Nonparametric Neighborhood Statistics for MRI Denoising

Suyash P. Awate and Ross T. Whitaker

UUCS-05-007

School of Computing
University of Utah
Salt Lake City, UT 84112, USA

Abstract

This paper presents a novel method for denoising MR images that relies on an optimal estimation, combining a likelihood model with an adaptive image prior. The method models images as random fields and exploits the properties of independent Rician noise to learn the higher-order statistics of image neighborhoods from corrupted input data. It uses these statistics as priors within a Bayesian denoising framework. This paper presents an information-theoretic method for characterizing neighborhood structure using nonparametric density estimation. The formulation generalizes easily to simultaneous denoising of multimodal MRI, exploiting the relationships between modalities to further enhance performance. The method, relying on the information content of input data for noise estimation and setting important parameters, does not require significant parameter tuning. Qualitative and quantitative results on real, simulated, and multimodal data, including comparisons with other approaches, demonstrate the effectiveness of the method.

Nonparametric Neighborhood Statistics for MRI Denoising

Suyash P. Awate and Ross T. Whitaker

School of Computing, University of Utah, Salt Lake City, UT 84112, USA
{suyash,whitaker}@cs.utah.edu

Abstract. This paper presents a novel method for denoising MR images that relies on an optimal estimation, combining a likelihood model with an adaptive image prior. The method models images as random fields and exploits the properties of independent Rician noise to learn the higher-order statistics of image neighborhoods from corrupted input data. It uses these statistics as priors within a Bayesian denoising framework. This paper presents an information-theoretic method for characterizing neighborhood structure using nonparametric density estimation. The formulation generalizes easily to simultaneous denoising of multimodal MRI, exploiting the relationships between modalities to further enhance performance. The method, relying on the information content of input data for noise estimation and setting important parameters, does not require significant parameter tuning. Qualitative and quantitative results on real, simulated, and multimodal data, including comparisons with other approaches, demonstrate the effectiveness of the method.

1 Introduction

Over the last several decades, magnetic resonance (MR) imaging technology has benefited from a variety of technological developments resulting in increased resolution, signal to noise ratio (SNR), and acquisition speed. However, fundamental trade-offs between resolution, speed, and SNR combined with scientific, clinical, and financial pressures to obtain more data more quickly, result in images that still exhibit significant levels of noise. In particular, the need for shorter acquisition times, such as in dynamic imaging, often undermines the ability to obtain images having both high resolution and high SNR. Furthermore, the efficacy of higher-level, post processing of MR images, including tissue classification and organ segmentation, that assume specific models of tissue intensity (e.g. homogeneous), are sometimes impaired by even moderate noise levels. Hence, denoising MR images remains an important problem. From a multitude of statistical and variational denoising formulations proposed, no particular one appears as a clear winner in all relevant aspects, including the reduction of randomness and intensity bias, structure and edge preservation, generality, reliability, automation, and computational cost. The paper proposes a method for denoising MR magnitude data modeling images as random fields, but unlike statistical methods in literature, *it does not rely on a specific, ad-hoc image prior*. Instead, it estimates

the higher-order signal statistics from the neighborhood statistics of the noisy input data by deconvolving the latter with the noise statistics. It then uses these statistics as priors within an optimal Bayesian denoising framework.

2 Related Work

A multitude of variational/nonlinear PDE-based methods have been developed for a wide variety of images and applications [15, 14], with some of these having applications to magnetic resonance imaging (MRI) [8, 11, 7]. However, such methods impose certain kinds of *models* on local image structure, and these models are often too simple to capture the complexity of anatomical MR images. Also they do not take into account the bias introduced by Rician noise. Furthermore, they usually involve manual tuning of *critical* free parameters that control the conditions under which the models prefer one sort of structure over another; this has been an impediment to the widespread adoption of these techniques.

The wavelet literature addresses image denoising extensively [16]. Healy *et al.* [9] were among the first to apply soft-thresholding based wavelet techniques for denoising MR images. Hilton *et al.* [10] applied a threshold-based scheme for functional MRI data. Nowak [13], operating on the square magnitude MR image, includes a Rician noise model in the threshold-based wavelet denoising scheme and thereby corrects for the bias introduced by the noise.

Several statistically based image processing algorithms rely on information theory such as the *mean-shift* algorithm [3]. It is a *mode seeking* process that operates only on image intensities (scalar/vector valued) and does not account for the neighborhood structure. As such it has been used for image segmentation, but not for reconstruction. Some MR nonuniformity correction methods are based on the quantification of information content in MR images [19, 12]. They follow from the observation that nonuniformities increase the entropy of the 1D *gray scale* probability density functions (PDFs). However, entropy measures on first-order image statistics are insufficient for denoising; thus this paper extends the information theoretic strategy to higher-order PDFs.

