IEEE Transactions on Medical Imaging

Diffusion Weighted Image Reconstruction: A Bayesian Approach

Journal:	*IEEE Transactions on Medical Imaging*
Manuscript ID	Draft
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Rowe, Daniel; Marquette University, Mathematical and Statistical Sciences; Medical College of Wisconsin, Department of Biophysics
Keywords:	Diffusion weighted imaging < Imaging modalities, Probabilistic and statistical methods < General methodology

Diffusion Weighted Image Reconstruction: A Bayesian Approach

Mustafa M. Farrah, *Member, IEEE*, Emad I. Abdelsalam, and Daniel B. Rowe

***Abstract*—Diffusion weighted imaging (DWI) is a Magnetic Resonance Imaging (MRI) technique based on measurement of random Brownian motion of water molecules within a voxel of tissue. Recently, DWI has become an important tool in in the diagnoses and early identification of medical conditions such as ischemic stroke, differentiation of epidermoid cysts from arachnoid cysts, assessment of cortical lesions in Creutzfeldt Jakob disease (CJD), assessment of active demyelination, and many other conditions. DWI suffers from a low Signal-to-Noise Ratio (SNR) due to low signal in the presence of noise during the imaging process. Although many techniques have been proposed to improve SNR of DWI, these techniques are time consuming and impractical. However, several denoising techniques such as a Principal Component Analysis (PCA), the non-local-means algorithm, and the Discrete Cosine Transform (DCT) have shown good performance. During a DWI scan, a sequence of images are acquired at varying b-values, a DWI factor. Higher b-values lead to stronger diffusion weighting but lower SNR images. An image with no diffusion, known as the b0 image is usually acquired. This b0 image has less noise than the other DWI images in the sequence, hence can be used to improve the quality of those images. In this work we developed a Bayesian approach to de-noise DWI images. Our approach uses the entropy between the b0 image and the other DWI images in the sequence to de-noise. Although our approach has similar processing time, it showed improved image SNR performance when compared with other leading techniques.**

***Index Terms*— MRI, DWI Image, Image Reconstruction.**

I. INTRODUCTION

Diffusion-weighted imaging (DWI) has been used to help detect early stroke, abscess, epidermic, cellular tumors and other neurological diseases [1]. The use of faster magnetic resonance imaging (MRI) techniques made it possible to acquire diffusion-weighted images of the whole-body. These techniques include multichannel coils, echo planar imaging, and stronger gradients [2]. DWI has also been applied in oncology to image the liver [3], prostate gland [4], breasts, as well as whole-body imaging [5].

DWI depends on the constant motion of water molecules. The motion depends on the thermal energy of the molecules and their change of concentration. The difference in the concentration of the molecules, allows the molecules to flow from the high concentration areas to low concentration areas. The flow depends on the change of concentration in the x direction and can be described by:

$$F = -D\frac{dC}{dx}$$

where F is the diffusion flux (the amount of substance per unit area per unit time), D is the diffusion coefficient (it controls the amount of substance that will flow through a small area during a small time interval), C is the concentration or the amount of substance per unit volume, and x is the position.

Ideal diffusion is described by Brownian motion and is given by:

$$C = C_0 exp\left[-\frac{x^2}{2\sigma^2}\right]$$

where C_0 is the concentration at time t=0. The concentration follows a Gaussian distribution in space. Diffusion is a random process of molecular movement which can define the microscopic movement of atoms or molecules in a solution or gas.

In living tissues, molecules of water, and other chemicals, flow freely through the various tissue of the body. However, in certain pathological conditions such as a tightly packed cells of tumors, the movement of molecules is restricted. MRI can be used to evaluate the diffusion restriction [6].

The movement of unrestricted molecules- in most fluids- is usually homogeneous; it usually spreads in all spatial directions. This is referred to as isotopic diffusion. Cell

M. M. Farrah, is with the Department of Electrical Engineering, University of Wisconsin Milwaukee, Milwaukee, WI 53203 (mfarrah@uwm.edu).
E. I. Abdelsalam is with the Department of Electrical Engineering, Al Hussein Technical University, Amman, Jordan 11831 (emad.abdelsalam@htu.edu.jo).
D. B. Rowe is with the Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI 53233 USA daniel.rowe@marquette.edu), and also with the Department of Biophysics, the Medical College of Wisconsin 53226 (dbrowe@mcw.edu).

membranes, vascular structures, axon cylinders, and fibers, are highly structured-they restrict the diffusion of molecules. When the movement of the molecules is not the same in all directions the diffusion is described as anisotropic [7].

Using MRI to Measure Diffusion

Stejskal and Tanner [8] described an MR experiment that could be applied to the detection and quantification of water diffusion in-vivo [3]. They used a spin echo sequence to measure the diffusion of water. A spin echo sequence is a 90° radio frequency (RF) pulse followed by a 180° refocusing pulse. They applied a symmetric pair of diffusion-sensitizing pulses around the 180° refocusing pulse as shown in Fig. 1.

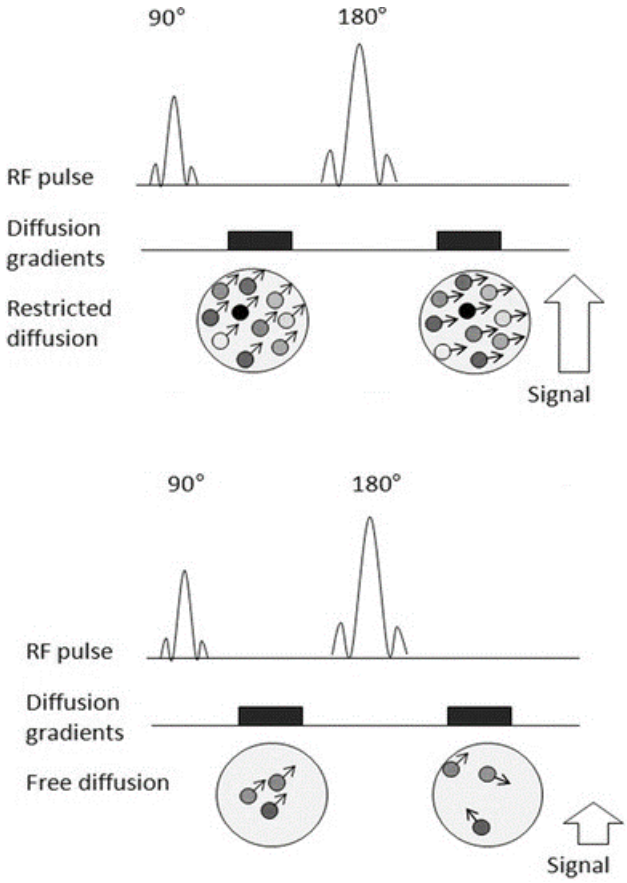


Figure 1 Restricted Diffusion vs. Free Diffusion[1].

