ ). (e-h) Estimated true image ( $\gamma=0.014$ ). (i-l) Estimated bias fields.

### Heart Data

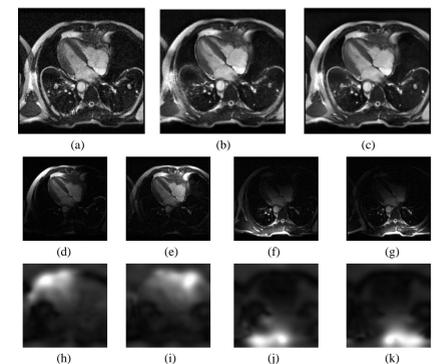


Fig. 3: (a) Body coil image. (b) True image estimate ( $\gamma=0$ ). (c) True image estimate ( $\gamma=1800$ ). (d-g) Surface coil images. (h-k) Estimated bias fields.

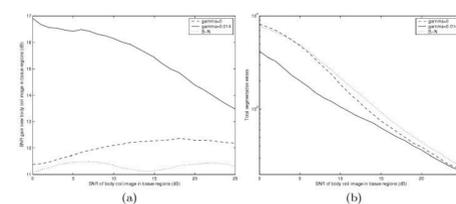


Fig. 2: (a) SNR gain from fusing images for various input noise levels. (b) Thresholding segmentation performance with varying SNR.

### Prostate Example

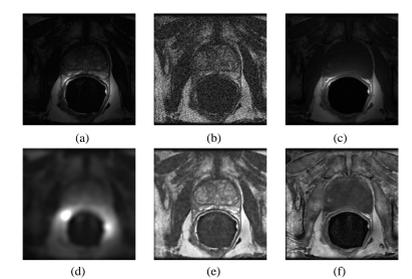


Fig. 5: (a) Surface coil T2-weighted image. (b) Body coil T2-weighted image. (c) Surface coil T1-weighted image. (d) Estimated bias field. (e) Estimated intrinsic T2-weighted image. (f) Estimated true T1-weighted image.

### Brain Example

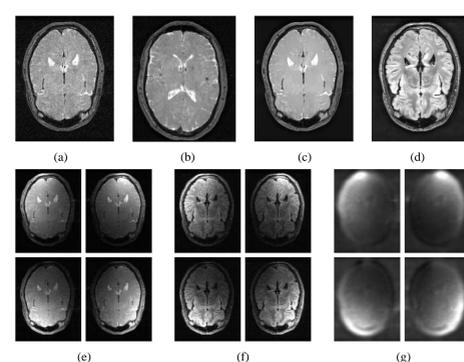


Fig. 4: (a) Body coil GRE image. (b) Estimated true GRE image ( $\gamma=0$ ). (c) Estimated true GRE image ( $\gamma=1000$ ). (d) Estimated true FLAIR image. (e) Surface coil GRE images. (f) Surface coil FLAIR images. (g) Estimated bias fields.

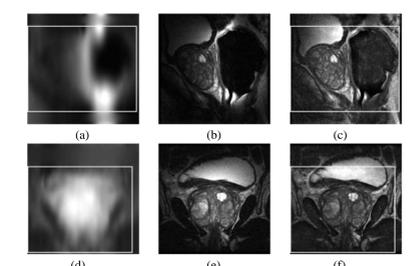


Fig. 6: White box indicates overlap with axial volume. Top row contains sagittal images, bottom row contains coronal images. (a,d) Interpolated bias fields. (b,e) Surface coil images. (c,f) Corrected images.

## Conclusion

- Results are visually appealing, but not dramatically superior to older methods such as filtering and dividing.
- Good convergence properties.
  - No problems with converging to poor local minima, even when initializing with random noise.
  - Guaranteed convergence to saddle point or local minimum, and every b- and f-step decreases the energy.
- Contributions
  - Non-parametric variational formulation of image fusion problem with statistical estimation flavor.
  - Demonstrably superior results on synthetic examples.
  - Simultaneous bias correction and denoising.
  - Seamless handling of multiple surface coils and multiple pulse sequences.
  - Efficient solver using coordinate descent, preconditioned CG, multigrid.