Another class of statistical methods are based on Markov random fields [24, 22]. The proposed method also exploits the Markov property of the images, but rather than imposing an ad-hoc image model, it *estimates* the relevant conditional PDFs from the input data. We show that incorporating spatial information, via neighborhood statistics, is effective for MRI denoising *and* that the process can be bootstrapped from the image data, making a very general algorithm with less tuning of critical free parameters.

Previous work in estimation theory has addressed the use of optimal image estimation using neighborhood probabilities [21]. That work focuses on *discrete* functions and relies on inverting the channel transition matrix (noise model) to give a closed form estimate for source statistics. The proposed method addresses continuous-valued signals, which is essential for medical imaging applications, and thus entails deconvolving nonparametric approximations to PDFs via entropy reduction. It also addresses the effect of noise in the neighborhoods that

are used to condition the estimate, hence making it more effective for reducing additive/multiplicative noise, which is important in medical image processing.

The method in this paper builds on our previous work in [1]. That work lays down the foundations for unsupervised learning of higher-order image statistics and proposes entropy reduction as a denoising heuristic for independent additive zero-mean Gaussian noise for single gray scale images. This paper uses entropy reduction coupled with the Rician noise model as a means to recover higher-order image statistics from noisy input data. It exploits such statistics for optimal Bayesian denoising of MR images, with a method for computing the expectation of the posterior. It also addresses the question of how to utilize multimodal data within this optimal framework.

3 Neighborhood Statistics for MRI Denoising

This section begins with an overview of the random-field image model and then describes the formulation that uses *a priori* information of higher-order (neighborhood) statistics within an optimal Bayesian estimation framework. The next section (Section 4) describes a way of bootstrapping this process by generating such priors from the noisy data itself.

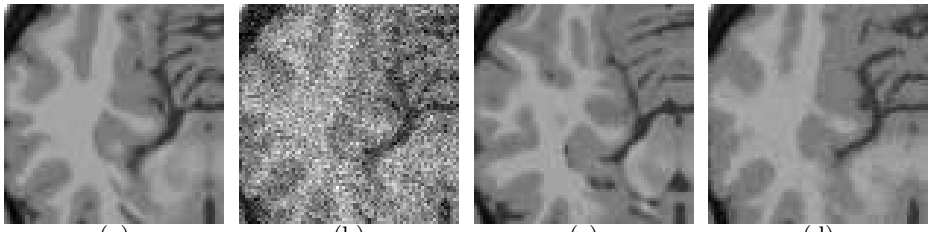
3.1 Random Field Image Model

A random field/process [5] is a family of random variables $X(\Omega;T)$, for some index set T, where, for each fixed $T=t$, the random variable $X(\Omega;t)$ is defined on the sample space Ω. If we let T be a set of points defined on a discrete Cartesian grid and fix $\Omega=\omega$, we have a realization of the random field called the *digital image*, $X(\omega,T)$. In this case $\{t\}_{t\in T}$ is the set of pixels in the image. For 2-dimensional images t is a two-vector. We use a shorthand to denote random variables $X(\Omega;t)$ by $X(t)$. We denote a specific realization $X(\omega;t)$ (the digital image), as a deterministic function $x(t)$.

If we associate with T a family of pixel neighborhoods $N=\{N_t\}_{t\in T}$ such that $N_t\subset T$, $t\notin N_t$, and $u\in N_t$ if and only if $t\in N_u$, then N is called a neighborhood system for the set T and points in N_t are called neighbors of t. We define a random vector $Y(t)=\{X(t)\}_{t\in N_t}$, denoting its realization by $y(t)$, corresponding to the set of intensities at the neighbors of pixel t. We denote the noiseless image by $X(\omega,T)$ and its associated set of neighborhood intensities by $Y(\omega,T)$. Correspondingly, for the observed noisy image, we use $\tilde{X}(\omega,T)$ and $\tilde{Y}(\omega,T)$. For the formulation in this paper, we assume the noiseless image to be generated from a stationary ergodic process (in practice this assumption can be relaxed, somewhat). For notational simplicity, we use the short hand for random variables $X(t)$ as X and their realizations $x(t)$ as x, dropping the index t.