The first gradient pulse which is applied before the 180° pulse introduces a phase shift that depends on the strength of the gradient pulse at the position of the spin (at $t = 0$). Then the 180° pulse is applied which will reverse the phase shift induced by the first gradient pulse. Then a second diffusion gradient pulse will be applied. The diffusion gradient causes the field intensity to vary with position. In the case of restricted diffusion, as shown on the top side of Fig. 1, the movement of the water molecules will be restricted. Therefore, after the second gradient pulse, all molecules will be at the same location resulting in high signal intensity. However, in the case of the free diffusion, as shown on the bottom-side of Fig. 1, water molecules will not be at the same location when applying the second diffusion gradient pulse. This will cause the molecules to accumulate different phase values. Hence, the molecules signals will not be completely rephased by the second gradient; which will result in a low signal intensity. Having said that, the degree of diffusion can be inferred by measuring the attenuation of the signal intensity.

One application of DWI is in the early detection of a stroke. Fig. 2 shows diffusion MRI images of a brain that was taken for a patient who had a stroke. The bright area on the left hand-side of the brain indicates high diffusion signal, which indicates restricted diffusion due to stroke [3].

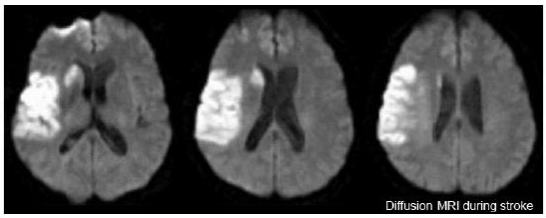


Figure 2 Diffusion MRI Brain Image during a stroke: The bright areas on the left side of the brain indicates restricted diffusion due a stroke[2].

Diffusion Gradients and the b-value:

In this section, the pulse sequence that was proposed by Stejskal and Tanner to detect and quantify the diffusion of water [8][9] is described. The proposed pulse sequence is shown in Fig. 3, where RF is the RF pulse, *GS* is the slice selective pulse, *GP* is the phase encoding pulse, and *GM* is the readout direction pulse.

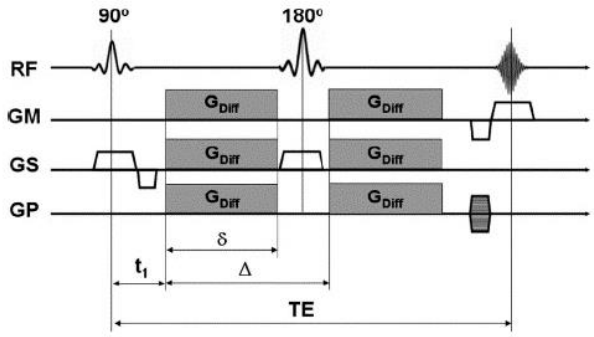


Figure 3 Diffusion Weighted Images Pulse Sequence[3].

This pulse sequence is a spin echo pulse with two diffusion gradient pulses added around the 180° pulse. The diffusion gradient pulse duration is δ, and the time between the start of the two pulses is Δ. The phase shift of a single static spin in the presence of a magnetic field gradient is given by:

$$\emptyset(t) = \gamma B_0 t + \gamma \int_0^t G(\tau).x(\tau)d\tau$$

where γ is the gyromagnetic ratio, B_0 is the strength of the external magnetic field, $G(\tau)$ is the strength of the diffusion gradient, and $x(\tau)$ is the spatial location of the spin. The amount

of phase shift depends on several factors including: the strength of the applied gradient, the time duration that the gradient will be applied for, and the spatial location of the spin. At the echo time *TE*, which represents the time in milliseconds between the application of the 90° pulse and the peak of the echo signal, this phase will be given by:

$$\emptyset(TE) = \gamma \int_{t_1}^{t_1+\delta} G(\tau).x(\tau)d\tau - \gamma \int_{t_1+\Delta}^{t_1+\Delta+\delta} G(\tau).x(\tau)d\tau$$

where δ is the duration of each sensitizing gradients and Δ is time between the start of the two pulses. The diffusion weighted signal at time TE is given by:

$$DWI(b,TE) = I_0 exp\left(-\frac{TE}{T_2}\right)\exp(-bD) \quad (1)$$

where I_0 is the signal intensity in the b_0 image which is a non-diffusion MRI image, and T_2 is the transverse relaxation time which determines the rate at which excited molecules go out of phase with each other. This signal depends on the apparent diffusion coefficient (ADC), which is a measure of the magnitude of diffusion within tissue, and a sensitization parameter known as the b-value which is given by:

$$b = \gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3})$$

and is measured in s/mm^2. The b-value can be set by changing the duration of the pulse, the time interval between the two pulses or the strength of the gradient. In practice, on clinical MR scanners, the diffusion sensitivity is varied by changing the strength of the sensitization gradient.

Diffusion weighted Images vs. Apparent Diffusion Coefficients:

The DWI signal in (1) has two signal factors, T_2 signal ($exp(-TE/T_2)$), and diffusion ($exp(-bD)$). In some cases, it is not obvious if a bright area in an image is due to restricted diffusion or due to the T_2 image contrast. Fig. 4 illustrates this situation. The bright bulbs on the left-hand side of the graph can indicate restricted area. The left-hand side of the graph is a T_2 image. It shows bright bulbs on the left-hand side. The middle image is from DWI. It also has the same bright bulbs which makes it hard to distinguish if these two bulbs are due to restricted diffusion or the T_2 images. The image on the right is the ADC where the bright bulbs correspond to regions without restricted diffusion.

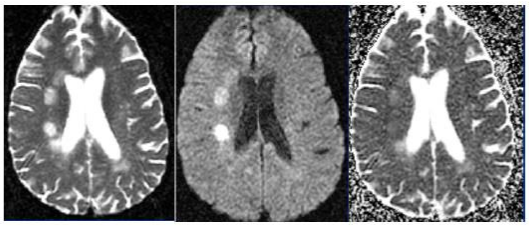


Figure 4 T2 shine through. Left image is a T2 image, middle image is a DWI, and right is an ADC[4].

To eliminate the T_2 factor, two images can be taken; the first one will be T_2 image with no diffusion. This can be done by setting the b-value equal to zero, so that Equation (1) becomes:

$$DWI(TE) = I_0 exp\left(-\frac{TE}{T_2}\right) \quad (2)$$

A second image at the desired b-value can be taken. If Equation 1 is divided by Equation 2, we obtain:

$$\frac{DWI(b,TE)}{DWI(TE)} = exp(-bD) \quad (3)$$

This division is calculated pixel by pixel in the two images. The *ADC* can be computed by:

$$D = -\frac{1}{b} log\left(\frac{DWI(b,TE)}{DWI(TE)}\right) \quad (4)$$

In Fig. 5, a diffusion weighted image is shown in the left side and the corresponding ADC image is shown in the right. The DWI image shows a bright bulb on the right side of the brain which can be due to restricted diffusion, ADC image shows the bulb dark, which indicates restricted diffusion is not caused by the T_2 effect.

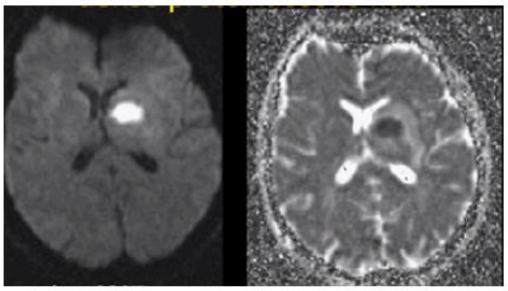


Figure 5 Diffusion Weighted Image vs. Apparent Diffusion Coefficient[5].

Optimum b-value and Its Relation to the Signal Intensity:

The sensitivity of the DWI sequence to water motion can be varied by changing the b-value which is proportional to the three factors mentioned before. If the b-value is increased, the contrast between different tissues will become more apparent; however, the image will have more noise. Fig. 6 illustrates the logarithm of relative signal intensity (SI) versus the b-value for tumor and normal tissue. The slope of the "tumor line" is less than that of the line representing normal tissue, which translates into a lower signal on the ADC map [3]. If we increase the b-value, the contrast between the tumor and the normal cells will also be increased, however, this decreases the SNR value, so there is a trade-off between the SNR and the contrast in the image.

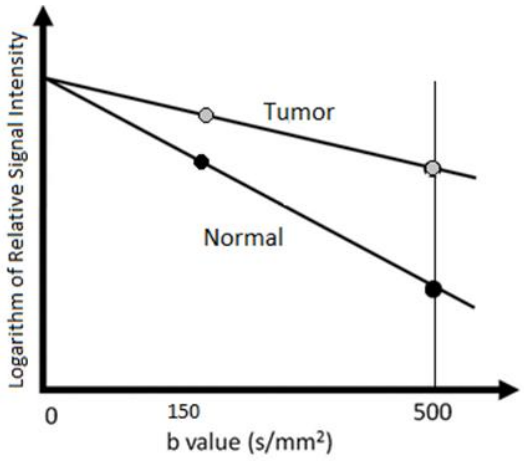


Figure 6 Effect of b-values[6].

If the b-value is increased, then the contrast between the normal tissue and the tumor will be increased, but the strength of the signals will be decreased as shown in Fig. 6. Since the noise remains the same, increasing the b-value will decrease the signal intensity and hence lower the signal to noise ratio (SNR).

Noise in MRI

Signals in MRI images are measured using quadrature detectors which measure the MRI signal from two orthogonal directions. These detectors give the real and imaginary part of the signals. The measured signals are contaminated with Gaussian noise. MRI images are formed by taking the Fourier transform of the measured signals. Therefore, the MRI image will be a complex image. The noise distribution of the real part as well as the imaginary part of the complex image, are Gaussian. The magnitude image is then formed by a nonlinear mapping. It has been shown by many studies that noise distribution in the magnitude image follows a Rice distribution [10].

The Rice distribution has a probability density function (pdf) given by:

$$f(x|v,\sigma) = \frac{x}{\sigma^2} exp\left(\frac{-(x^2+v^2)}{2\sigma^2}\right) I_0\left(\frac{xv}{\sigma^2}\right) \quad (5)$$

where $I_0(x)$ is the modified Bessel function of the first kind given by:

$$I_\alpha(x) = \frac{1}{\pi}\int_0^\pi exp(x\cos(\theta))\cos(\alpha\theta)d\theta - \frac{\sin(\alpha\pi)}{\pi}\int_0^\pi \exp(-x\cosh t - \alpha t)\,dt \quad (6)$$

and non-centrality parameter v and scale parameter $\sigma > 0$. The noise distribution in MRI contains the modified Bessel function of the first kind with zeroth order, which is given by

$$I_0(x) = \frac{1}{\pi}\int_0^\pi exp(x\cos(\theta))\cos(0\theta)d\theta - \frac{\sin(0\pi)}{\pi}\int_0^\pi exp(-x\cosh t - 0t)\,dt$$

$$I_0(x) = \frac{1}{\pi}\int_0^\pi exp(x\cos(\theta))d\theta \quad (7)$$

In Fig. 7 we show the Rice distribution for different values of $v = N/\sigma$ where N is the mean of the signal and σ is the standard deviation.