3.2 Bayesian Estimation with Higher-Order Statistical Priors

The proposed strategy relies on several pieces of technology that interact to provide accurate, practical models of image statistics. For clarity the discussion



(a) (b) (c) (d)

Fig. 1. Insets of (a) the noiseless image, (b) the noisy image (SNR 12db), (c) one of the two images forming the higher-order *prior*, and (d) the denoised image (SNR 23db).

begins at a high level allowing for certain available models and estimates; successive sections discuss how each of these pieces is developed from the input data. Our goal is to estimate the true intensity x from the observed noisy intensity $\tilde{x}$ by exploiting the neighborhood intensities. We begin with the simplest case where we know the *uncorrupted* neighborhood intensities y. We consider Bayesian estimation with the prior $P(X|Y = y)$ and the likelihood $P(\tilde{X} = \tilde{x}|X)$. Assuming again, for simplicity, that we know the prior, Bayes rule gives the posterior as

$$P(X|\tilde{X} = \tilde{x}, Y = y) = \frac{1}{\eta} P(\tilde{X} = \tilde{x}|X) P(X|Y = y) \tag{1}$$

where $\eta = P(\tilde{X} = \tilde{x}|Y = y)$ is a normalization factor. For a squared error loss function the optimal estimate is the posterior mean $\hat{x} = E[X|\tilde{X} = \tilde{x}, Y = y]$.

In practice, two problems undermine this strategy. The first concerns obtaining the conditional PDFs that give the priors for an image. We propose to model these nonparametrically using Parzen windowing with samples of image neighborhoods, as described in subsequent sections. These samples can come from either a suitable database of high SNR images (e.g. different images of the same modality and anatomy) or from the noisy input image itself, using a bootstrapping process described in Section 4. The second problem is that, even if we know the priors, we know only $\tilde{y}$ for the input data (not y). To address this issue, we start with $\tilde{y}$ as an approximation for y and iterate on the posterior estimates to a fixed point where the posterior estimate for each pixel is consistent with the prior given by the estimates of its neighbors. Thus, as the iterations proceed, the noise in the pixel intensities reduces and the neighborhoods give progressively better estimates of the prior. The proposed algorithm is therefore:

1. The input image I comprises a set of intensities $\{\tilde{x}\}_{t \in T}$ and neighborhoods $\{\tilde{y}\}_{t \in T}$. These values form the initial values ($I^0 = I$) of a sequence of images $I^0, I^1, I^2, \ldots$, with corresponding intensities $\hat{x}^0, \hat{x}^1, \hat{x}^2, \ldots$ and neighborhoods $\hat{y}^0, \hat{y}^1, \hat{y}^2, \ldots$.
2. Compute the likelihood PDF $P(\tilde{X} = \tilde{x}|X)$, as described in Section 3.4.
3. For each pixel in the current image I^m, estimate the higher-order prior $P(X|Y = \hat{y}^m)$, as described in Section 3.3.
4. Construct a new image I^{m+1} with intensities $\hat{x}^{m+1}$ as the posterior mean $\hat{x}^{m+1} = E[X|\tilde{X} = \tilde{x}, Y = \hat{y}^m]$.

5. If $\| I^{m+1} - I^m \| > \delta$ (small threshold), go to Step 3, otherwise I^{m+1} is the output.

Figure 1 shows a demonstration of this concept on simulated MRI data from the BrainWeb [2] project. We corrupt a T1 image with Rician noise and use two other similar, but not identical, images as priors. We use 9×9 neighborhoods. Figure 1(c) is one of the two images representing the nonparametric prior model (Parzen windows, 500 local random samples for each t), and Figure 1(d) is the output image. This example shows the power of the prior—the denoised image exhibits structures that are barely visible in the noisy version. The coming sections describe the underlying technology in this estimation process, and give an algorithm for generating data-driven prior models *without an example*.

3.3 Modeling the Prior: Nonparametric Density Estimation

Bayesian estimation using higher-order statistics entails the estimation of higher-order conditional PDFs. Despite theoretical arguments suggesting that density estimation beyond a few dimensions is impractical, the empirical evidence from the statistics literature is more optimistic [17, 1]. The results in this paper confirm that observation. Moreover, stationarity implies that the random vector (X, Y) exhibits identical marginal PDFs, leading to more accurate density estimates [17]. In addition, the neighborhoods in natural images have a lower-dimensional topology in the high-dimensional feature space [4] that aids in density estimation.

We use the Parzen-window nonparametric density estimation technique [6] with an n-dimensional Gaussian kernel $G_n(z, \Psi_n)$, where n is the neighborhood size. Having no a priori information on the structure of the PDFs, we choose an isotropic Gaussian, i.e. $\Psi_n = \sigma_P I_n$, where I_n is the $n \times n$ identity matrix. Using optimal values of the Parzen-window parameters is critical for success, and that can be difficult in such high-dimensional spaces; we have developed a method for automatically choosing this parameter, as described Section 4.3.

For a stationary ergodic process, the estimated prior is

$$P(X|\tilde{Y} = \tilde{y}_i) = \frac{\sum_{t_j \in A_i} G_n(\tilde{y}_i - y_j, \Psi_n) G_1(x_j, \Psi_1)}{\sum_{t_j \in A_i} G_n(\tilde{y}_i - y_j, \Psi_n)} \tag{2}$$

where the set A_i is a small subset of T, chosen at random for each t_i, and x_j and y_j are shorthand for $x(t_j)$ and $y(t_j)$ respectively. This results in a stochastic approximation for the conditional PDFs and the corresponding posteriors.