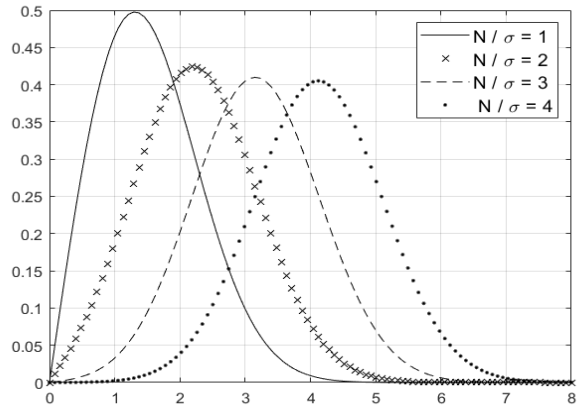


Figure 7 The Rician Distribution for Several Signal-to-Noise-Ration Values, N/σ.

The graph shows that when the SNR is greater than 3 ($N/\sigma \geq 3$) the Rice distribution can be approximated by the Gaussian distribution [10].

II. Contemporary Denoising Methods

Several methods have been proposed to denoise MRI images. We will compare our work to three methods. The first method uses the local 3D Discrete Cosine Transform (DCT) with thresholding and averaging proposed by Guleryuz (2007) [26]. This method makes use of the sparseness of an image in another transformation. Sparseness means that the image can be represented by a small number of basis functions. This happens if the representation of the image in another domain or basis, such as the DCT, has many near zero coefficients that can be disregarded without affecting the image itself. In this method a sliding-window DCT was performed and several thresholding methods were used. The second method that we will compare our work to is the non-local means filter denoising proposed by Buades [2]. In this method, the image is assumed to have many similarities at different locations and hence similar pixels can be at different location in the image and not necessary close to each other. When denoising a pixel, the image is divided into blocks and then pixels with similar neighborhoods tend to have similar original values and hence similar blocks will be averaged to denoise the pixel.

The third method is the overcomplete local principal

component analysis as proposed by Manjón [25]. In his method, instead of dividing each image into patches, a series of *k* images were used to form a matrix to be used to reduce the noise. Manjon used a series of *k* images to denoise DWI images. He utilized the idea that corresponding pixels in a series of *k* DWI images in *k* direction are very similar and hence these patches can be used to reconstruct a matrix that will be filtered.

For a series of *k* images, a block of 4×4×4 voxels is converted into a column of matrix X. Hence, the size of matrix *X* will be 64×*k*. Then the PCA algorithm will be applied to matrix *X*, to compute its approximation $\hat{X}$ from which the patches can be reconstructed.

III. Diffusion Weighted Imaging Reconstruction Using Bayesian Approach

In the previous Section, we explained the principle of DWI and the limitations that are imposed by the noise. We also discussed three methods that have been applied to de-noise DWI. In this Section, we propose a new Bayesian method to reconstruct DWI images. In this approach [11], the joint entropy between the DWI image and the prior image is used as a measure of similarity. Since the proposed method is based on the Bayesian approach and the joint entropy, we call it Entropy Bayesian Approach (EBA). The EBA algorithm is noise level independent in the image, i.e. no matter how much noise the image has, the algorithm can improve it. This will make it possible to use this method with very high b-values where other methods are limited. If complex-valued images are available, EBA can be readily applied and should be more accurate since the approximation on the noise can be removed. The EBA method can be applied to a single DWI utilizing a prior image, while other methods, such as the PCA, require a sequence of images. We will explain the Bayesian approach in the next Section, the image entropy in the next Section and then the EBA algorithm.

Image Entropy

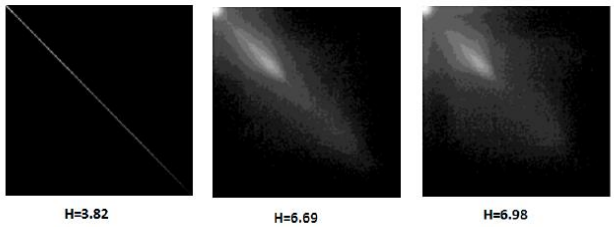
Entropy is a measure of the uncertainty in a distribution of a random variable. Shannon defined the entropy $H(x)$ of a discrete random variable X with possible values x_1, ..., x_n and probability mass function $P(X)$ as:

$$H(X) = -\sum_i P(x_i)\, log_b\, P(x_i) \tag{8}$$

where *b* is the number of bits used to represent the outcome of the experiment. The joint entropy $H(X,Y)$ of a pair of discrete random variables with a joint distribution $p(x,y)$ is a simple extension that measures the uncertainty in the joint distribution of a pair of random variables [12]. Joint entropy $H(X,Y)$ is defined as:

$$H(X,Y) = -\sum_{i,j} p(x_{ij}, y_{ij}) \ln p(x_{ij} y_{ij}) \tag{9}$$

For digital images, the joint entropy of two images can be computed using the joint histogram of the two images. The joint histogram is computed using the intensities of the corresponding pixels of the two images, and the value in the histogram represents the number of occurrences of intensity value pairs [13][14].



(a) Joint histogram of aligned images (b) Joint histogram of slightly misaligned images (c) Joint histogram of two images that are misaligned

Figure 8 Joint Entropy of Images.

Fig. 8 shows an example of the joint histogram of two images and the corresponding joint entropy. When the two images are identical and aligned with each other, the joint histogram will not be spread out and it has the smallest entropy.
In Fig 8 (b) and (c), the joint histograms of misaligned images are shown, the more the two images are misaligned or different, the more the joint histogram will be spread out and the joint entropy will be higher.

Bayes' Theorem

Bayes' theorem [15][10] allows us to write $p(x|y) = p(y|x)p(x)/p(y)$ where the likelihood, $p(y|x)$, describes the data formation process for a particular underlying image while the prior, $p(x)$, encodes any prior beliefs about the properties of such underlying images. In this work, we propose a new method to reconstruct DWI using a Bayesian approach by incorporating prior information from another image that has high SNR. The prior image can be a T_2 weighted image or it can be another DWI with lower b-value that has higher SNR. While the Bayesian approach has been used in many applications such as image registration [16][17][18], it has not been successfully used in DWI yet.

Unlike other applications with real images, DWI has complex-valued data which require a different approach. Numerous solutions have been proposed to solve functions with complex variables [19] [20]. Here, we treat the complex-valued image as two real images, one corresponds to its real part, and another corresponds to its imaginary part. Joint entropy has previously been used in many applications such as image registration [21][22][23][24]. In EBA, it is applied as a measure to quantify similarity. We assume that the DWI is corrupted with Gaussian noise. A typical model of a signal corrupted by noise is

$$y_{ij} = x_{ij} + \varepsilon_{ij} \tag{10}$$

where y_{ij} is the observed image pixel, x_{ij} is the true image pixel, and ε_{ij} is random Gaussian noise with a mean of 0 and a

standard deviation of σ. A valid assumption about the observed image is that the real and imaginary parts of each pixel follow independent Gaussian distributions with an identical standard deviation. This assumption is valid from the physical measurement process because they are measured by quadrature coils which constituent two pairs of identical coils oriented perpendicular to each other. As these coils are identical and uncoupled, the real signal from one pair of coils will not interfere with the imaginary signal from the second pair of coils. The distribution of the intensity of each image can be described by:

$$p(y_{ij}|x_{ij}) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left(\frac{-1}{2\pi\sigma^2}(y_{ij} - x_{ij})^2\right) \quad (11)$$

Using Bayes theorem, the posterior distribution can be found as:

$$p(x_{ij}|y_{ij}) = p(y_{ij}|x_{ij}).p(x_{ij})\,/p(y_{ij}) \quad (12)$$

where $p(y_{ij}|x_{ij})$ is the likelihood function, and $p(x_{ij})$ is the prior density. We assume the prior density follows a Gibbs distribution of the form:

$$p(x_{ij}|y_{ij}) = \frac{1}{w} e^{-\beta h(x_{ij}|g_{ij})} \quad (13)$$

where $h(x_{ij}|g_{ij})$ is the joint entropy between x_{ij}, g_{ij}:

$$h(x_{ij}|g_{ij}) = p(x_{ij}, g_{ij}).\ln p(x_{ij}, g_{ij}) \quad (14)$$

and w is a normalization constant. The prior image is the b0 image with pixel g_{ij}. Substituting Equations 11,12, 13, in Equation 14 we get

$$p(x_{ij}|y_{ij}) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left(\frac{-1}{2\pi\sigma^2}(y_{ij} - x_{ij})^2\right) .\frac{1}{w}\exp(-\beta p(x_{ij}|g_{ij}).\ln p(x_{ij}|g_{ij})) \quad (15)$$

where β is a parameter that controls the joint entropy term. If $\beta = 0$, the reconstructed image will be the same as the DWI. As β increases, the reconstructed image will become similar to the prior image or the b_0 image. Since the distribution of each pixel is identically independently distributed, we can write $p(x_{ij}|y_{ij})$ as shown in Equation 15 for all possible values of i and j.

If we take the natural logarithm of Equation 14 we obtain

$$\ln p(x_{ij}|y_{ij}) = \ln(2\pi\sigma^2)^{-\frac{1}{2}} + \frac{1}{w} - \frac{1}{2\sigma^2}\sum_{ij}(y_{ij} - x_{ij})^2 - \beta p(x_{ij}|g_{ij}) \ln p(x_{ij}|g_{ij}) \quad (16)$$

Focusing on an individual pixel and neglecting subscripts, the solution of this equation

$$\hat{x} = \underset{x}{argmax}\, p(x\,|y) \quad (17)$$

where $\underset{x}{argmax}\, p(x|y)$ is the set of points x for which $p(x|y)$ attains its largest value, can be found using nonlinear optimization technique like nonlinear conjugate gradient [6][31]. We used the Matlab built-in function fminunc to find the solution of Equation 16. The first two terms in Equation 16 are constants and the solution of the equation does not depend on them. Hence, they can omitted without affecting its solution.

The inputs of the program are as follows:

y: the DWI image.

g: the b_0 image

σ: the standard deviation of the DWI image which is computed by converting the DWI into a vector and calculating its standard deviation of the vector.

β: is the control parameter that is specified for each image and is in the range of .01 to 0.5

x_0: the initial value of the reconstructed image, which is taken to be the average of the DWI image and the b_0 image.

The output of the program is x which is the new reconstructed DWI image.

We use the default values from Matlab for the input parameters below:

DiffMinChange: the minimum change in variables for finite-difference gradients with the default of 10−8

MaxIter : the maximum number of iterations with the default value of 400.

IV. Experimental Results

The performance of the EBA algorithm has been compared to the three denoising methods discussed earlier. The data for comparisons is obtained from Laboratoire Bordelais de Recherche en Informatique1. There are 21 slices with each slice having 21 directions. The comparisons have been performed on a laptop computer with a 4-core 2 GHz i7 processor and 8 Giga Bytes Memory. The performance of the algorithms are evaluated basing on the processing time, the mean squared error of difference between the DWI and the filtered image, and on fidelity in preserving the fine details or the edges in the image. For each algorithm, three slices are used for evaluation. The first denoising method to compare is the DCT algorithm.

EBA Comparison with DCT:

To compare the EBA method to the DCT method, three slices (slices 3, 5 and 20) of a DWI are used. Figures 9, 10, and 11 show the results. In these figures, images (a) and (d) are the noisy DWI, image (b) is the denoised image with EBA algorithm, and image (c) is the difference between (a) and (b). Image (c) has some areas where the anatomy of the image can be seen. They appear when the prior and the DWI image are not totally aligned. The mean squared errors (MSE) are 8.58×10^{-4} for the slice 3, 7.39×10^{-4} for slice 5, and 7.325×10^{-4} for slice 20. The difference image in (c) shows that the noise was

suppressed from all areas of the image. Image (e) in Figures 9, 10, and 11 are the denoised DWI images with DCT algorithm. Areas in (e) look noisier than those of the EBA.