3.4 Approximating the Rician Likelihood

The Rician PDF of the MRI intensities does not lend itself to analytical, closed-form representations of quantities, such as the likelihood and the posterior expectation, which we need for each iteration of this algorithm. In practice we have found that the shape of the PDF is less important than having good estimates of variance and bias. Therefore, we develop a method of approximating Rician

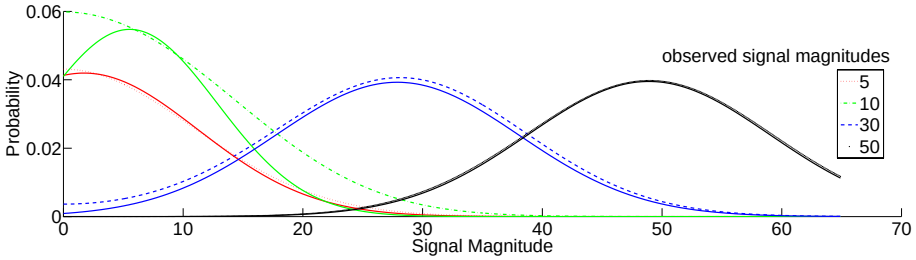


Fig. 2. Gaussians (solid lines) approximating likelihood functions (non-solid lines) for different observed signal magnitudes (underlying noise $\equiv N(0, 100)$).

noise (via the likelihood) by additive Gaussian noise with a signal-dependent mean and variance. For the underlying independent noise $N(0, \sigma^2)$, and σ estimated using the method described by Nowak [13], the likelihood is

$$P\left(\tilde{X} = \tilde{x} | X = x\right) = \frac{\tilde{x}}{\sigma^2} \exp(-\frac{\tilde{x}^2 + x^2}{2\sigma^2}) I_0(\frac{\tilde{x}x}{\sigma^2}) \tag{3}$$

where $I_0(\cdot)$ is the zero-order modified Bessel function of the first kind. For a discrete set of observed signal magnitudes $\tilde{x}$, we fit a Gaussian to the likelihoods via a Levenberg-Marquardt optimization scheme. In this way, we create (in a preprocessing step) a lookup table mapping $\tilde{x}$ to the parameters of the Gaussian approximation, and interpolate the parameters between sample points as needed in subsequent likelihood calculations. At high SNR the means are close to $\tilde{x}$ while at low SNR the means are substantially lower. Figure 2 shows the likelihood PDFs and the approximated Gaussians for various observed signal magnitudes.

3.5 Computing the Posterior Mean

Equations 1 and 2, and the Gaussian approximated likelihood, give the posterior

$$P(X|\tilde{X} = \tilde{x}, \tilde{Y} = \tilde{y}_i) = \frac{1}{\eta} \frac{\sum_{t_j \in A_i} G_n(\tilde{y}_i - y_j, \Psi_n) G_1(x_j, \sigma_P^2)}{\sum_{t_j \in A_i} G_n(\tilde{y}_i - y_j, \Psi_n)} G_1(\tilde{x}_L, \tilde{\sigma}_L^2), \tag{4}$$

where $\tilde{\sigma}_P^2$ is the Parzen-window kernel variance, and $\tilde{x}_L$ and $\tilde{\sigma}_L^2$ are the mean and variance of the Gaussian approximation to the likelihood (from the lookup table). The posterior mean is given by a sum of expectations of Gaussian products:

$$E[X|\tilde{X} = \tilde{x}, \tilde{Y} = \tilde{y}_i] = \frac{\sum_{t_j \in A_i} G_n(\tilde{y}_i - y_j, \Psi_n) K_{ij} M_{ij}}{\sum_{t_j \in A_i} G_n(\tilde{y}_i - y_j, \Psi_n) K_{ij}}; \tag{5}$$

$$K_{ij} = \frac{\exp(-A_{ij}(C_{ij} - B_{ij}^2/4))}{\sqrt{2\pi(\tilde{\sigma}_P^2 + \tilde{\sigma}_L^2)}}; M_{ij} = \frac{B_{ij}}{2};$$

$$A_{ij} = \frac{\tilde{\sigma}_P^2 + \tilde{\sigma}_L^2}{2\tilde{\sigma}_P^2 \tilde{\sigma}_L^2}; B_{ij} = 2\frac{x_j \tilde{\sigma}_P^2 + \tilde{x}_L \sigma_L^2}{\tilde{\sigma}_P^2 + \tilde{\sigma}_L^2}; C_{ij} = \frac{x_j^2 \tilde{\sigma}_P^2 + \tilde{x}_L^2 \sigma_L^2}{\tilde{\sigma}_P^2 + \tilde{\sigma}_L^2};$$

where we exploit the property that the Gaussian is its own *conjugate*.