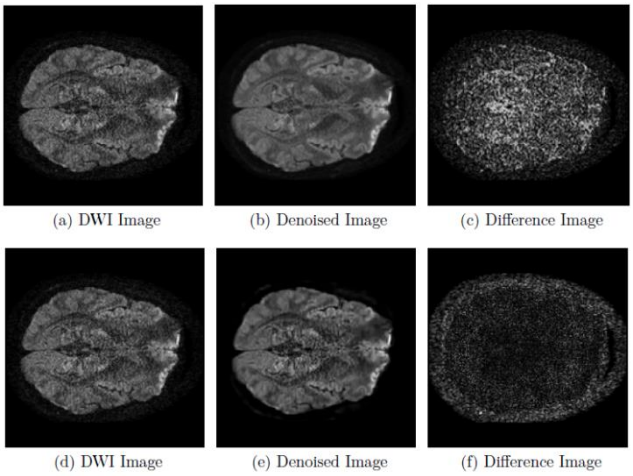


Figure 9 EBA ((b) and (c)) and DCT ((e) and (f)) Comparison, Slice 3.

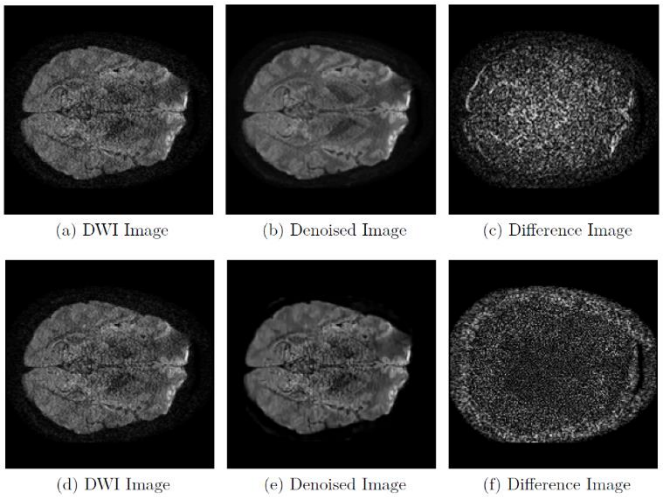


Figure 10 EBA ((b) and (c)) and DCT ((e) and (f)) Comparison, Slice 5.

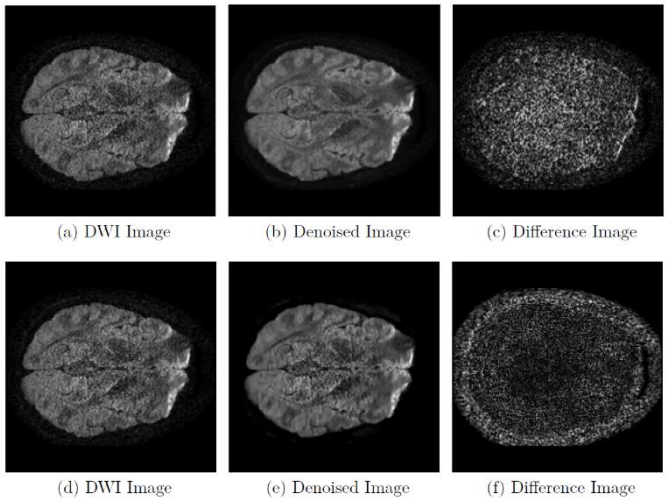


Figure 11 EBA ((b) and (c)) and DCT ((e) and (f)) Comparison, Slice 20.

The difference between the DWI image and the denoised image with DCT techniques is shown in image (f). The MSEs of the DCT algorithm are 9.06×10^{-4} for slice 3, 9.63×10^{-4} for slice 5, and 9.26×10^{-4} for slice 20. Consistently, the MSE of the DCT slices is higher than the MSE of the EBA slices. The difference images (f) in Figures 9, 10, and 11 show some dark areas in the middle, which indicate that there is no difference between the DWI image and the denoised image, i.e. the denoising algorithm is ineffective in these areas. The computation time of the DCT was 4 times higher than the computation time of the EBA.

EBA Comparison with NLM:

Next we compare the performance of the EBA method to the NLM algorithm. In NLM, noise is reduced according to the similarity in the image. The NLM algorithm is applied to the same slices used before and the results are compared to the EBA methods. Figures 12, 13, and 14 show the results for comparing EBA to NLM.

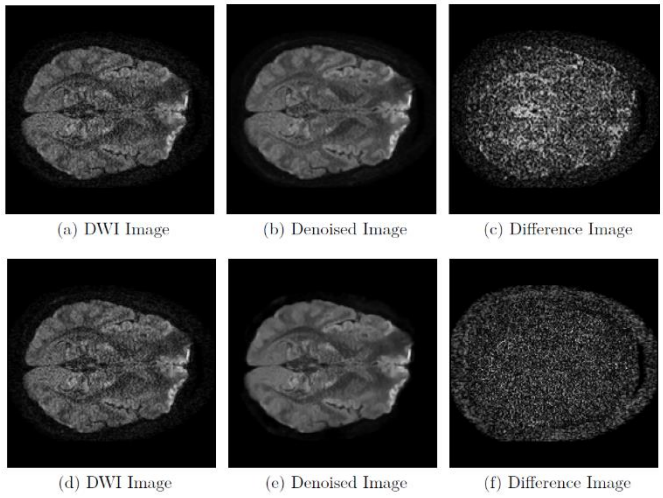


Figure 12 EBA ((b) and (c)) and NLM ((e) and (f)) Comparison, Slice 3.

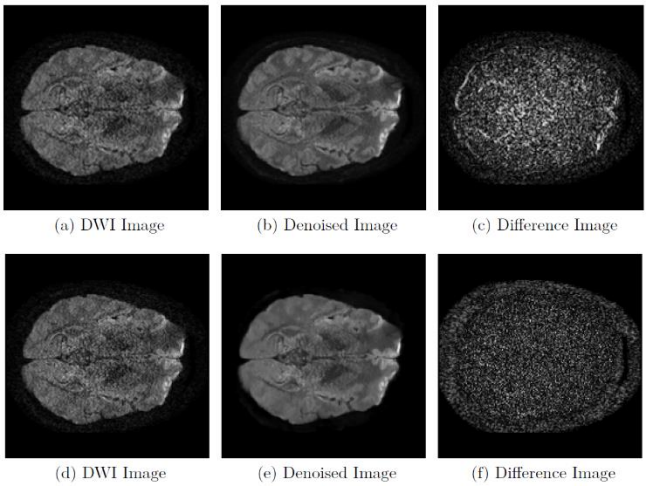


Figure 13 EBA ((b) and (c)) and NLM ((e) and (f)) Comparison, Slice 5.