4 Bootstrapping Neighborhood Statistics from Noisy Input Data

So far we discussed denoising with higher-order statistical priors. In the absence of noiseless/high-SNR example images, we must estimate these from the noisy input image. If we wish to construct an approximation to the prior (neighborhood statistics) from the input data, we must address the affects of noise on this PDF. We approximate Rician noise as (nonstationary) additive Gaussian. Hence the proposed method derives from the effects of additive Gaussian noise on PDFs. Additive noise in the signal corresponds to a *convolution* of the PDFs of the signal and noise. Therefore, for probability densities, noise reduction corresponds to deconvolving the PDF of the input data by the PDF of the noise.

4.1 Estimating Neighborhood Statistics

Rician noise affects the conditional PDFs in two ways: (a) it introduces a bias (shift), and (b) it increases its entropy $h(\tilde{X}|\tilde{Y}=\tilde{y})$ [18]. Hence, we propose entropy reduction coupled with bias correction in an attempt to recover the PDFs. Of course, entropy reduction might also partly eliminate the normal variability in the image. However, we are motivated by the observation that noiseless images tend to have very low entropies relative to their noisy versions. Thus, entropy reduction first affects the noise substantially more than the image statistics. We propose bias correction by shifting intensities $\tilde{x}$ towards their likelihood mean $E[\tilde{X}=\tilde{x}|X]$. For the case of zero noise these two values coincide, thereby eliminating the need for any correction. Otherwise, we move $\tilde{x}$ towards its likelihood mean with a force proportional to the difference. Thus, to restore the conditional PDFs of the input, we minimize the functional

$$\sum_{t\in T}\left[\lambda_1\left(h(\tilde{X}|\tilde{Y}=\tilde{y})\right)+\lambda_2\left(\tilde{x}-E[\tilde{X}=\tilde{x}|X]\right)^2/2\right]. \tag{6}$$

The first term in the functional sharpens the conditional PDFs, and the second term aids in bias correction. We use an iterative gradient-descent optimization scheme with finite forward differences. The PDF restoration proceeds as follows:

1. The input image I comprises a set of intensities $\{\tilde{x}\}_{t\in T}$. These values form the initial values of a sequence of images $I^0, I^1, I^2, \ldots$.
2. Using the current image I^m, construct a new image I^{m+1} with intensities $\tilde{x}^{m+1}=\tilde{x}^m-\lambda_1\partial h/\partial\tilde{x}^m-\lambda_2\left(\tilde{x}^m-E[\tilde{X}=\tilde{x}^m|X]\right)$.
3. If the estimated noise level (as per the method in [13]) in I^{m+1} is zero, then stop. Otherwise, go to Step 2.

We call the final image generated by this process as the *PDF-restored* image. This image forms the *example* image, from which samples are taken to model the prior conditional probabilities in Equation 2. In practice, the results are somewhat insensitive to the values of λ_1 and λ_2, and we choose λ_1, as described in Section 5, related to a mean-shift update.

4.2 Entropy Minimization via Gradient Descent

Entropy is the expectation of negative log-probability, and therefore we can approximate it with the sample mean [20]. For a stationary ergodic process, we approximate the entropy of the conditional PDF as

$$h(\tilde{X}|\tilde{Y}=\tilde{y}_i) \approx -\frac{1}{|T|}\sum_{t_i \in T} \log\left[\frac{\sum_{t_j \in A_i} G_{n+1}(\tilde{w}_i - \tilde{w}_j, \Psi_{n+1})}{|A_i| P(\tilde{Y}=\tilde{y}_i)}\right] \tag{7}$$

where $\tilde{w}_i = (\tilde{x}_i, \tilde{y}_i)$, A_i is a small subset of T, chosen at random; as done in Section 3.3 for computing the prior. A variety of practical issues associated with this strategy, are discussed in Section 4.3. The gradient descent for w_i is

$$\frac{\partial \tilde{x}_i}{\partial t} = -\frac{1}{|T|}\frac{\partial \tilde{w}_i}{\partial \tilde{x}_i}\sum_{t_j \in A_i} \frac{G_{n+1}(\tilde{w}_i - \tilde{w}_j, \Psi_{n+1})}{\sum_{t_k \in A_i} G_{n+1}(\tilde{w}_i - \tilde{w}_k, \Psi_{n+1})} \Psi_{n+1}^{-1}(\tilde{w}_i - \tilde{w}_j) \tag{8}$$

where $\partial \tilde{w}_i / \partial \tilde{x}_i$ projects the $n+1$ dimensional vector $\tilde{w}_i$ onto the dimension associated with the element $\tilde{x}_i$. In previous work [1] we have shown that, a timestep of $|T|\sigma_P^2$ corresponds to a mean-shift procedure on the conditional PDFs; that is, each data value moves to the weighted average of the sample data.