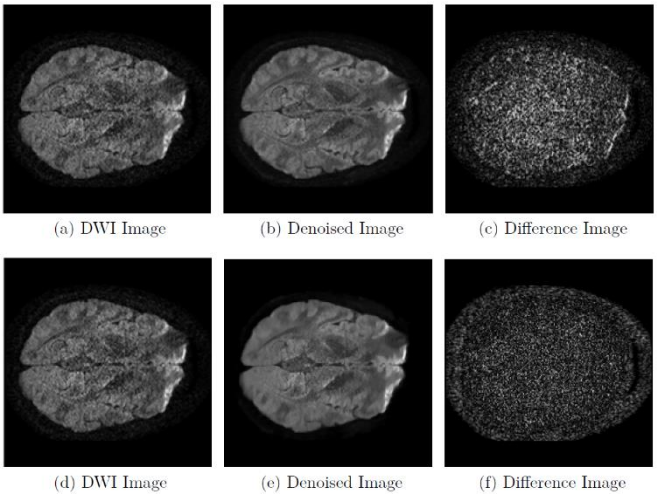


Figure 14 EBA ((b) and (c)) and NLM ((e) and (f)) Comparison, Slice 20.

The first three images in these figures, images (b) and (c), are the results of the EBA algorithm. Images (e) show the filtered images using the NLM algorithm. These images show that the noise is removed and the fine details are preserved in the image except in boundary areas. Images (f) show the differences between images (d) and (e). They show that the noise is removed from all parts of the image which is not the same case when the NLM algorithm is used. The MSEs for the NLM algorithm are 13×10^{-4} for slice 3, 12×10^{-4} for slice 5, and 13×10^{-4} for slice 20. These numbers are higher than the EBA numbers. Finally, the processing time for the NLM algorithm was about 30 s, which is the same as that of the EBA algorithm.

EBA Comparison with PCA:

Finally, we compared the EBA method to the PCA. Again the same DWI images were used. The results are shown in Figures 15, 16 and 17. The images in (a) and (d) show the DWI before denoising. Images (b) and (c) are the same ones shown in the previous Section. The denoised image using PCA is shown in (e). It does not show the details in the middle as that of the EBA algorithm in image (b). PCA acts like a low pass filter in some areas of the image.

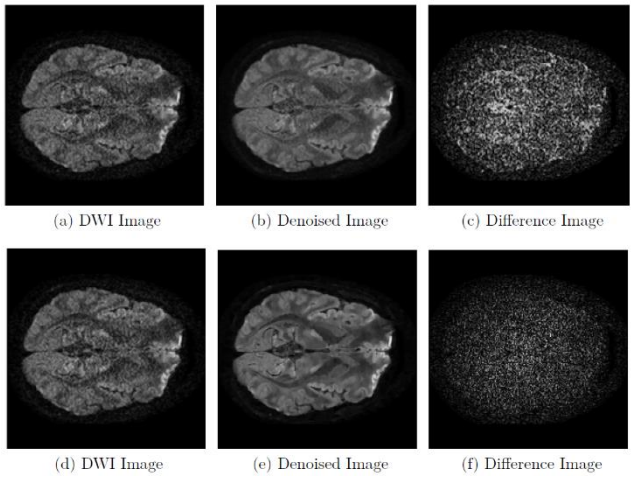


Figure 15 EBA ((b) and (c)) and PCA ((e) and (f)) Comparison, Slice 3.

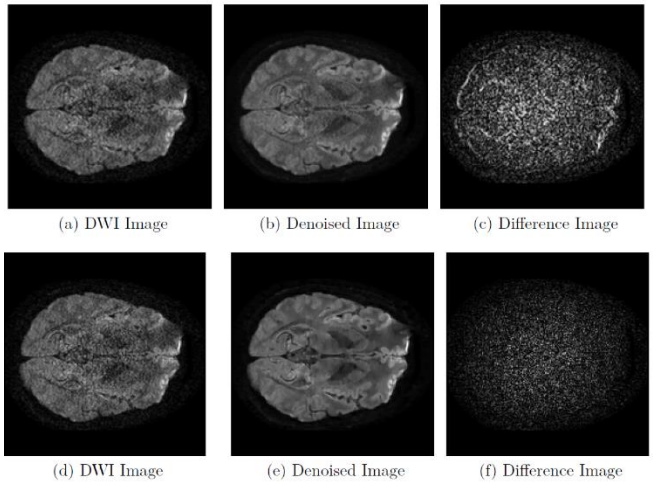


Figure 16 EBA ((b) and (c)) and PCA ((e) and (f)) Comparison, Slice 5.

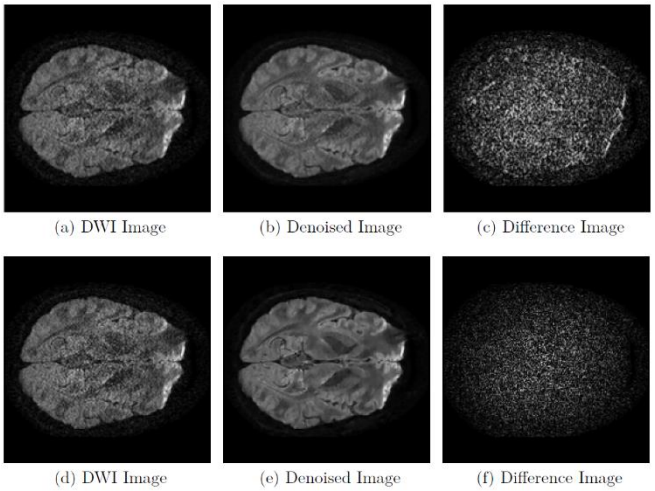


Figure 17 EBA ((b) and (c)) and PCA ((e) and (f)) Comparison, Slice 20.

The differences between images (d) and (e) are shown in images (f). It appears that the noise is removed from the image without affecting its details. The MSEs for all the three slices are 13×10^{-4}, which are higher than that of EBA algorithm. The processing time for this algorithm was about 30 sec which is the same as the processing time for the EBA algorithm.

Comparison Summary:

Here, we summarize our comparisons of the EBA algorithm to three different denoising methods in Table 1.

TABLE I
COMPARISON RESULTS BETWEEN EBA, DCT, NLM, AND PCA

Method	Mean Squared Error			Processing Time
	Slice 3	Slice 5	Slice 20	
EBA	8.58×10^{-4}	7.39×10^{-4}	7.325×10^{-4}	~30 Sec
DCT	9.06×10^{-4}	9.63×10^{-4}	9.26×10^{-4}	~120 Sec
NLM	13×10^{-4}	12×10^{-4}	13×10^{-4}	~30 Sec
PCA	13×10^{-4}	13×10^{-4}	13×10^{-4}	~30 Sec

The processing time of the EBA is about the same as that of the DCT and PCA. However, the processing time for the DCT algorithm is 4 times more than those of the other methods. The MSE of the EBA algorithm is the least among all the denoising methods. The EBA and the PCA perform better in keeping the fine details of the image, the NLM comes next and the DCT is the last. We can conclude that the EBA method is very competitive with the most predominant denoising methods for DWI.

REFERENCES

[1] A. A. Malayeri, R. H. El Khouli, A. Zaheer, and M. A. Jacobs. (2011, Oct). Principles and applications of diffusion-weighted imaging in cancer detection, staging, and treatment follow-up. *RadioGraphics*. 31(6), pp.1773–1791.

[2] A. Buades, B. Coll, and J. M. Morel (2005, Jan.). A Review of image denoising algorithms, with a new one. M*ultiscale Modeling & Simulation*, A SIAM Interdisciplinary Journal. 4 (2), pp. 490–530.

[3] D. M. Koh and D. J Collins. (2007, Jun.). Diffusion-weighted MRI in the body: applications and challenges in oncology. *American Journal of Roentgenology*. 188(6), pp. 1622–1635.

[4] K. Hosseinzadeh and S. D. Schwarz. (2004, Sep.). Endorectal diffusion-weighted imaging in prostate cancer to differentiate malignant and benign peripheral zone tissue. *Journal of Magnetic Resonance Imaging*. 20(4), pp. 654–661.

[5] T. C Kwee, T. Takahara, R. Ochiai, R. AJ Nievelstein, and P. R Luijten. (2008, Apr.). Diffusion-weighted whole-body imaging with background body signal suppression (DWIBS): features and potential applications in oncology. *European radiology*. 18(9), pp. 1937–1952.

[6] D. L. Bihan, et al. (2001, Apr.). Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging*. 13(4). pp. 534–546.

[7] C. Beaulieu. (2002, Nov.). The basis of anisotropic water diffusion in the nervous system–a technical review. *NMR in Biomedicine*. 15(7-8), pp. 435–455.

[8] E. O. Stejskal and J. E. Tanner (2004, Jul.). Spin diffusion measurements: spin echoes in the presence of time-dependent field gradient. *Journal of Chemical Physics*. 42(1), pp. 288–292.

[9] R. Bammer. (2003, Mar.). Basic principles of diffusion-weighted imaging. *European Journal of Radiology*. 45(3), pp. 169–184.

[10] H. Gudbjartsson and S. Patz. (2008, Feb.). The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine*. 34(6), pp. 910–914.

[11] J. Sheng, D. Liang, J. Tang, D. Huo, and L. Ying. (2009, Apr.). Bayesian reconstruction of DW-PROPELLER images using joint entropy. Proceedings of International Society of Magnetic Resonance in Medicine Scientific Meeting. 17, pp. 2861.

[12] T. M. Cover and J. A. Thomas. (2006, Jul.). *Elements of information theory* (*Wiley Series in Telecommunications and Signal Processing*). Wiley-Interscience, pp. 19–21.

[13] R. C. Gonzalez, R. E. Woods, S. L. Eddins. *Digital Image Processing Using MATLAB*. New Jersey, Prentice Hall, 2004, pp. 93–94.

[14] J. P. W. Pluim, J.B.A. Maintz, and M. A. Viergever. (2003, Aug.). Mutual-information-based registration of medical images: a Survey. *IEEE Trans. Med. Imag.* 22(8), pp. 986–1004.

[15] S. M. Smith, et al. (2004, Sep). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 23, pp. S208–S219.

[16] J. Tang and A. Rahmim. (2009, Nov.). Bayesian PET image reconstruction incorporating anato-functional joint entropy. *Physics in Medicine and Biology*. 54(23), pp. 7063–7075.

[17] F. M. A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. (1997, Apr.). Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.* 16(2), pp. 187–198.

[18] B. Zitova and J. Flusser. (2003, Aug.). Image registration methods: a Survey *Image and Vision Computing*. 21(11), pp. 977–1000.

[19] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, 2013.

[20] L. Sorber, M. V. Barel, and L. D. Lathauwer. (2012, Jul.). Unconstrained optimization of real functions in complex variables. *SIAM Journal on Optimization.* 22(3), pp. 879–898.

[21] Y. M. Zhu. (2002, Aug.). Volume image registration by cross-entropy optimization. *IEEE Trans. Med. Imag.* 21(2), pp. 174–180.

[22] J. Nuyts. (2007, Jan.). The use of mutual information and joint entropy for anatomical priors in emission tomography. Presented at IEEE Nuclear Science Symposium. 6, pp. 4149–4154

[23] W. Wein, S. Brunke, A. Khamene, M. R. Callstrom, and N. Navab. (2008, Jun.). Automatic CT-Ultrasound registration for diagnostic imaging and image-guided intervention. *Medical Image Analysis*, 12(5), pp. 577–585.

[24] M. Holden, et al. (2000, Feb.). Voxel similarity measures for 3-D serial MR brain image registration. *IEEE Trans. Med. Imag.* 19(2), pp. 94–102.

[25] J. V. Manjón, P. Coupé, L. Concha, A. Buades, D L. Collins, and M. Robles. (2013, Sep.). Diffusion weighted image denoising using overcomplete local PCA. *PloS One*, 8(9), pp. e73021.

[26] O. G Guleryuz (2007, Nov). Weighted averaging for denoising with overcomplete dictionaries. *IEEE Tran. Image Process.* 16(12), pp. 3020–3034.