4.3 Implementation Issues

This section discusses several practical issues that are crucial for the effectiveness of the entropy reduction and prior estimation on image neighborhoods. A more detailed discussion on these issues is given in [1].

Parzen-window kernel width: Parzen-window density estimates, using finitely many samples, are greatly sensitive to the value of the Gaussian kernel σ_P [6]. The particular choice of σ_P is related to the sample size $|A_i|$ in the stochastic approximation. We automatically compute an *optimal* σ_P, that minimizes the average entropy of all conditional PDFs in the image, via a Newton-Raphson optimization scheme. Our experiments show that for sufficiently large $|A_i|$ additional samples do not significantly affect the estimates of entropy and σ_P, and thus $|A_i|$ can also be generated automatically from the input data.

Stationarity and local sampling strategies: In practice, image statistics are not homogeneous, and statistics for most images are more accurately modeled as piecewise stationary ergodic. Thus the set A_i of samples used to evaluate entropy and process pixel t_i should consist of pixels that are spatially near t_i. To achieve this, we choose a unique set of samples for each pixel t_i at random from a Gaussian distribution on the image coordinates, centered at t_i with standard deviation 30. Thus, the set A_i comprises pixels biased to be more near t_i. This strategy gives consistently better results than uniform sampling, and we have found that the it performs well for virtually any choice of the standard deviation that encompasses more than several hundred pixels. For this sampling strategy, $|A_i|$ is automatically computed to be 500 for all examples in the paper.

Neighborhood shape and size: Larger neighborhoods generally yield better results but take longer to compute. Typically 9×9 neighborhoods suffice, and we use them for the results in this paper. To obtain rotational invariance we use a metric in the feature space (neighborhood mask) that controls the influence of each neighborhood pixel by making distances in this space less sensitive to neighborhood rotations. Likewise image boundaries are handled through anisotropic metrics that do not distort the neighborhood statistics of the image.

Computation: The computational complexity of the proposed method is significant: $O(|T||A_i|E^D)$ where D is the image dimension and E is the extent of the neighborhood along a dimension. This is exponential in E, and our current results are limited to 2D images. The literature suggests some potential improvements (e.g. [23]). However, the purpose of this paper is to introduce the theory and methodology—algorithmic improvements are the subject of future work.

5 Experiments and Results

We show results using (a) real T1 noisy data, as well as (b) simulated MR data (181×217 pixels) obtained via BrainWeb [2] for unimodal and multimodal denoising. We simulate Rician noise by adding zero-mean Gaussian noise to the real and imaginary parts of the simulated MR data and taking the magnitude. For entropy minimization in the functional 6, the time step $\lambda_1 = |T|\sigma^2$ ($\equiv$ mean-shift update) can lead to oscillations, because of interactions of neighborhoods

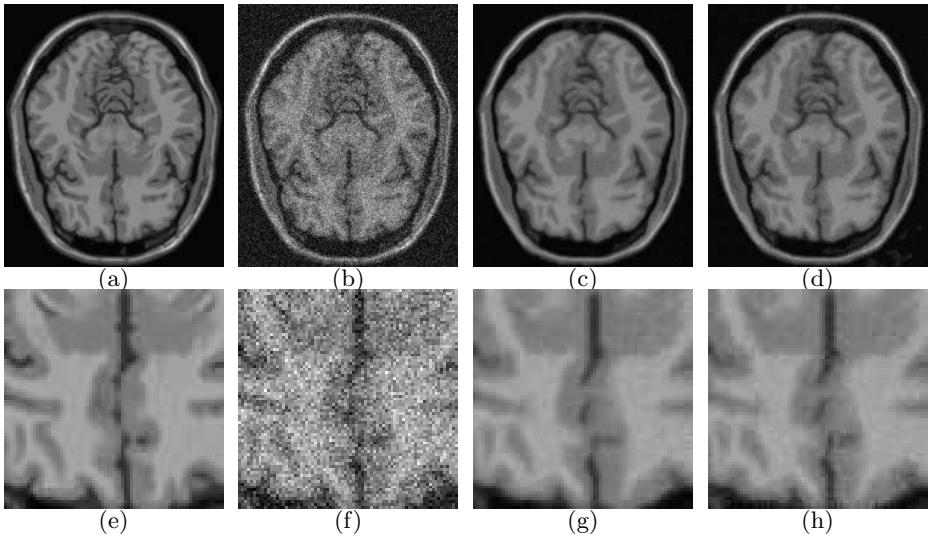


Fig. 3. (a) Noiseless T1 image. (b) Noisy image (gray matter SNR 12db, normalized squared error 1.0). (c) PDF-restored image (13 iterations) (as described in Section 4.1). (d) Denoised image (5 iterations, gray matter SNR 23db, normalized squared error 0.16). (e)-(h) show zoomed insets of images (a)-(d).

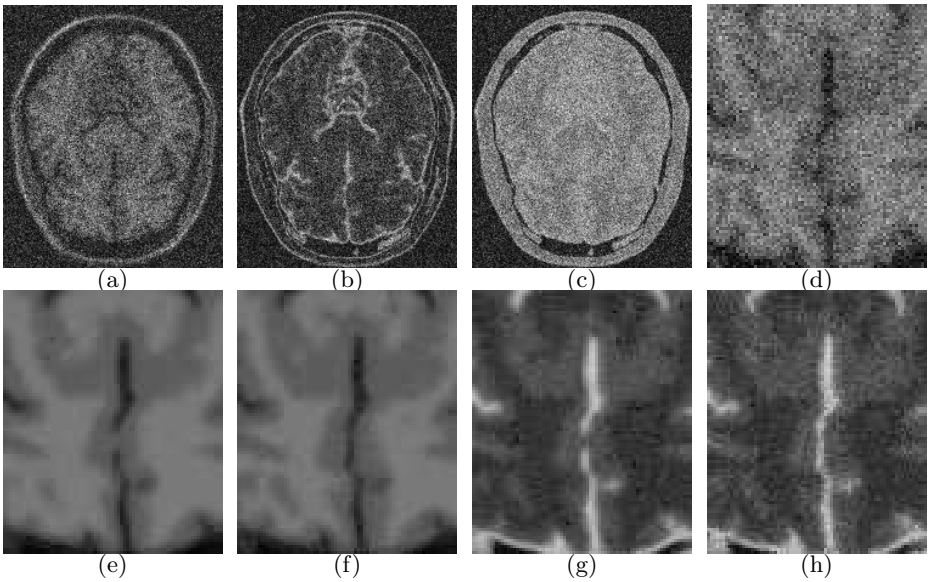


Fig. 4. Multimodal denoising. (a)-(c) Noisy T1, T2, PD images (signal intensity range 0:100, underlying noise $N(0, 400)$) (d) Zoomed inset of noisy T1 image. (e),(f) Zoomed insets of PDF-restored (as described in Section 4.1) and denoised T1 images. (g),(h) Zoomed insets of PDF-restored and denoised T2 images.

from one iteration to the next. We have found that a time step of $\lambda_1 = 0.2|T|\sigma^2$ alleviates this effect. We fix $\lambda_2 = 0.2$. We compute SNR as $20\log(x/\sigma)$ where x is the signal magnitude and the (estimated) underlying noise PDF is $N(0, \sigma^2)$. Each iteration on these data sets takes about 2 minutes on a Pentium-IV machine.

Multimodal denoising entails a simultaneous denoising of T1, T2, and PD images in a coupled manner, treating the combination of images as an image of vectors with the PDFs in the *combined* probability space. Although this paper shows results with multimodal images that are well aligned, we have evidence that the denoising is fairly robust to minor misregistration errors. The results show that incorporating more information in the denoising framework, via images of multiple modalities, produces consistently better results.

Figure 3 shows a denoising example using T1 data (SNR 12db) for the gray matter. With a normalized sum of squared pixel errors for the noisy image as 1.0, the denoised image has a squared error of 0.16. In general, the PDF-restored image (as described in Section 4.1) appears more smooth than the denoised image and may have less error. However the restoration of the neighborhood PDFs can produce some loss of structure, and the subsequent Bayesian estimation, which retains a fidelity to the input data, helps retain some of those details. We can see this behavior in the regions corresponding to the cerebro spinal fluid. This is even more clear in the next denoising example in Figure 4. With the same underlying noise PDF the normalized squared errors for the T2 and PD

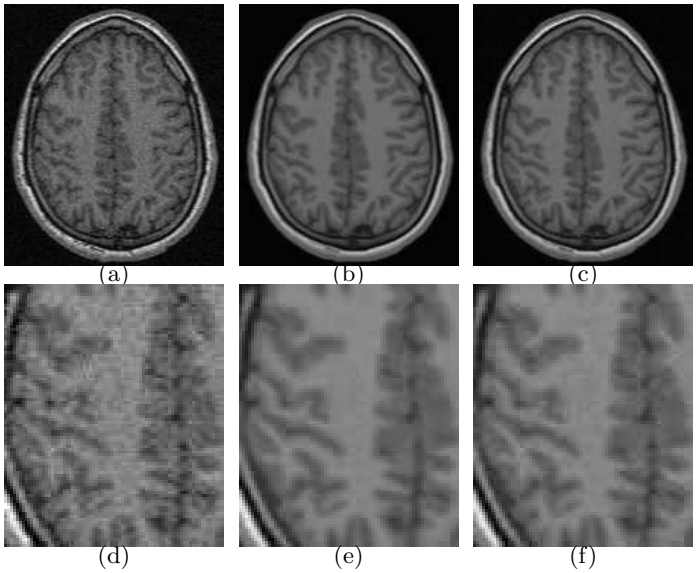


Fig. 5. (a) Real noisy image. (b) PDF-restored image (as described in Section 4.1). (c) Denoised image. (d)-(f) are zoomed insets of (a)-(c).

modalities are 0.3 and 0.19, respectively. Performing multimodal denoising with T1, T2, and PD data gives improved normalized squared errors of 0.10, 0.29, and 0.16, respectively.

Figure 4 shows T1, T2 and PD images (signal intensity range 0:100) with the underlying noise PDF as $N(0, 400)$. The SNR is 6db for the gray matter. Here, with a normalized squared error for the noisy image as 1.0, the squared error for the T1, T2, and PD denoised images are 0.08, 0.32, and 0.09 respectively. The squared error for T1 is significantly better than results in [13] for an equivalent gray matter SNR. Multimodal denoising, using T1, T2 and PD all together, gives normalized squared errors as 0.06, 0.17 and 0.07, respectively. Figure 5 shows results using real T1 noisy MRI data.

Acknowledgments

This work was supported by the NSF grant EIA0313268 and the NSF CAREER grant CCR0092065.

References

1. S. Awate and R. Whitaker. Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. *To appear in Proc. IEEE Int. Conf. Computer Vision Pattern Recog. 2005.*

2. D. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes, and A. Evans. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imag.*, 17(3):463–468, 1998.
3. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, 2002.
4. V. de Silva and G. Carlsson. Topological estimation using witness complexes. *Symposium on Point-Based Graphics*, 2004.
5. E. Dougherty. *Random Processes for Image and Signal Processing*. Wiley, 1998.
6. R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, 2001.
7. A. Fan, W. Wells, J. Fisher, M. Çetin, S. Haker, R. Mulkern, C. Tempany, and A. Willsky. A unified variational approach to denoising and bias correction in mr. In *Info. Proc. Med. Imag.*, pages 148–159, 2003.
8. G. Gerig, R. Kikinis, O. Kubler, and F. Jolesz. Nonlinear anisotropic filtering of mri data. *IEEE Trans. Med. Imag.*, 11(2):221–232, 1992.
9. D. Healy and J. Weaver. Two applications of wavelet transforms in magnetic resonance imaging. *IEEE Trans. Info. Theory*, 38(2):840–860, 1992.
10. M. Hilton, T. Ogden, D. Hattery, G. Jawerth, and B. Eden. Wavelet denoising of functional MRI data. pages 93–114. 1996.
11. M. Lysaker, A. Lundervold, and X. Tai. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Trans. Imag. Proc.*, 2003.
12. J. Mangin. Entropy minimization for automatic correction of intensity nonuniformity. In *IEEE Work. Math. Models Biomed. Imag. Anal.*, pages 162–169, 2000.
13. R. Nowak. Wavelet-based rician noise removal for magnetic resonance imaging. *IEEE Trans. Imag. Proc.*, 8:1408–1419, '99.
14. S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.
15. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, July 1990.
16. J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Imag. Proc.*, 12(11):1338–1351, 2003.
17. D. Scott. *Multivariate Density Estimation*. Wiley, 1992.
18. C. Shannon. A mathematical theory of communication. *Bell System Tech. Journal*, 27:379–423, July 1948.
19. J. Sled, A. Zijdenbos, and A. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Trans. Med. Imag.*, 17:87–97, 1998.
20. P. Viola and W. Wells. Alignment by maximization of mutual information. In *Proc. Int. Conf. Comp. Vision*, pages 16–23, 1995.
21. T. Weissman, E. Ordentlich, G. Seroussi, S. Verdu, and M. Weinberger. Universal discrete denoising: Known channel. *HP Labs Tech. Report HPL-2003-29*, 2003.
22. W. Wells, E. Grimson, R. Kikinis, and F. Jolesz. Adaptive segmentation of mri data. In *Proc. Int. Conf. on Comp. Vision*, pages 59–69, 1995.
23. C. Yang, R. Duraiswami, N. Gumerov, and L. Davis. Improved fast gauss transform and efficient kernel density estimation. In *Proc. Int. Conf. Comp. Vision*, pages 464–471, 2003.
24. Y. Zhang, M. Brady, and S. Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imag.*, 20(1), 2001.